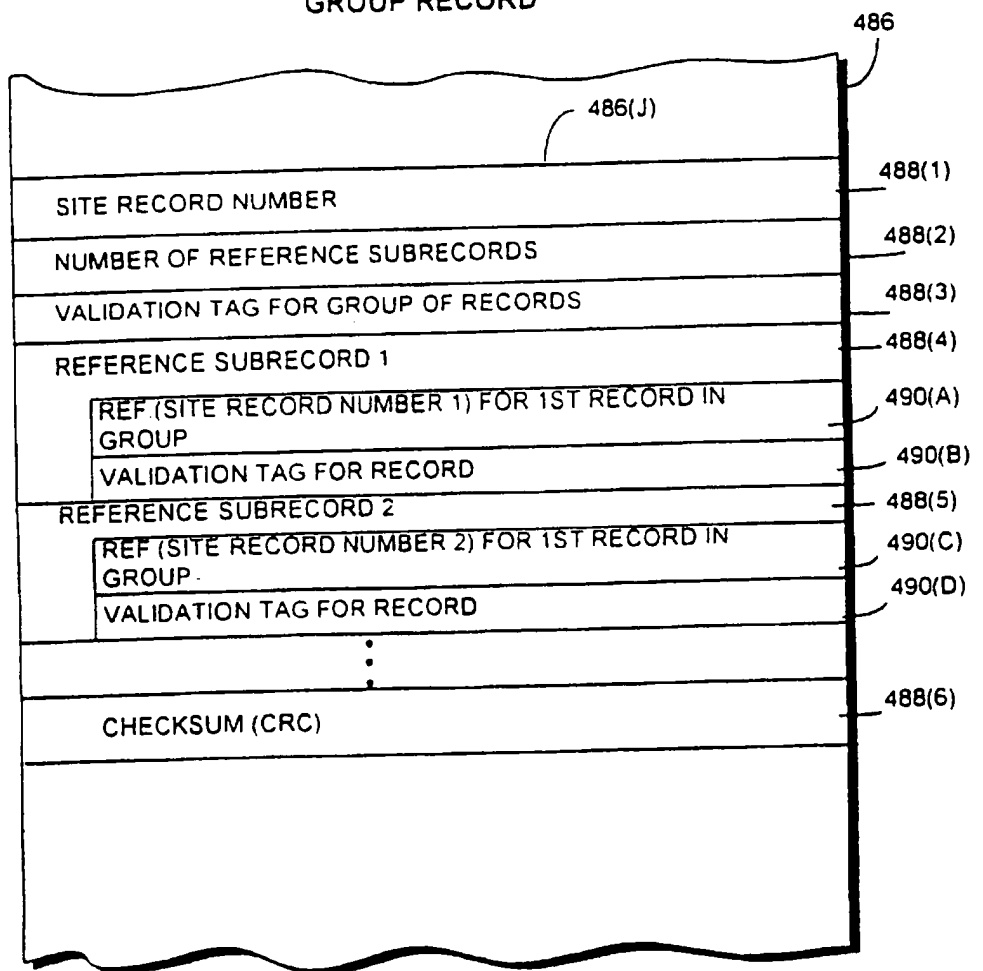


53/163

FIG. 34B

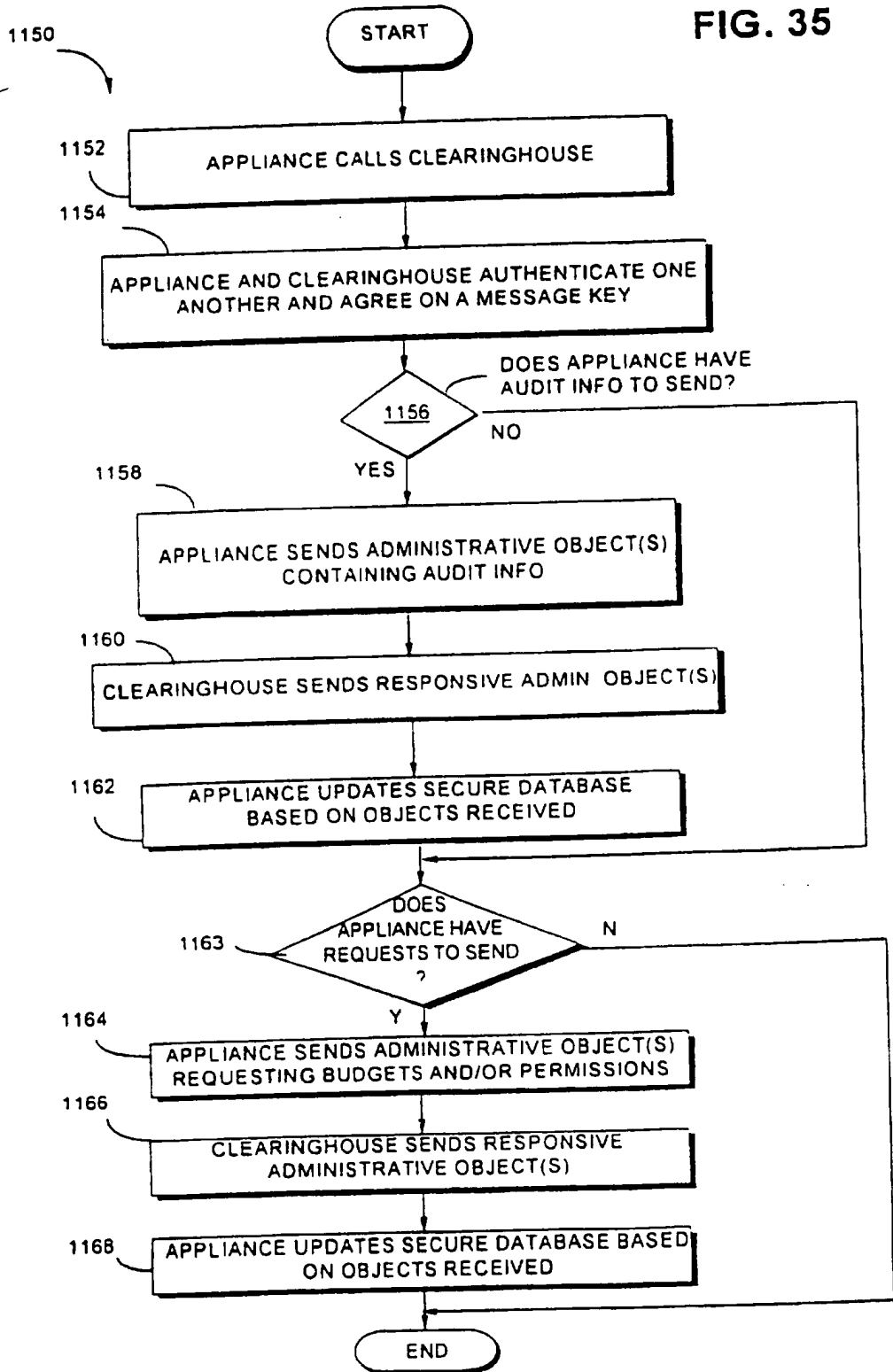
GROUP RECORD



SUBSTITUTE SHEET (RULE 26)

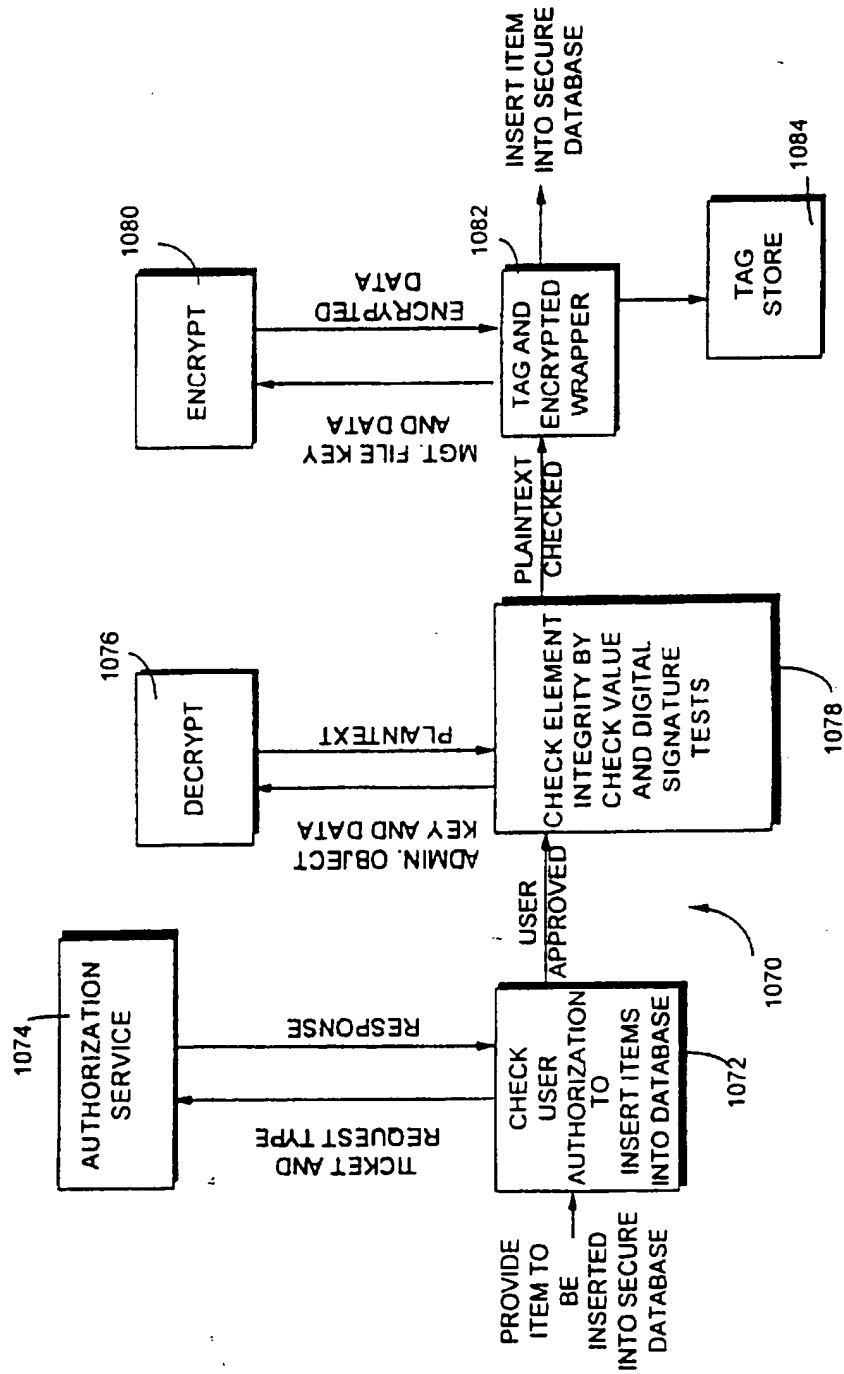
54/163

FIG. 35



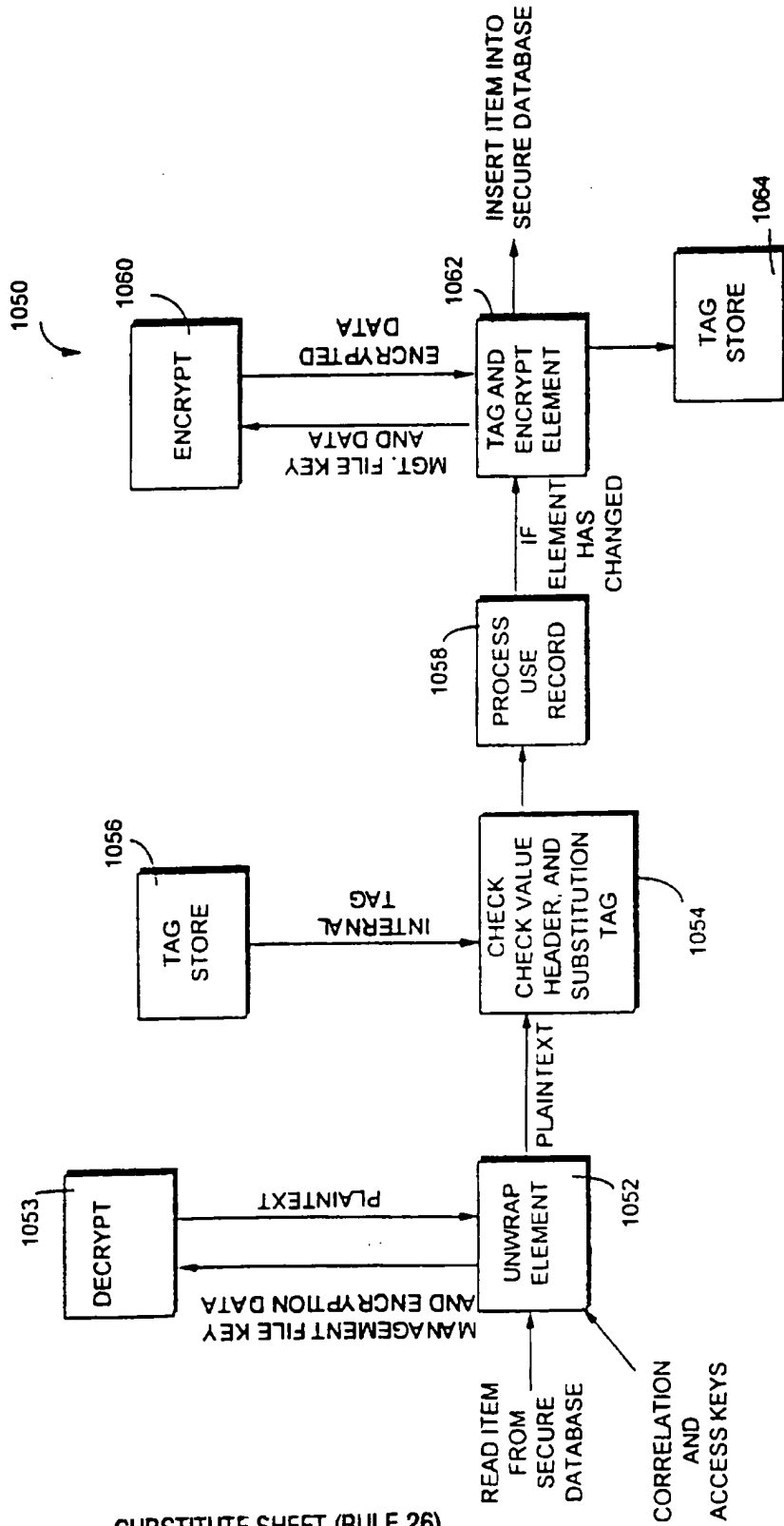
SUBSTITUTE SHEET (RULE 26)

FIG. 36



SUBSTITUTE SHEET (RULE 26)

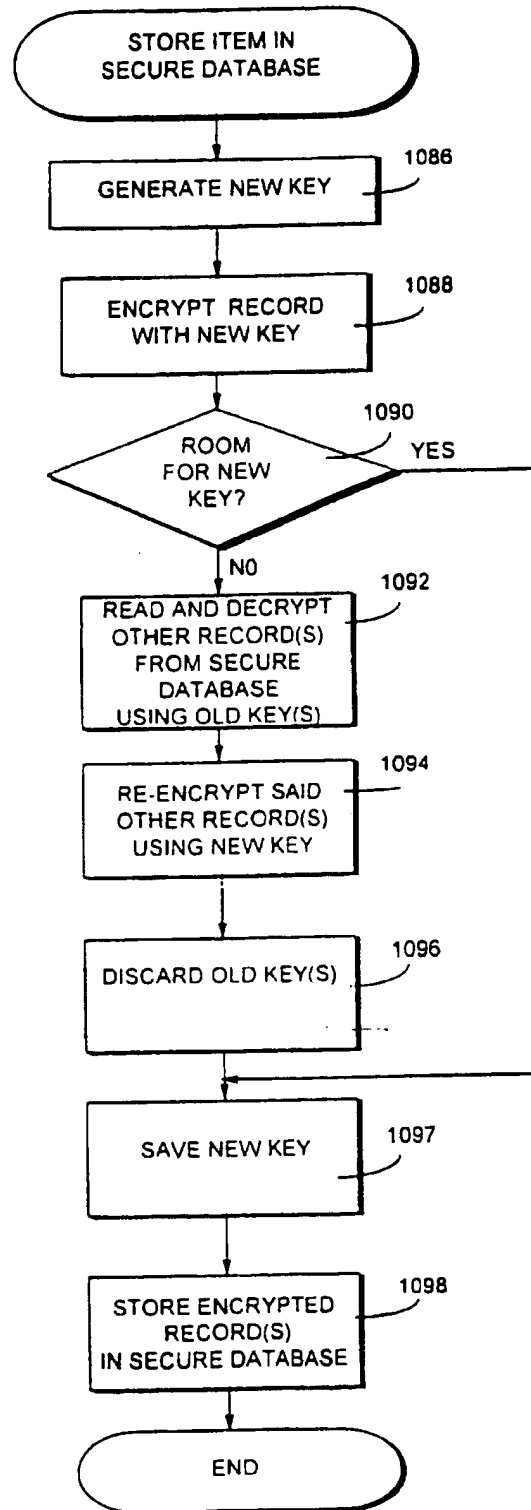
FIG. 37



SUBSTITUTE SHEET (RULE 26)

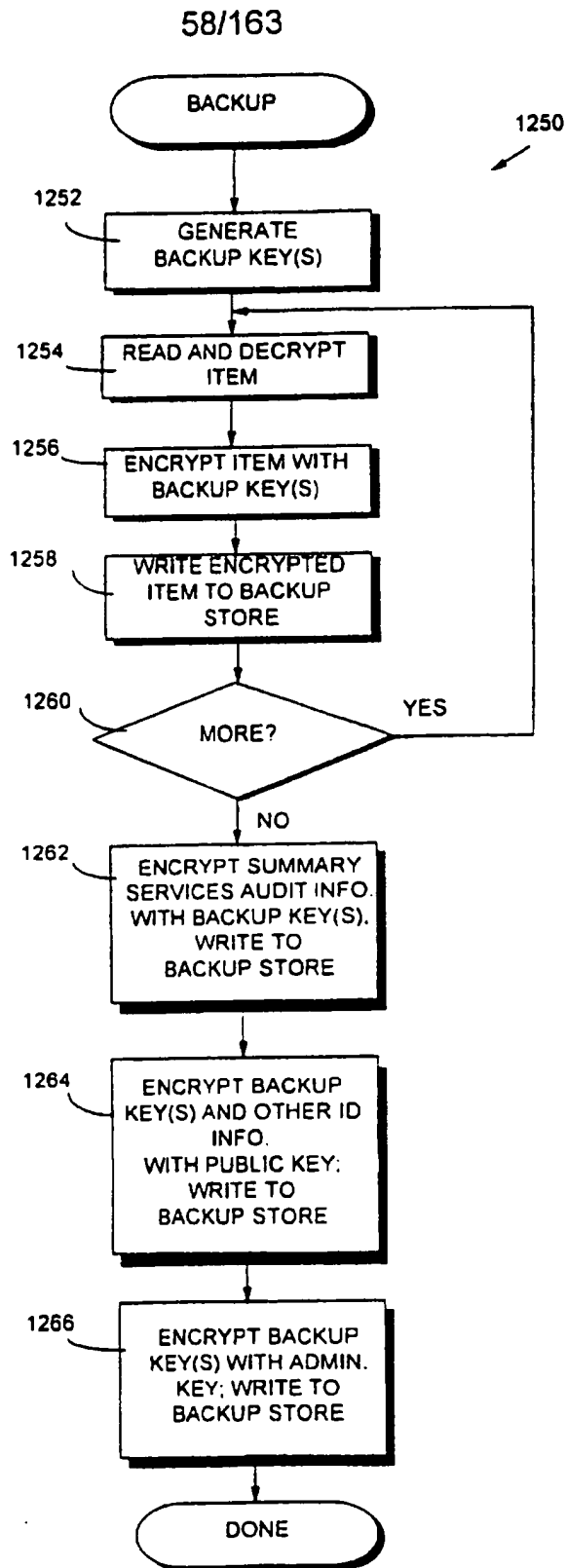
57/163

FIG. 38



SUBSTITUTE SHEET (RULE 26)

FIG. 39
BACKUP

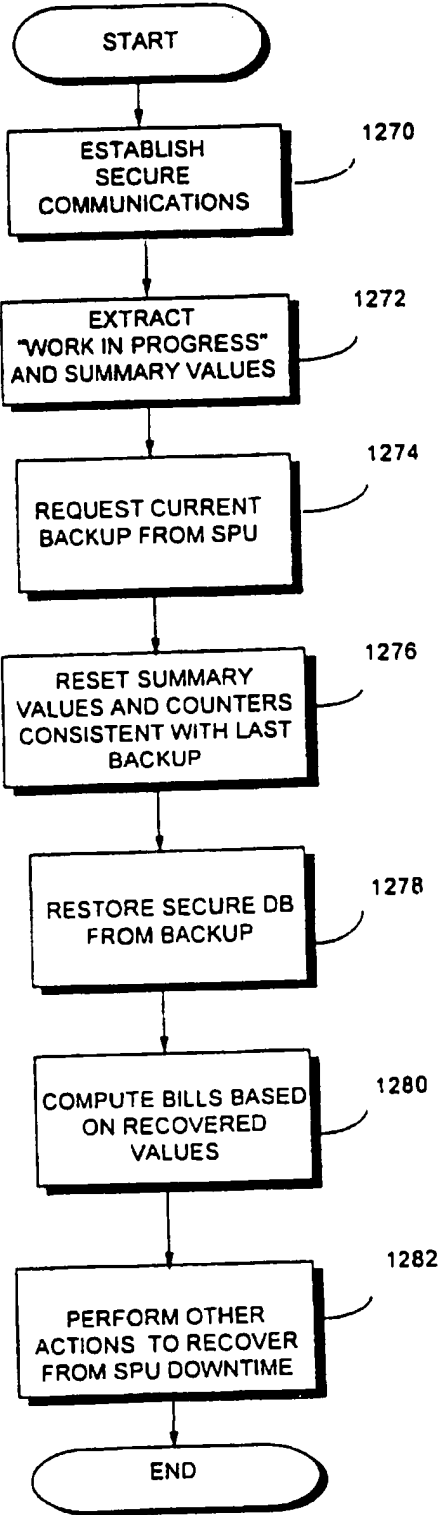


SUBSTITUTE SHEET (RULE 26)

59/163

FIG. 40
RECOVER SECURE DATABASE

1268
↘



SUBSTITUTE SHEET (RULE 26)

60/163

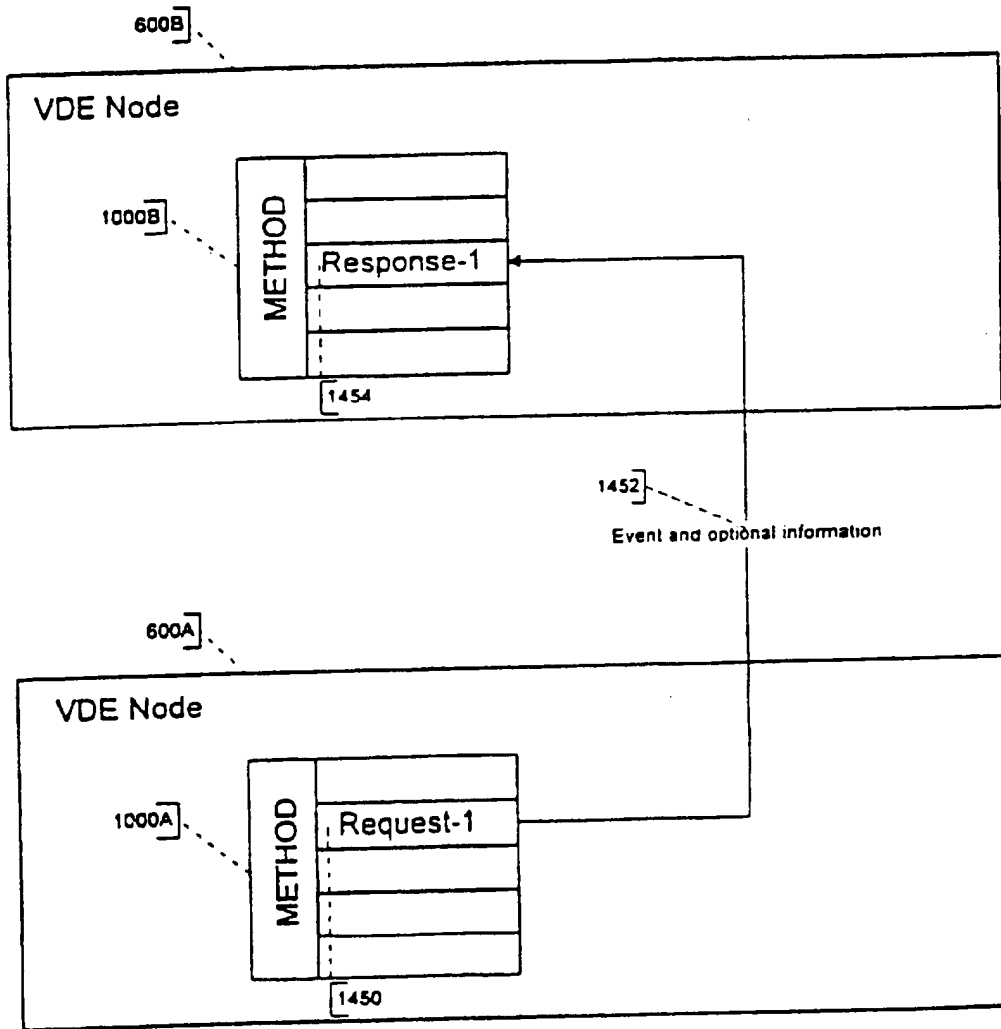


Figure 41a

SUBSTITUTE SHEET (RULE 26)

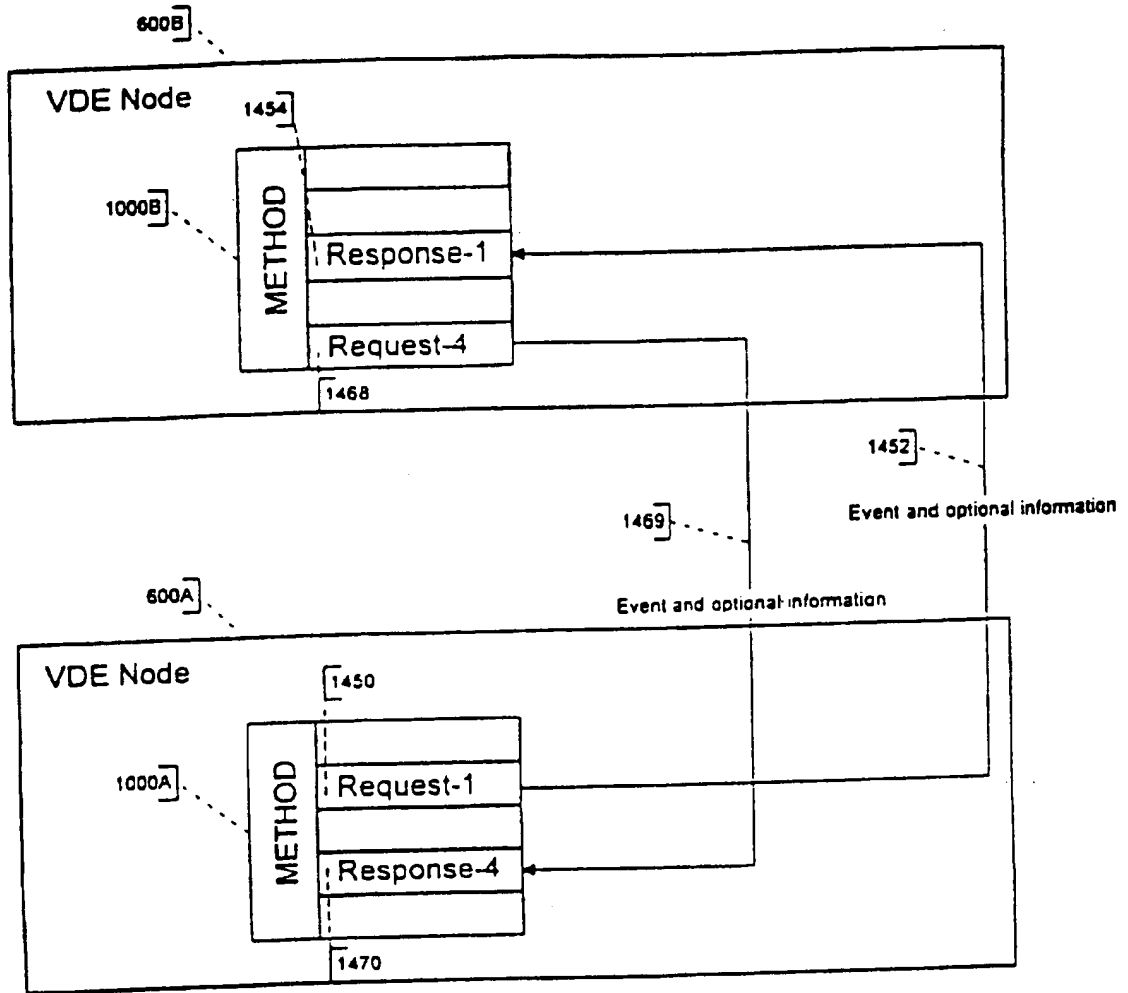


Figure 41b

SUBSTITUTE SHEET (RULE 26)

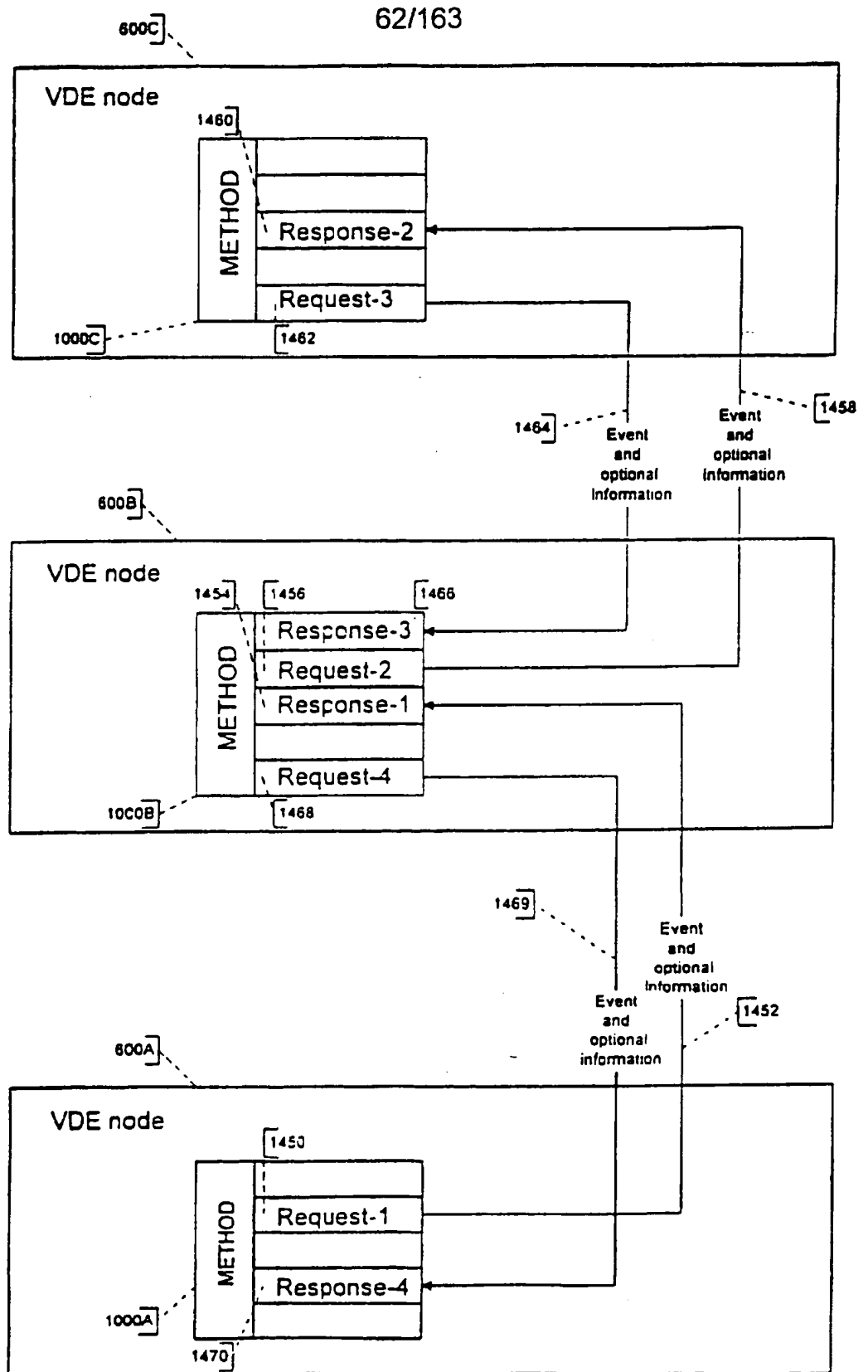


Figure 41c

SUBSTITUTE SHEET (RULE 26)

63/163

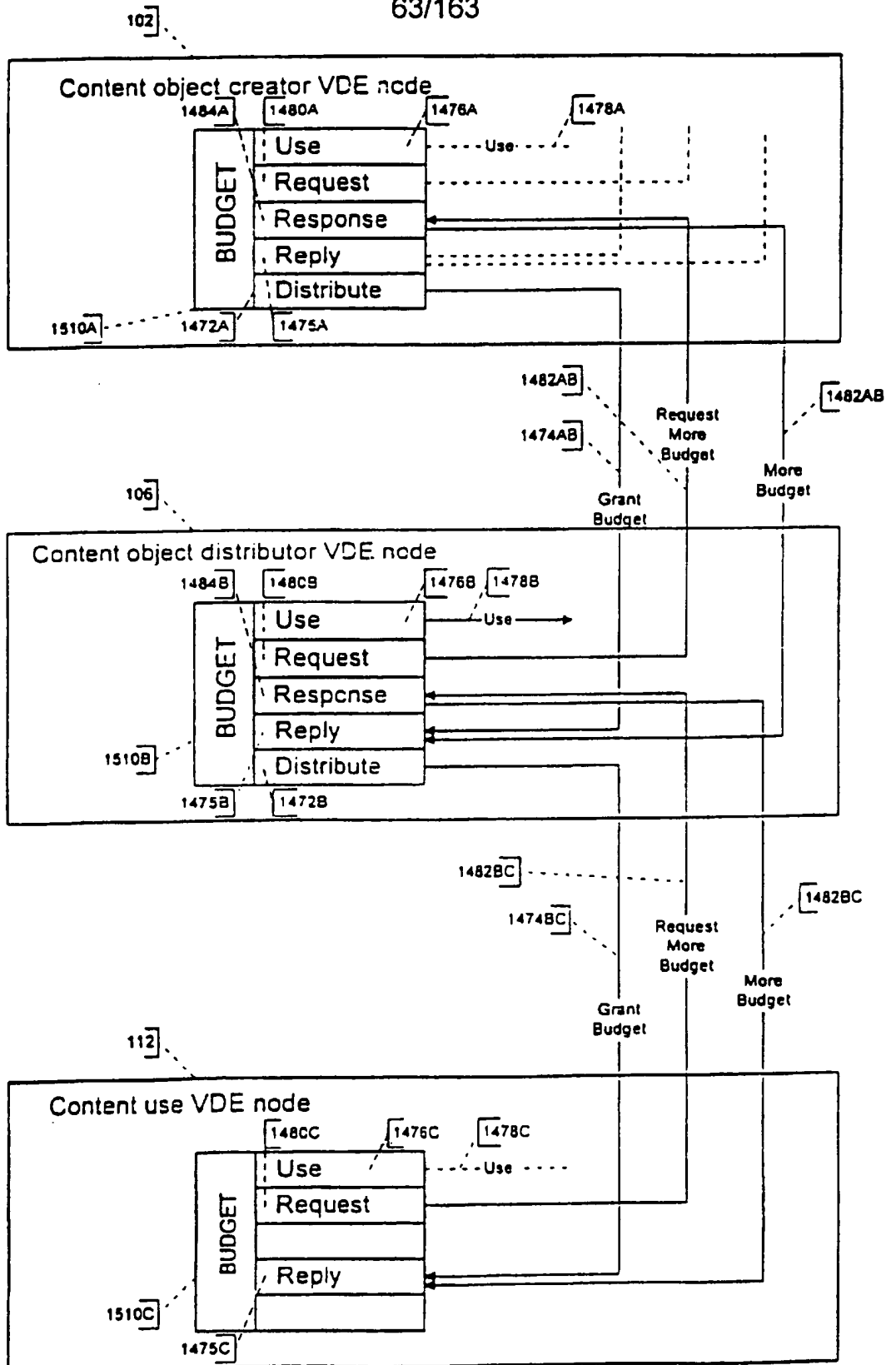


Figure 41d

SUBSTITUTE SHEET (RULE 26)

64/163

BUDGET Method Use Process Flow

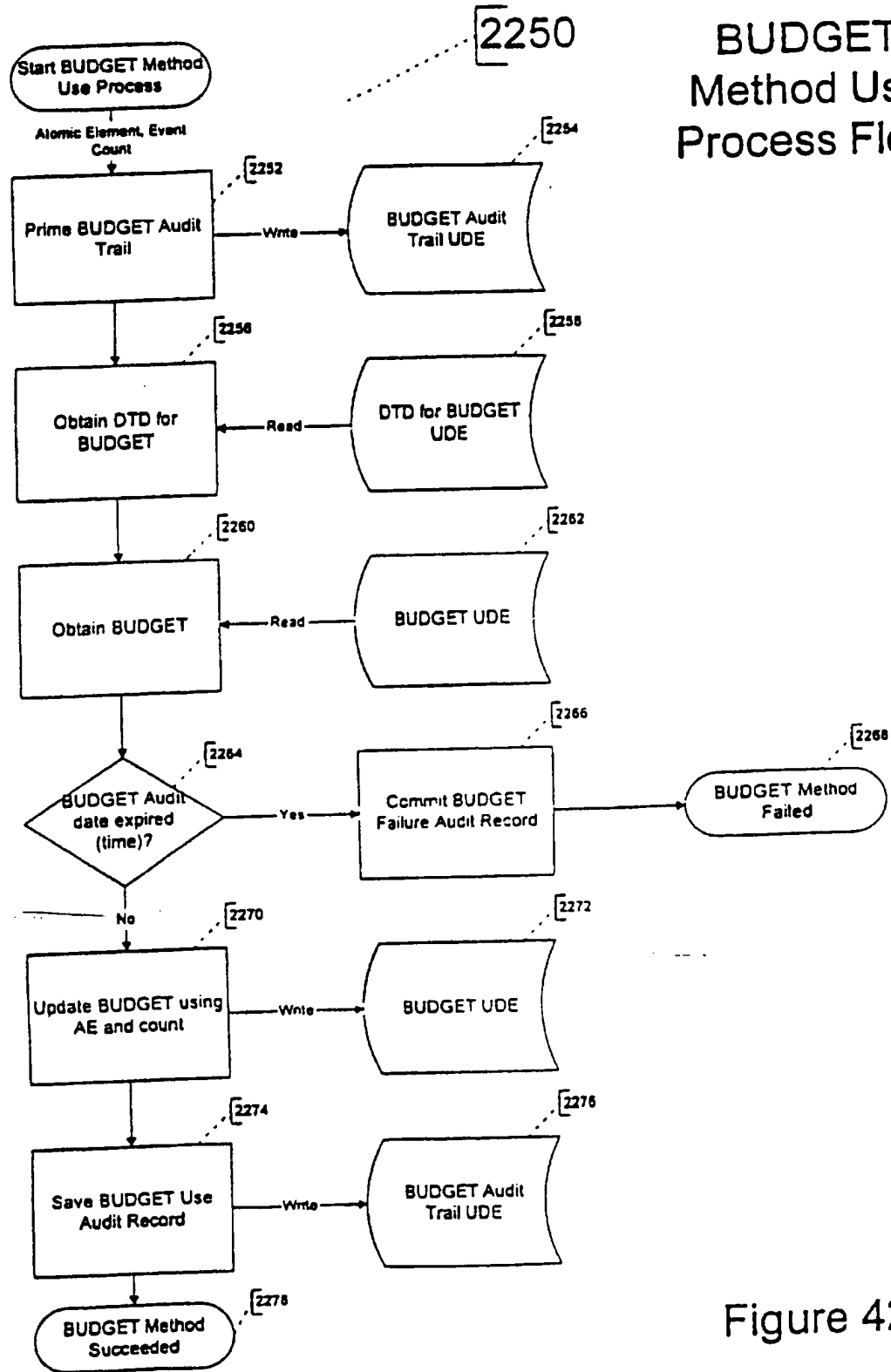


Figure 42a

SUBSTITUTE SHEET (RULE 26)

65/163

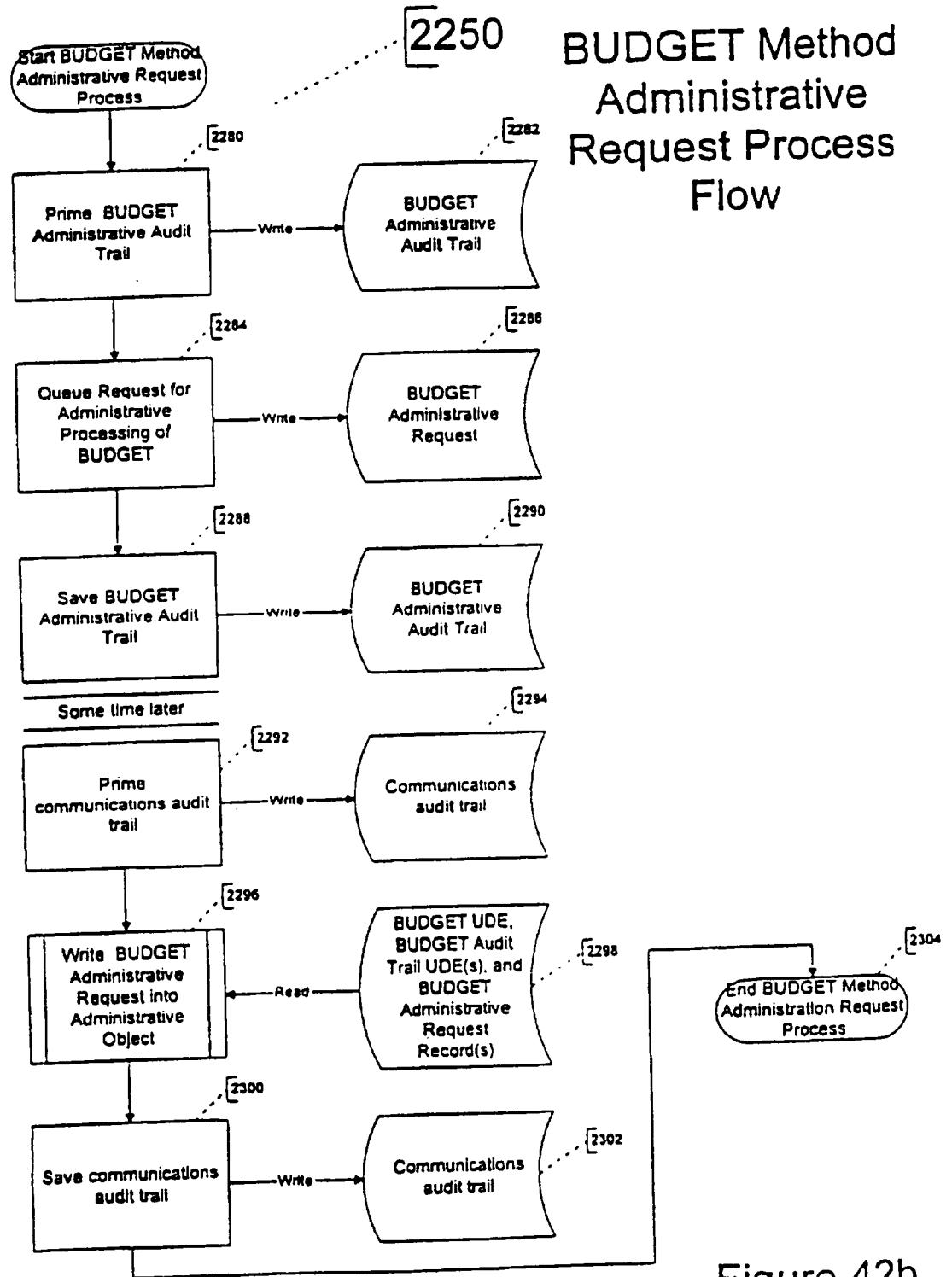


Figure 42b

66/163

BUDGET Method Administrative Response Process Flow

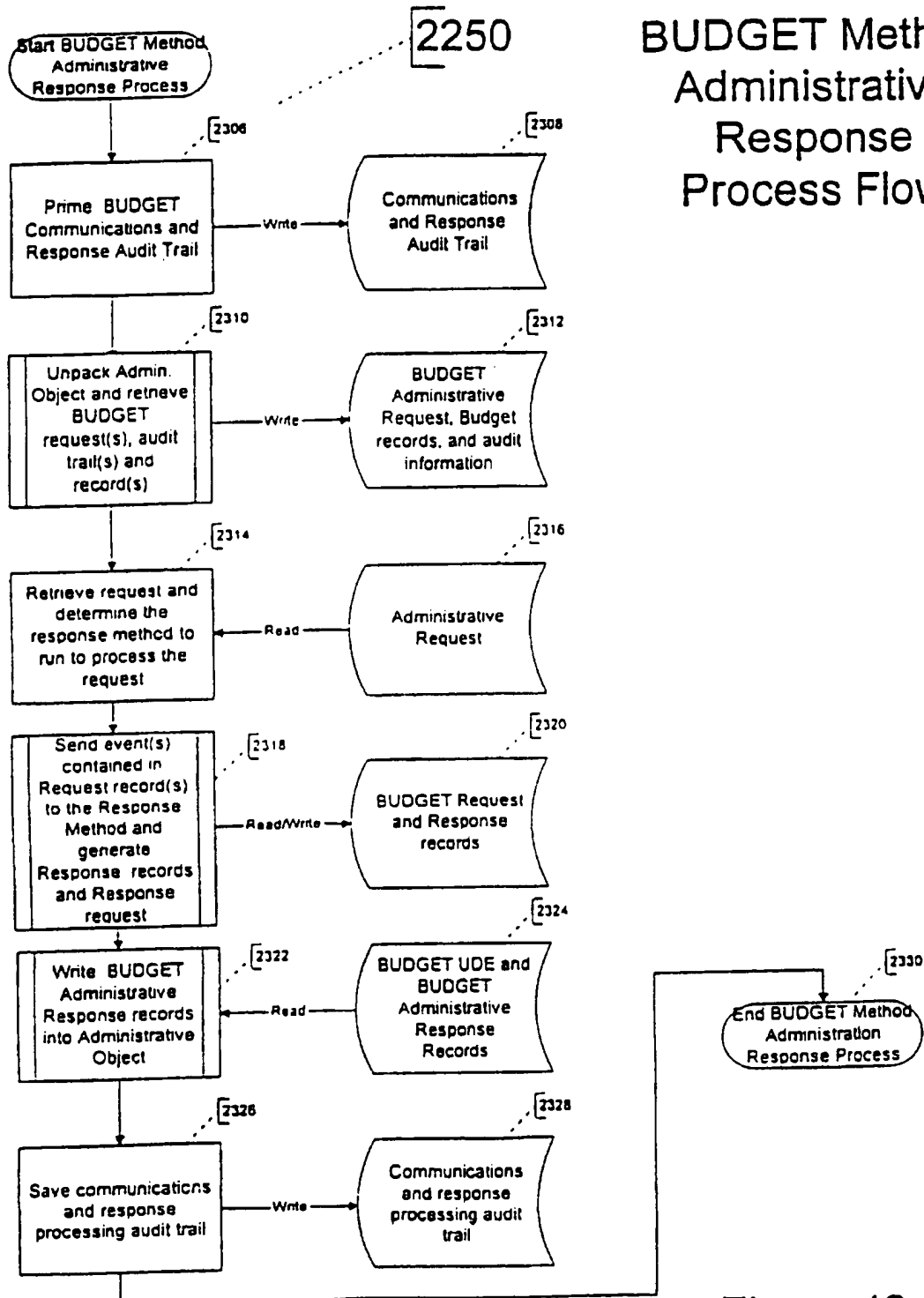


Figure 42c

SUBSTITUTE SHEET (RULE 26)

67/163

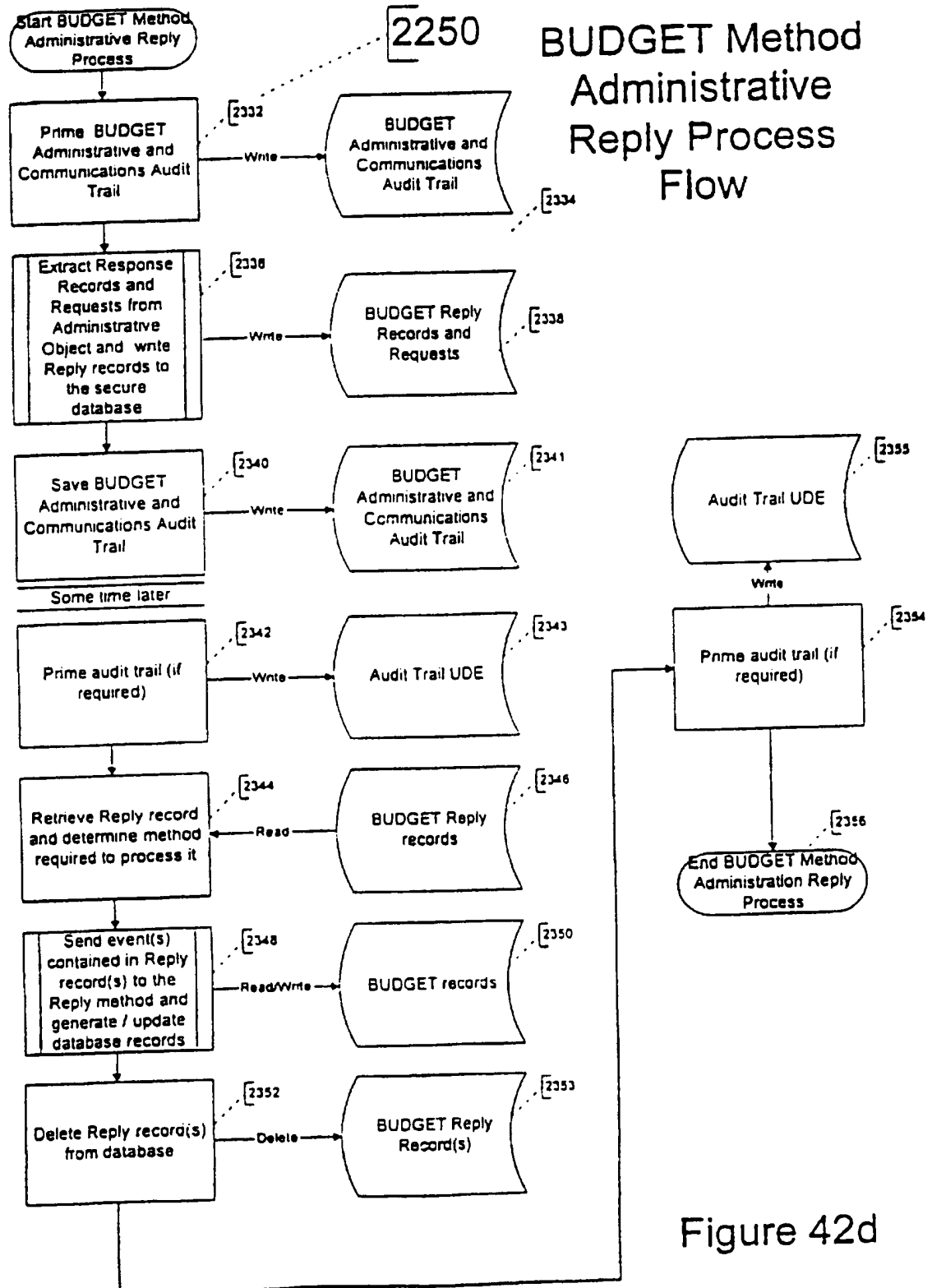
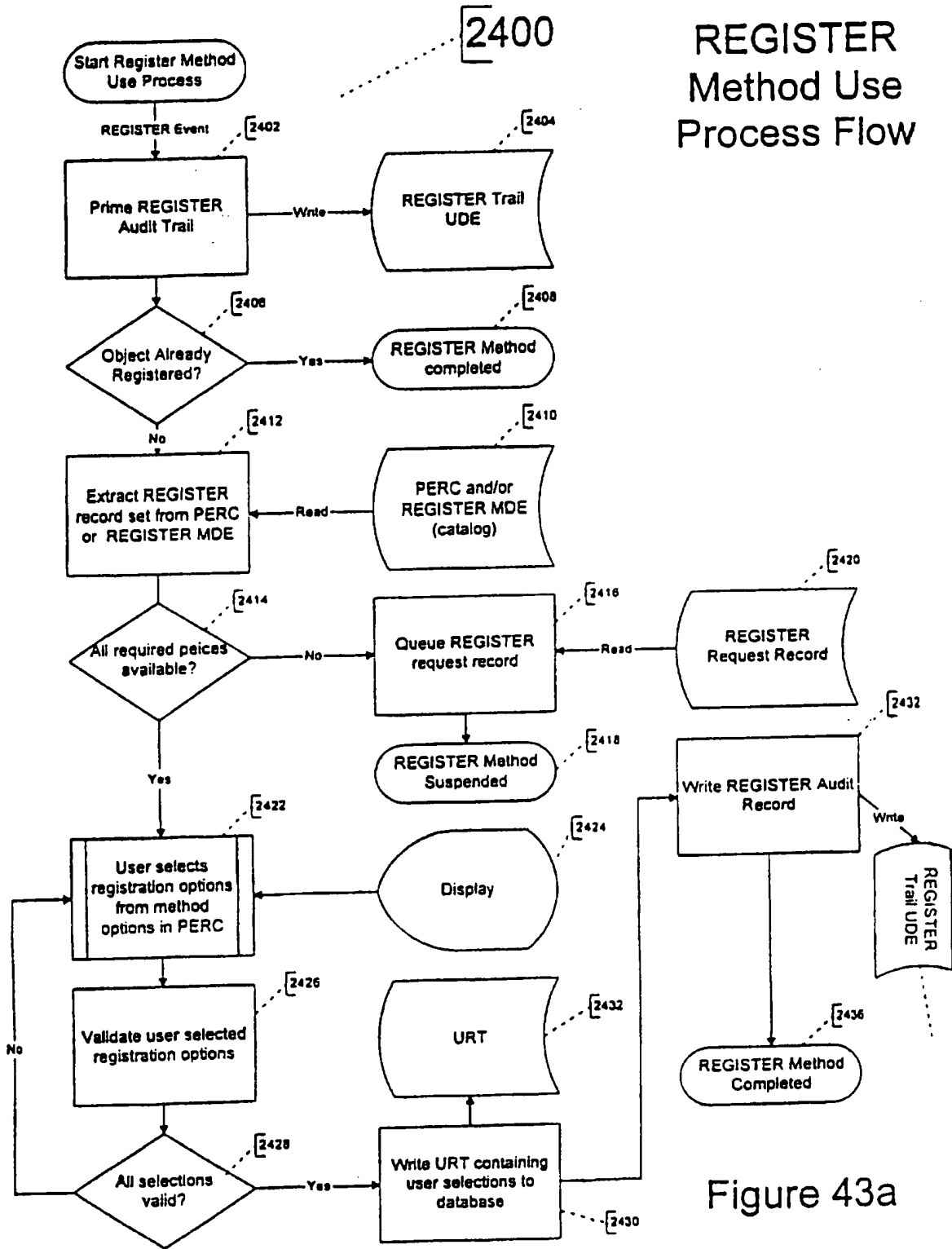


Figure 42d

68/163

REGISTER Method Use Process Flow



SUBSTITUTE SHEET (RULE 26)

Figure 43a

69/163

REGISTER Method Administrative Request Process Flow

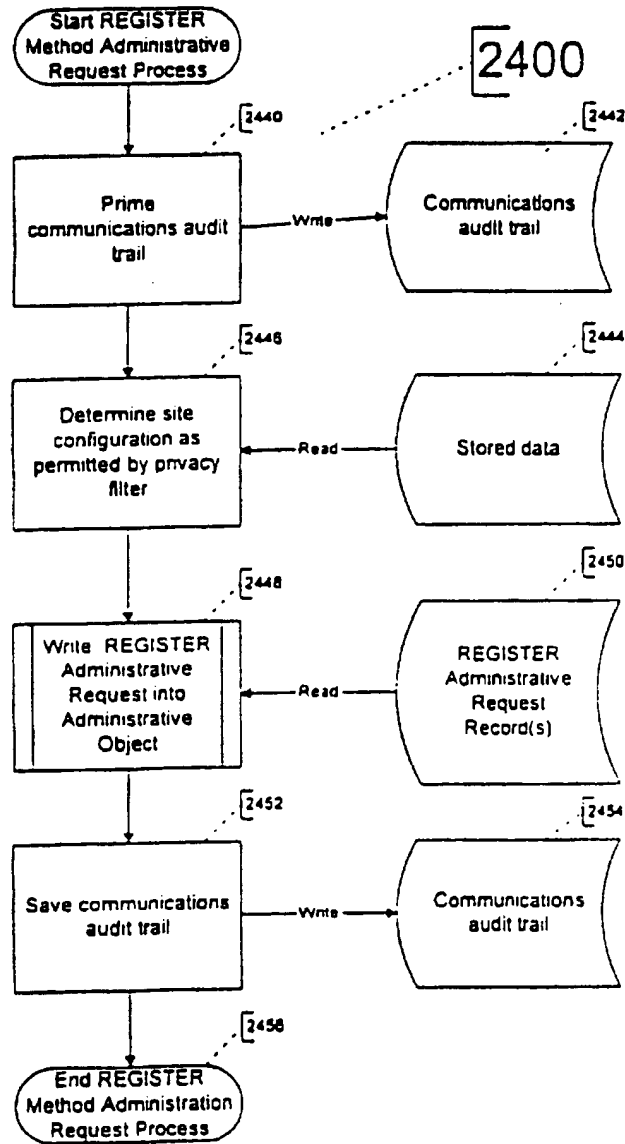


Figure 43b

70/163

REGISTER Method Administrative Response Process Flow

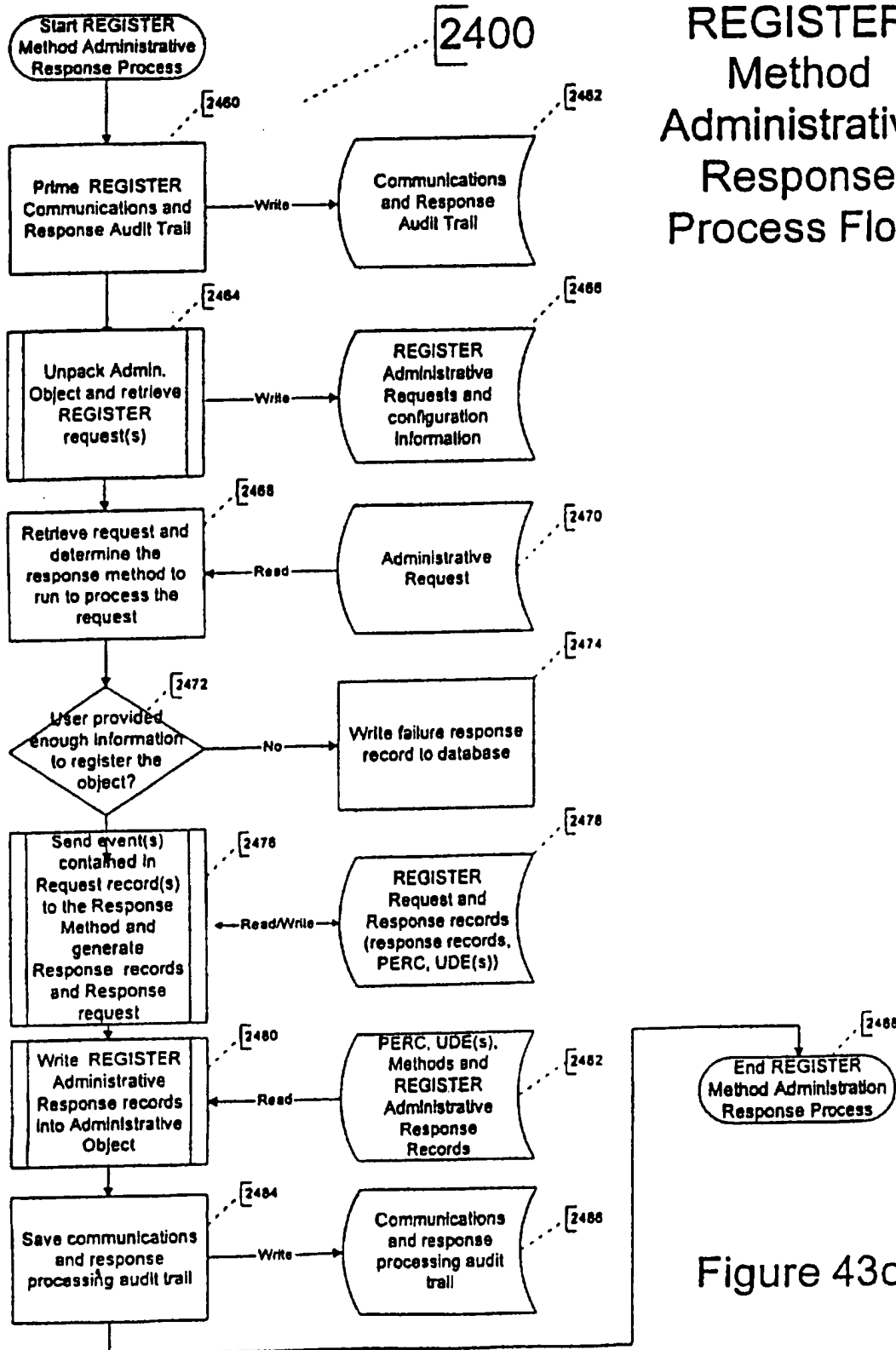


Figure 43c

71/163

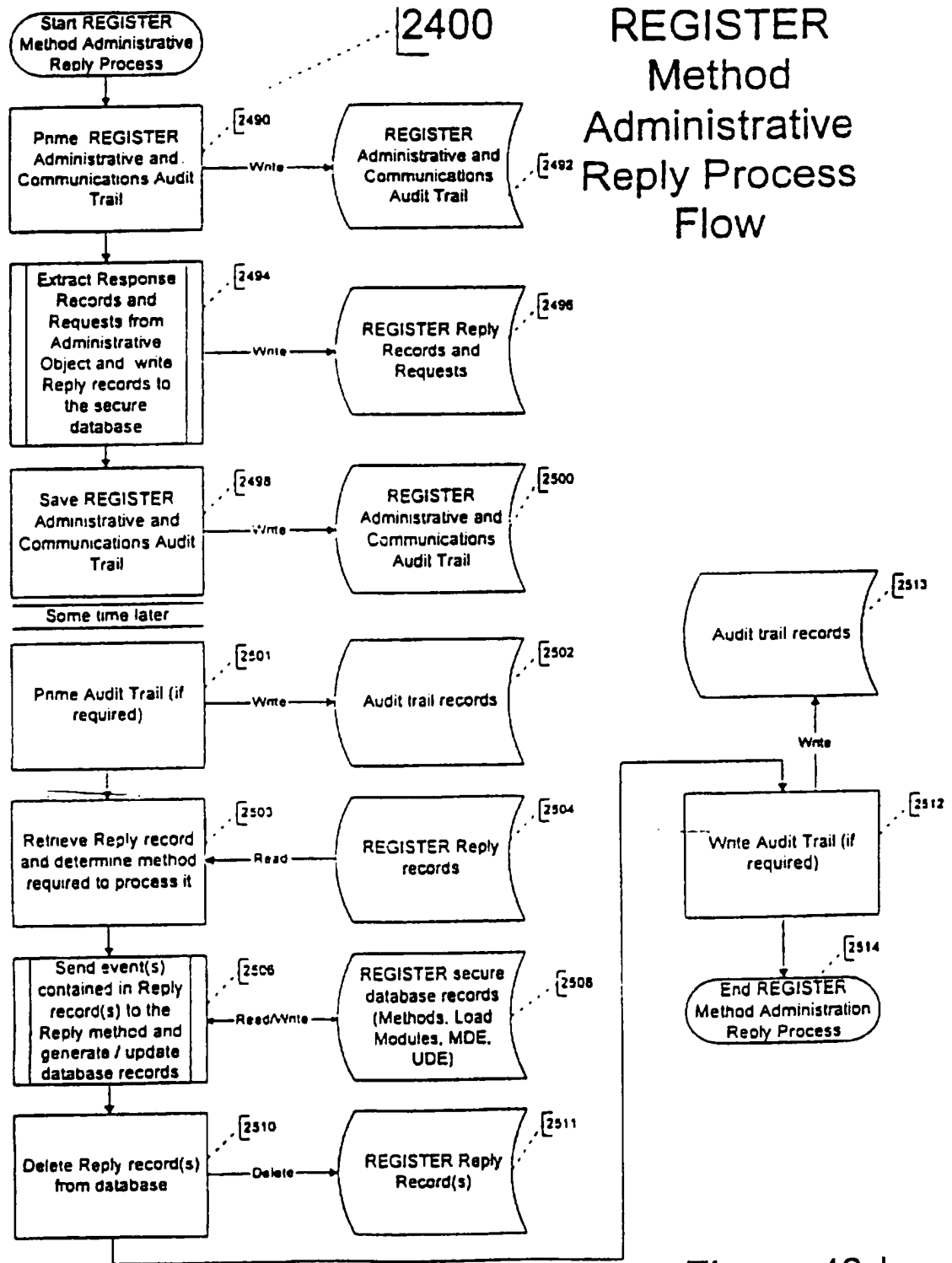


Figure 43d

SUBSTITUTE SHEET (RULE 26)

72/163

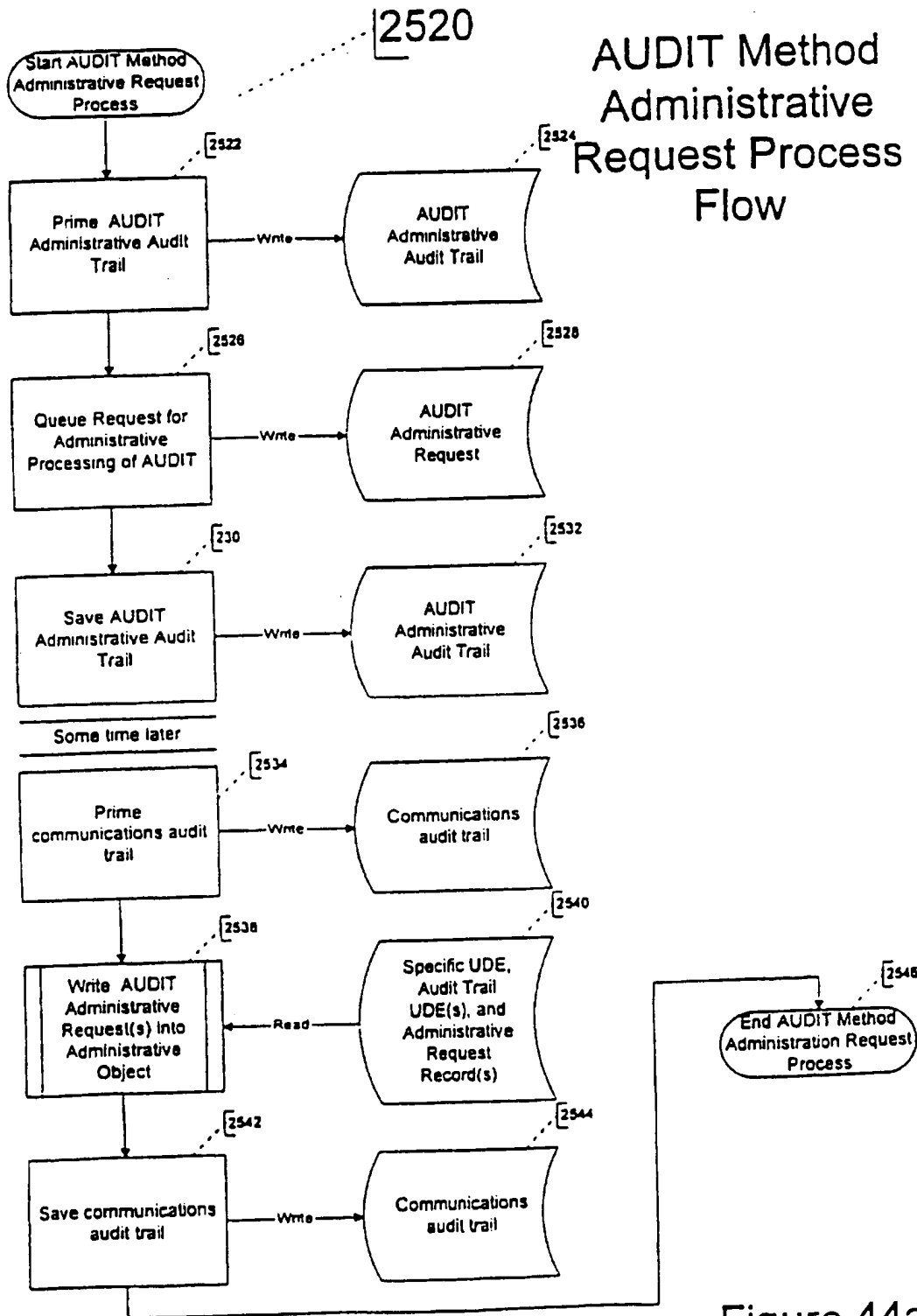


Figure 44a

SUBSTITUTE SHEET (RULE 26)

73/163

AUDIT Method Administrative Response Process Flow

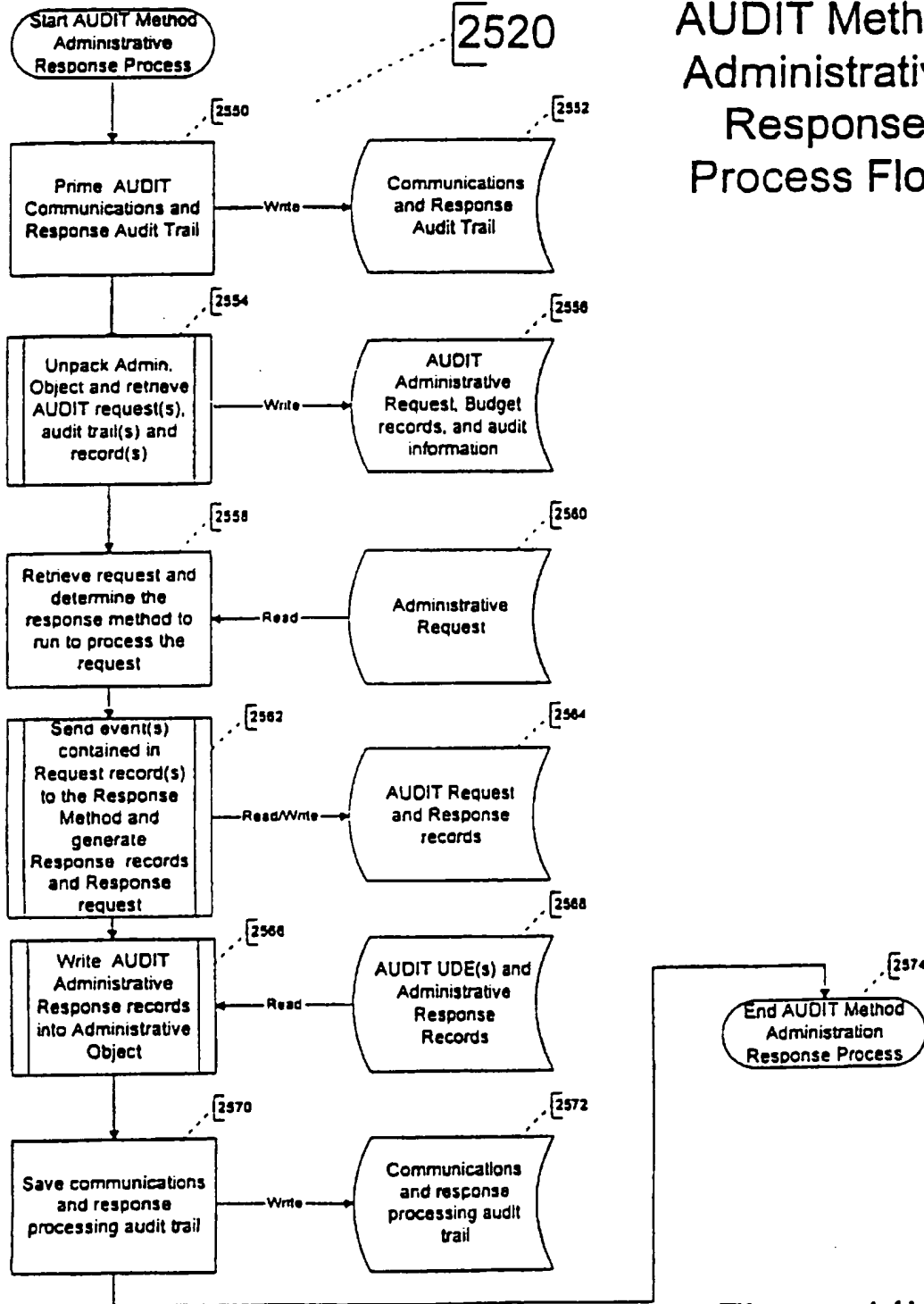
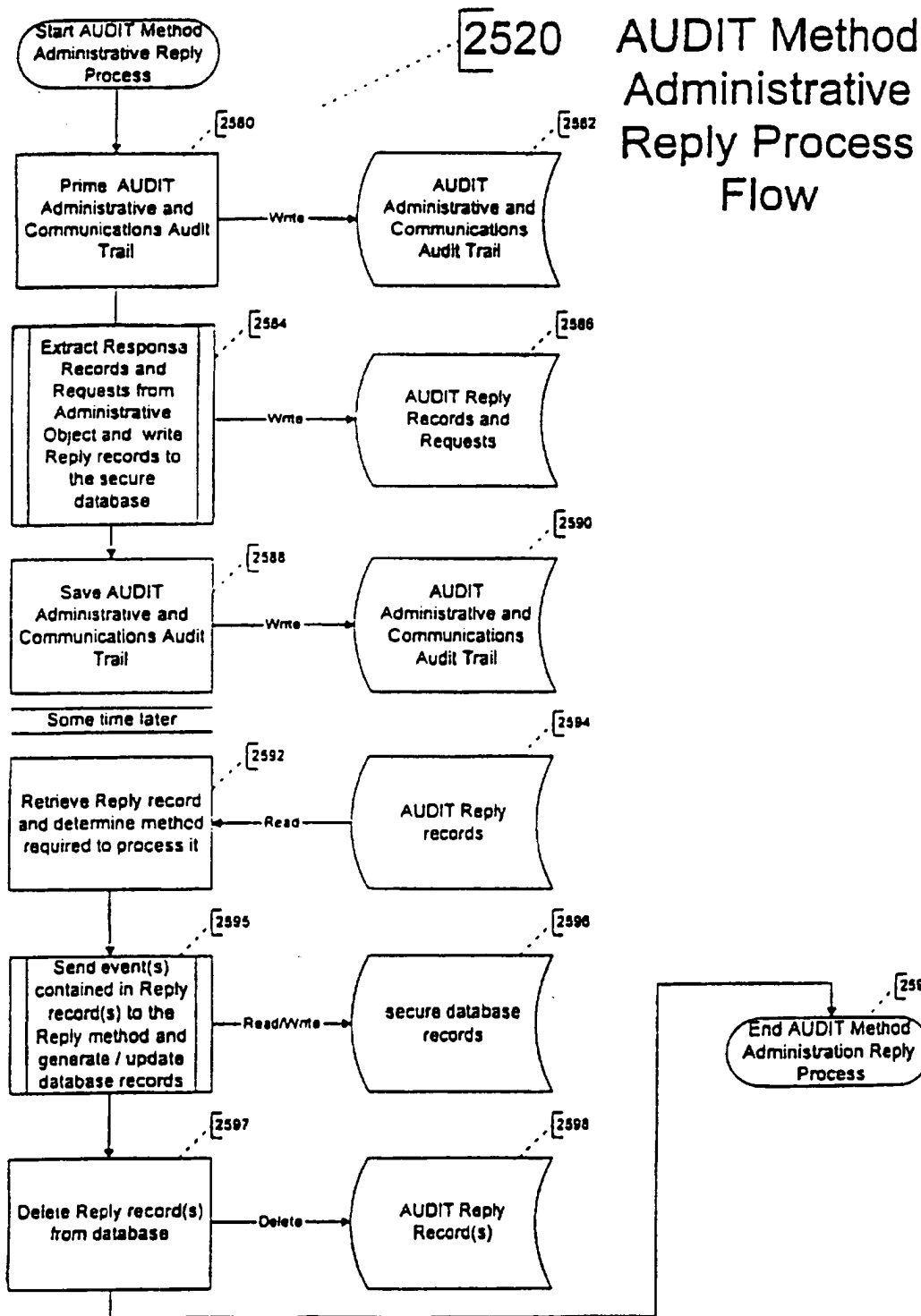


Figure 44b

74/163



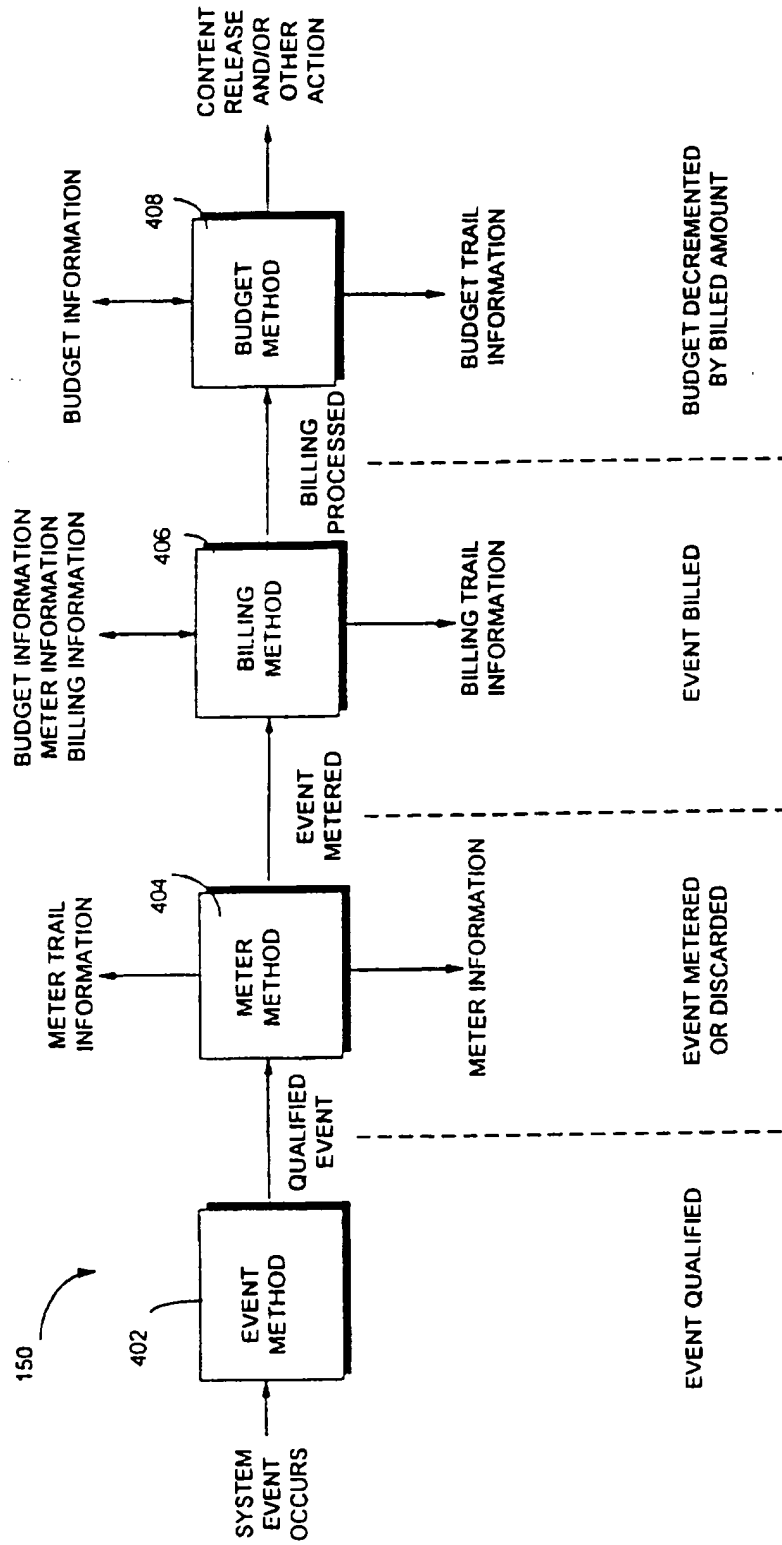
AUDIT Method Administrative Reply Process Flow

Figure 44c

SUBSTITUTE SHEET (RULE 28)

75/163

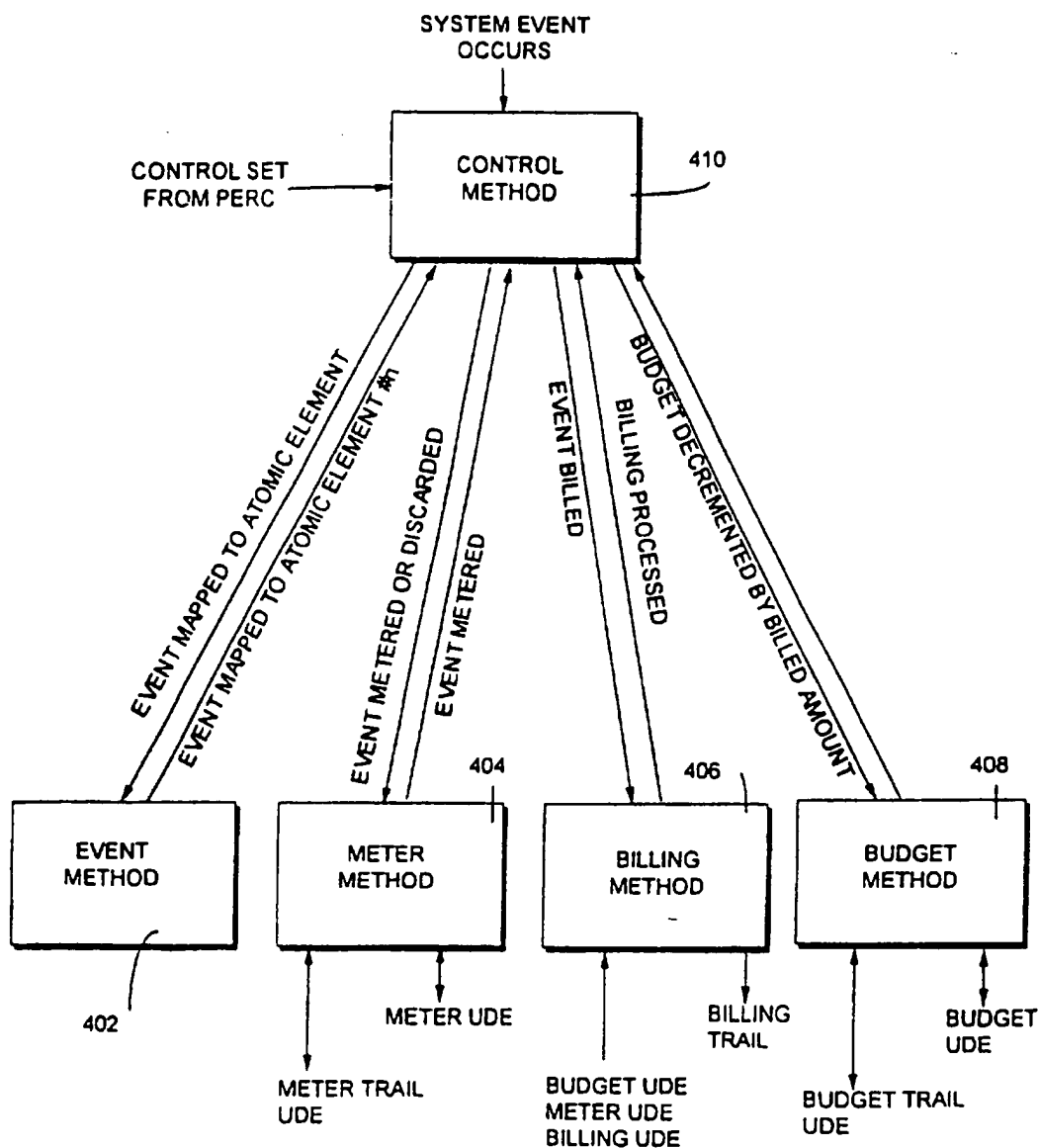
FIG. 45



SUBSTITUTE SHEET (RULE 26)

76/163

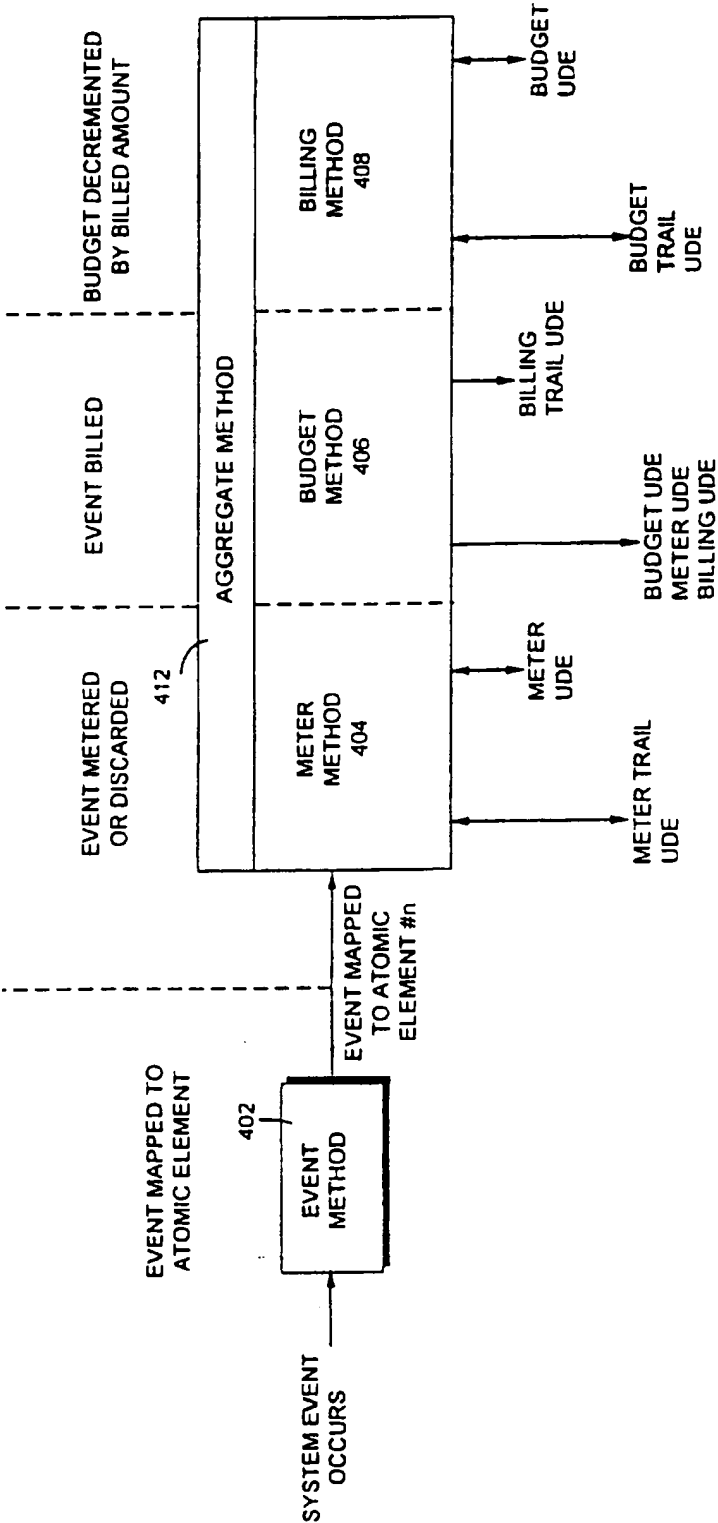
FIG. 46



SUBSTITUTE SHEET (RULE 26)

77/163

FIG. 47



78/163

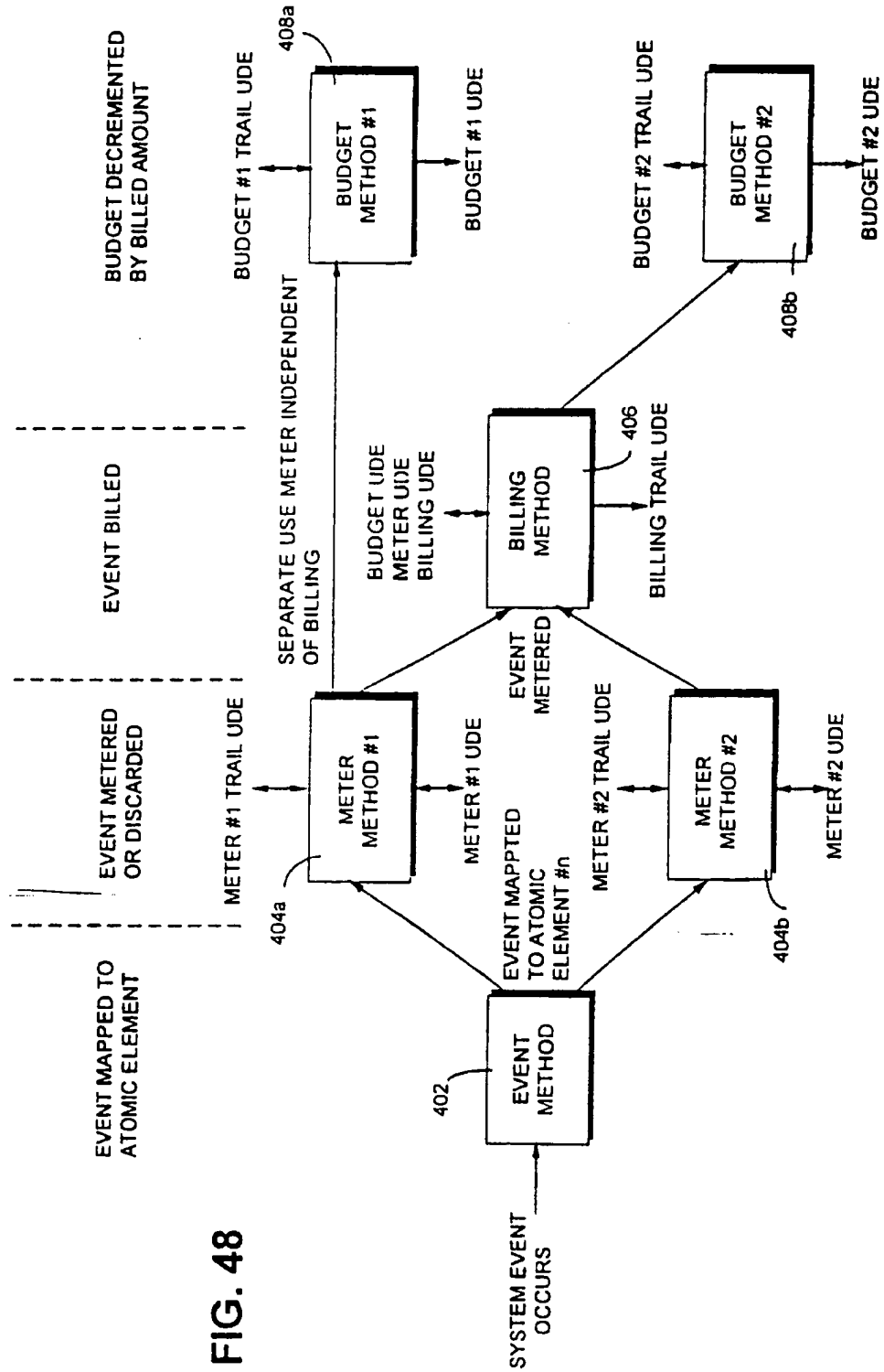


FIG. 48

SUBSTITUTE SHEET (RULE 26)

79/163

OPEN Method Use Process Flow

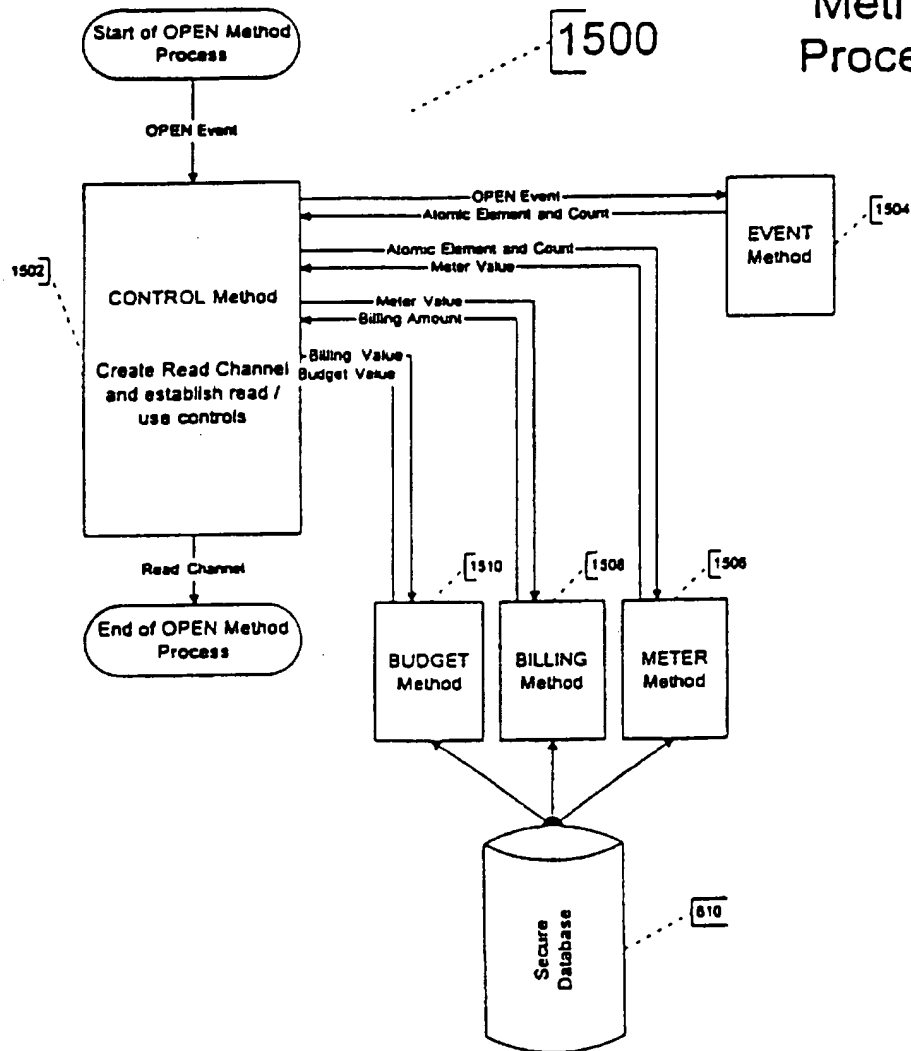


Figure 49

80/163

1500

1502

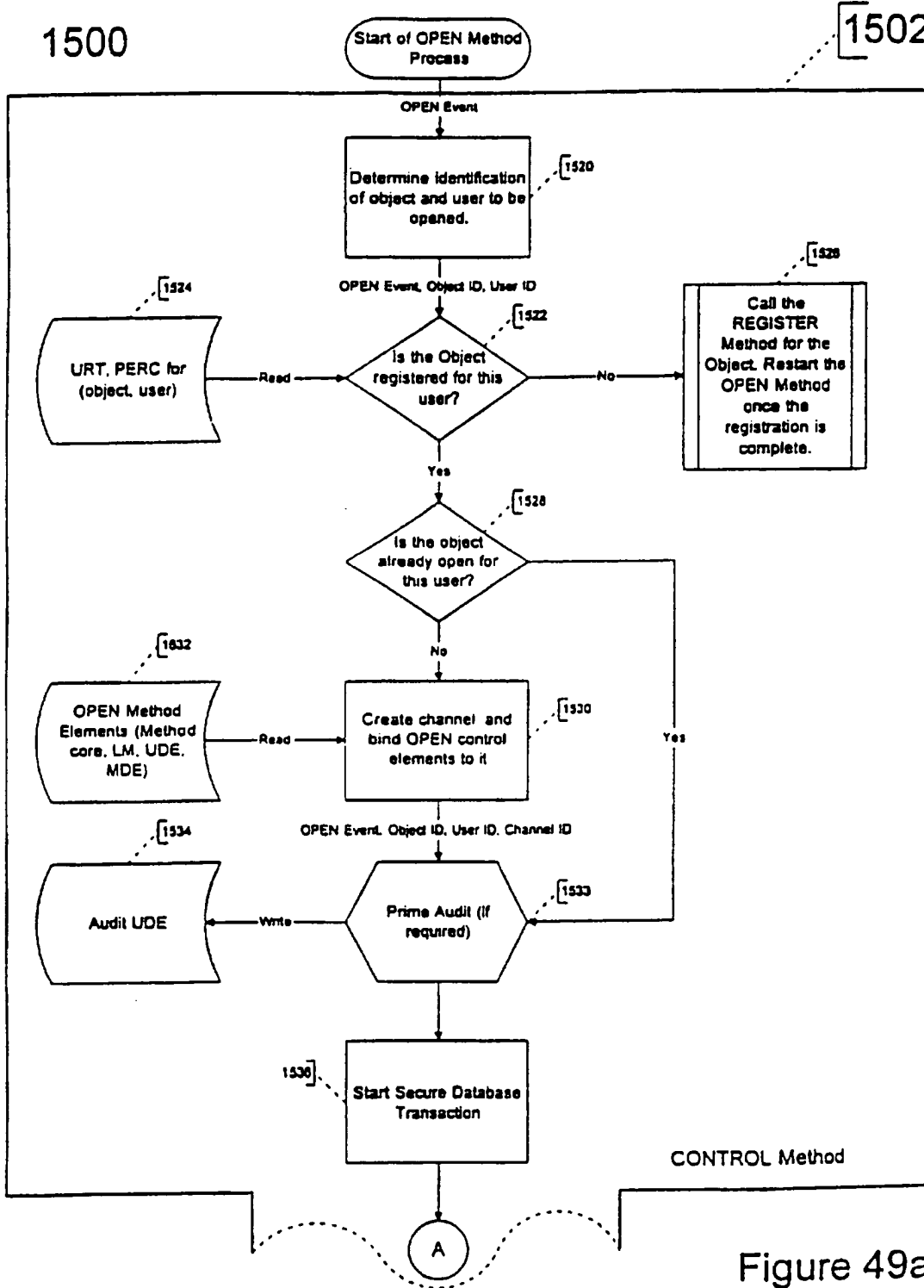


Figure 49a

81/163

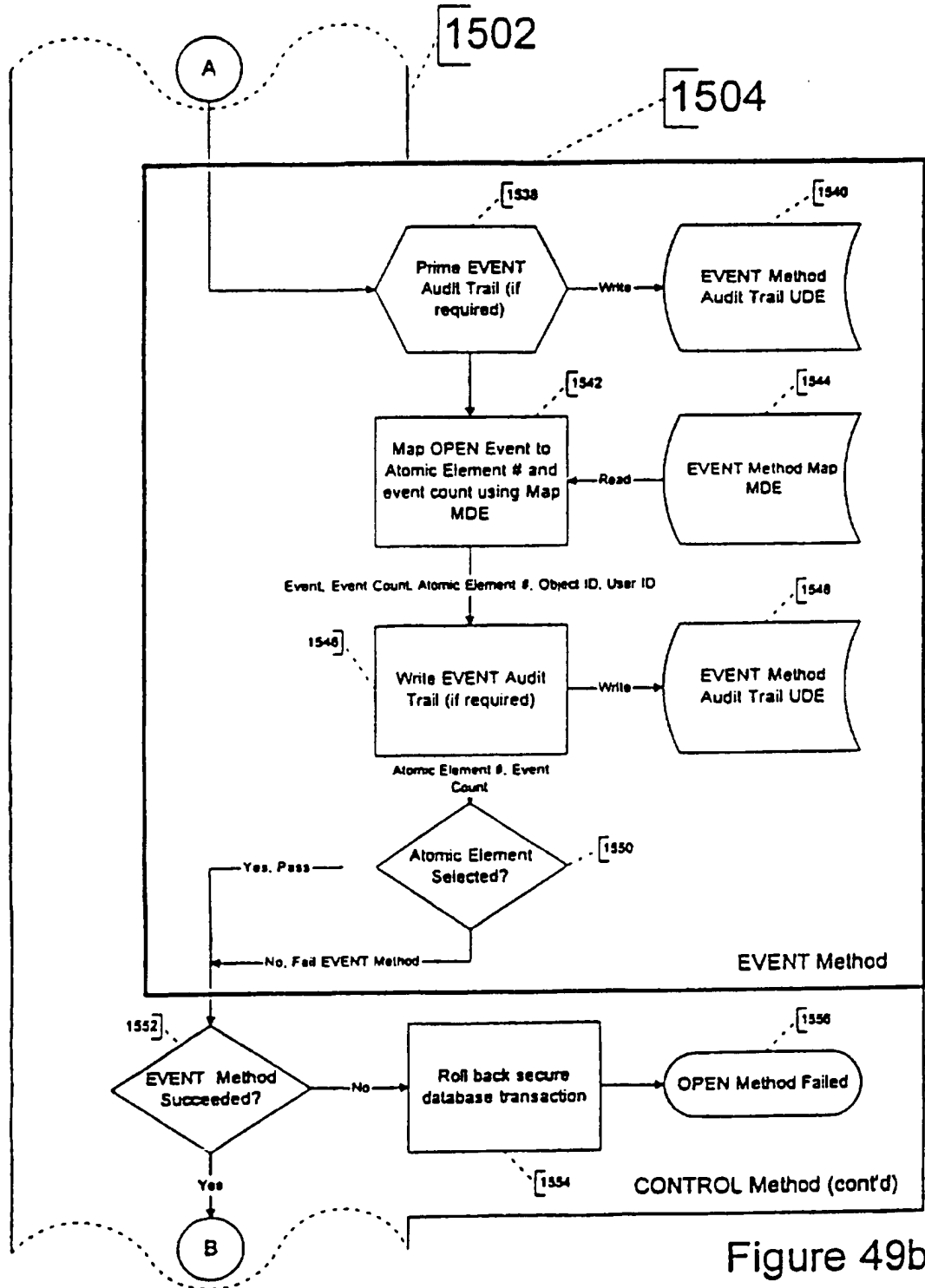


Figure 49b

82/163

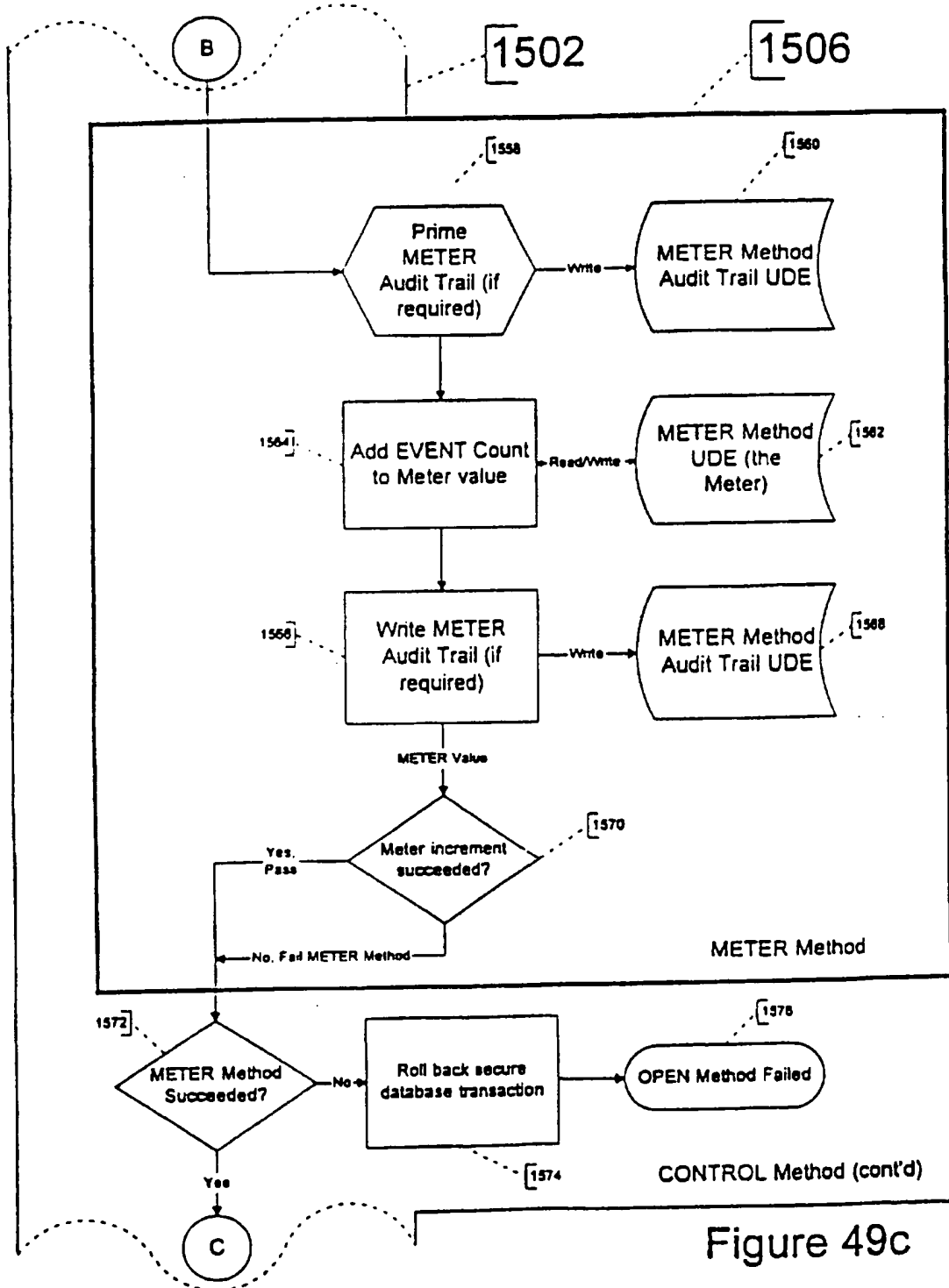


Figure 49c

SUBSTITUTE SHEET (RULE 26)

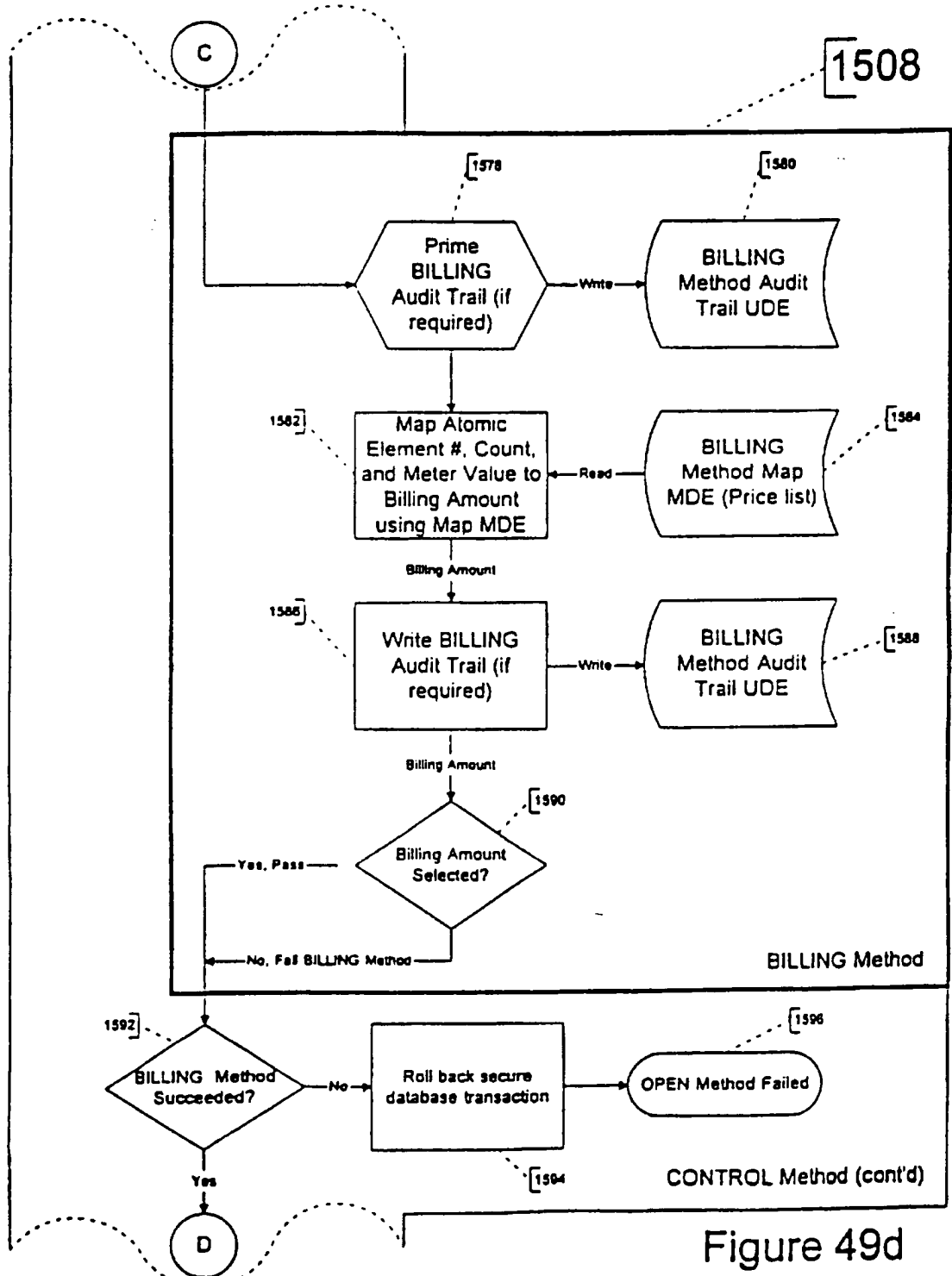


Figure 49d

SUBSTITUTE SHEET (RULE 26)

84/163

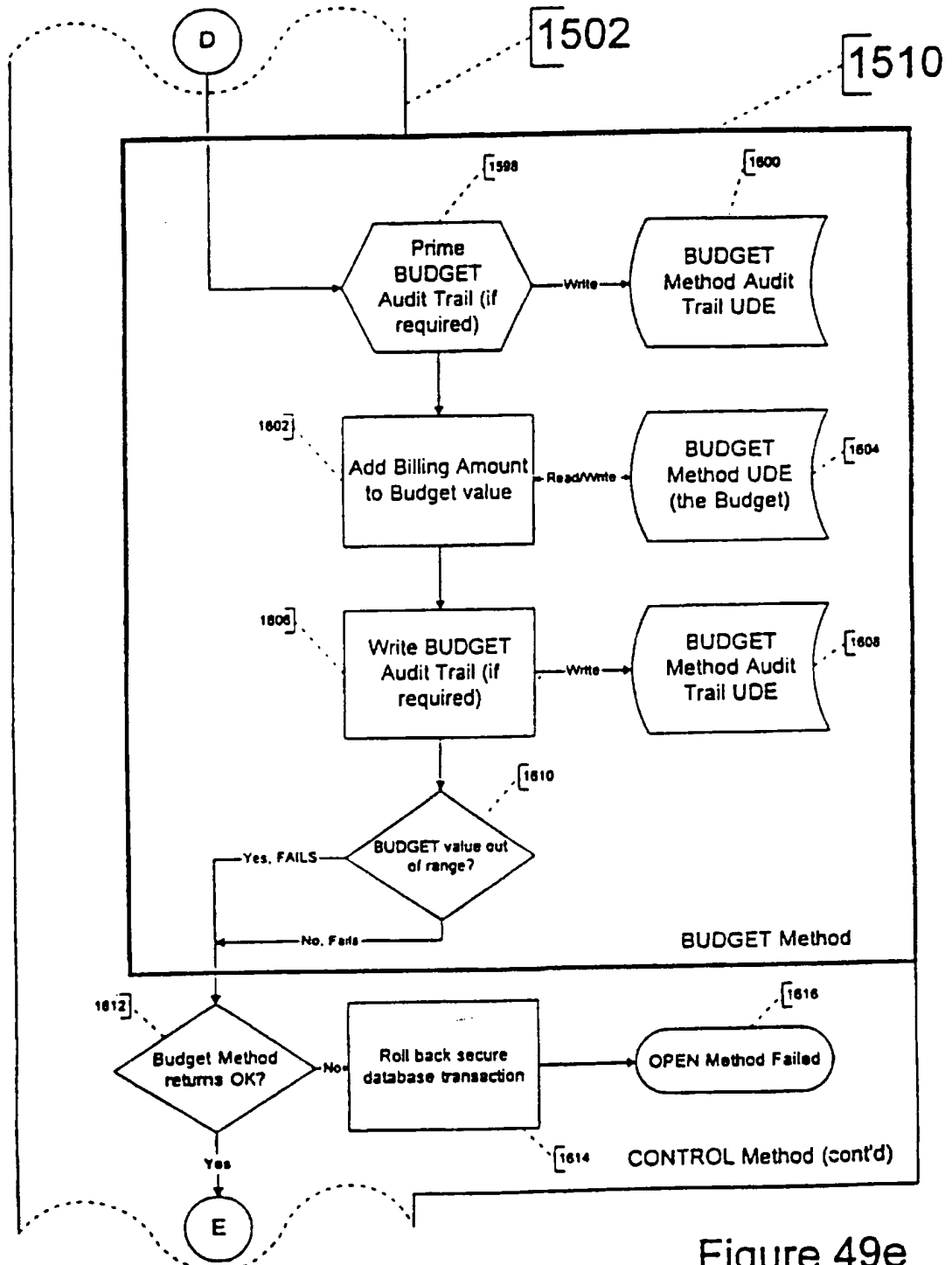


Figure 49e

SUBSTITUTE SHEET (RULE 26)

85/163

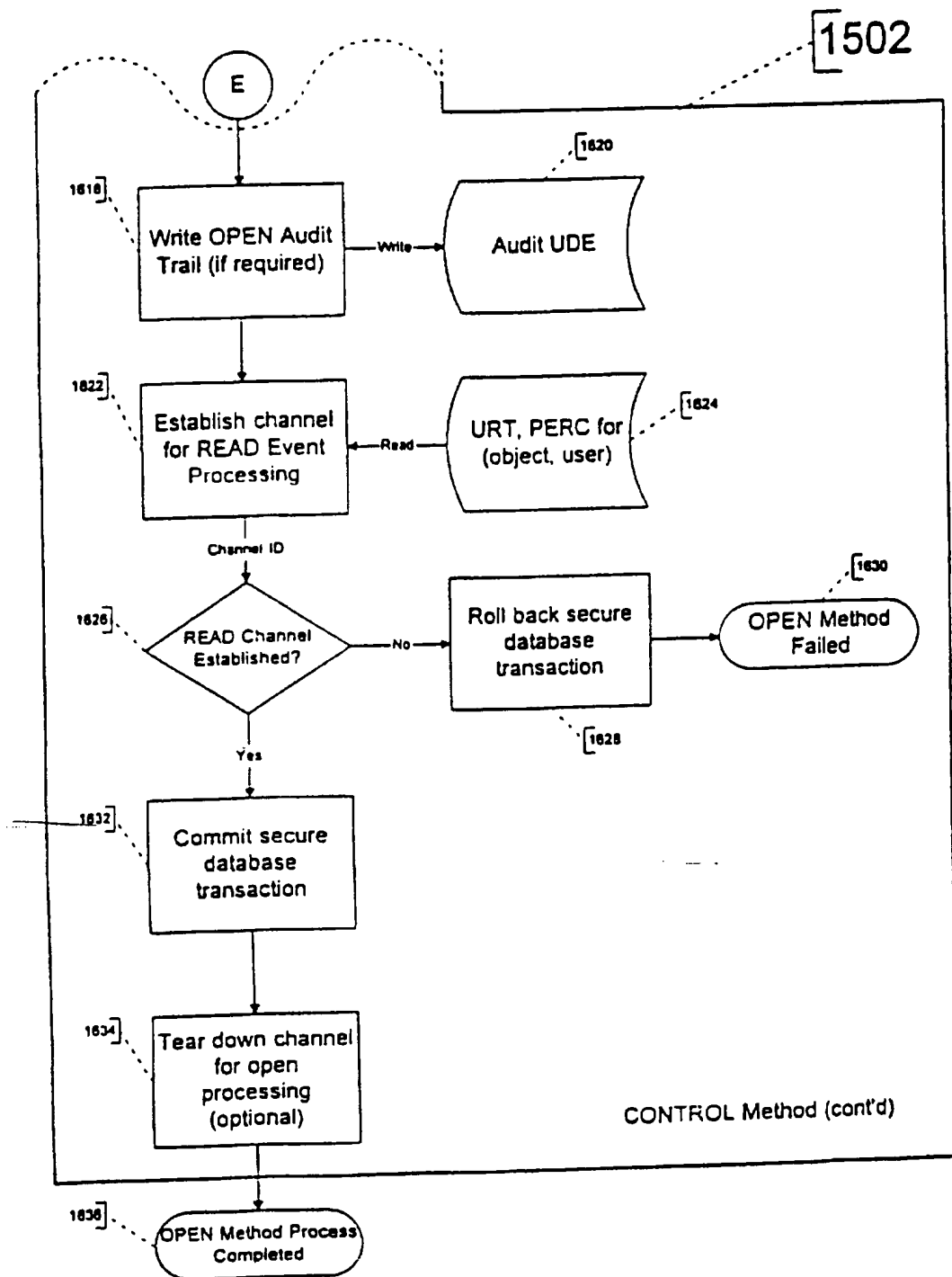


Figure 49f

86/163

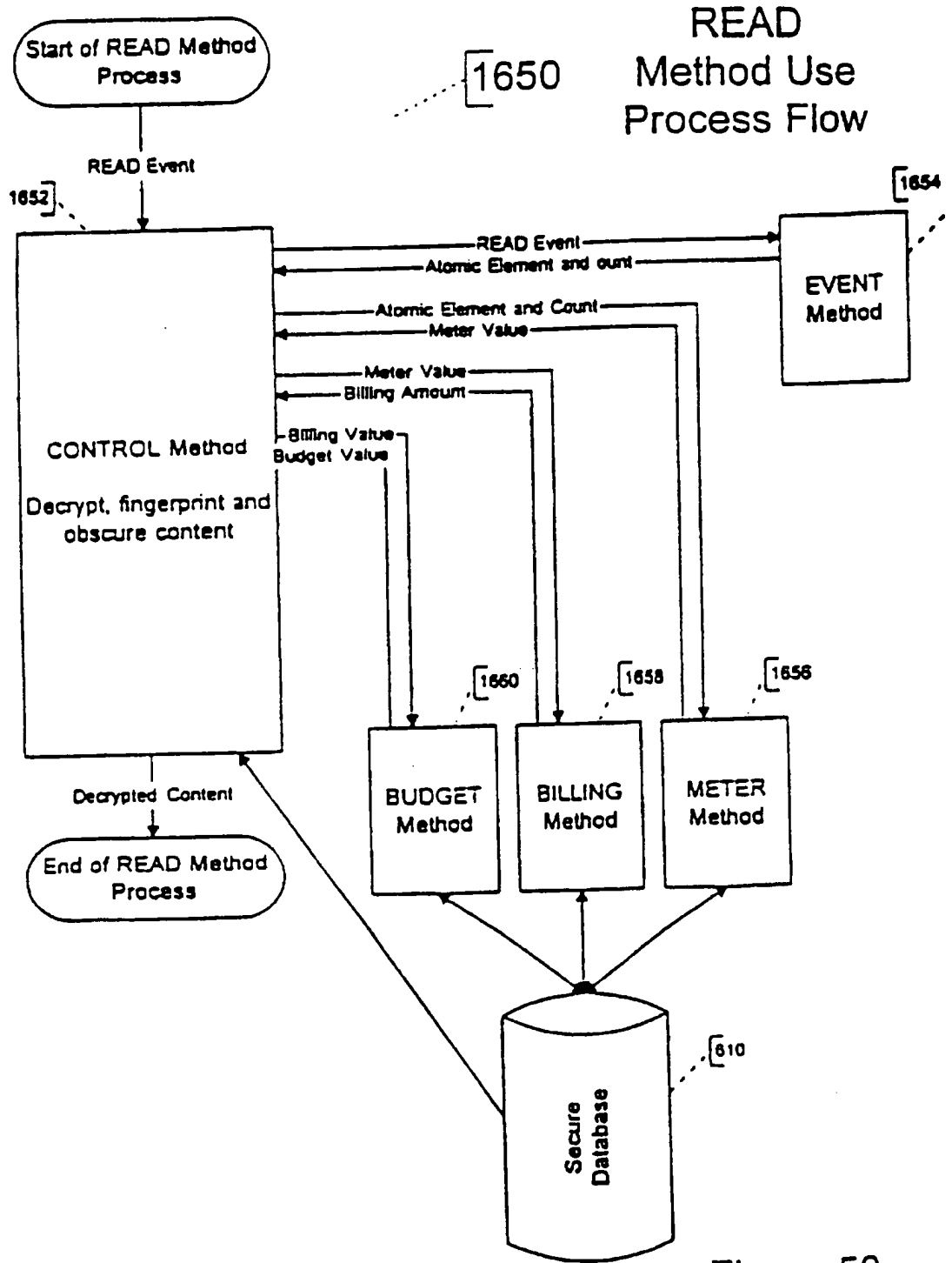


Figure 50

87/163

1650

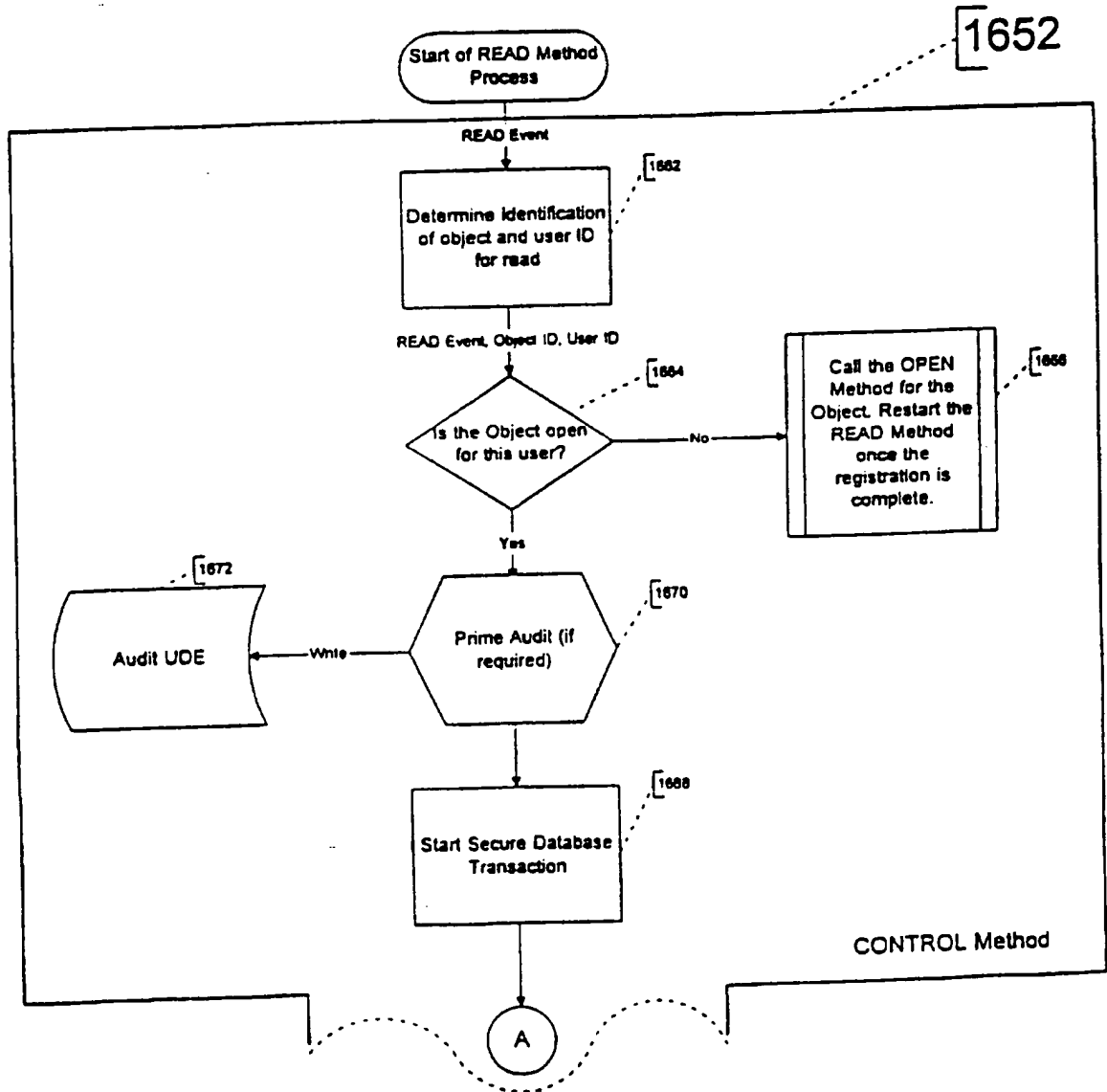


Figure 50a

SUBSTITUTE SHEET (RULE 26)

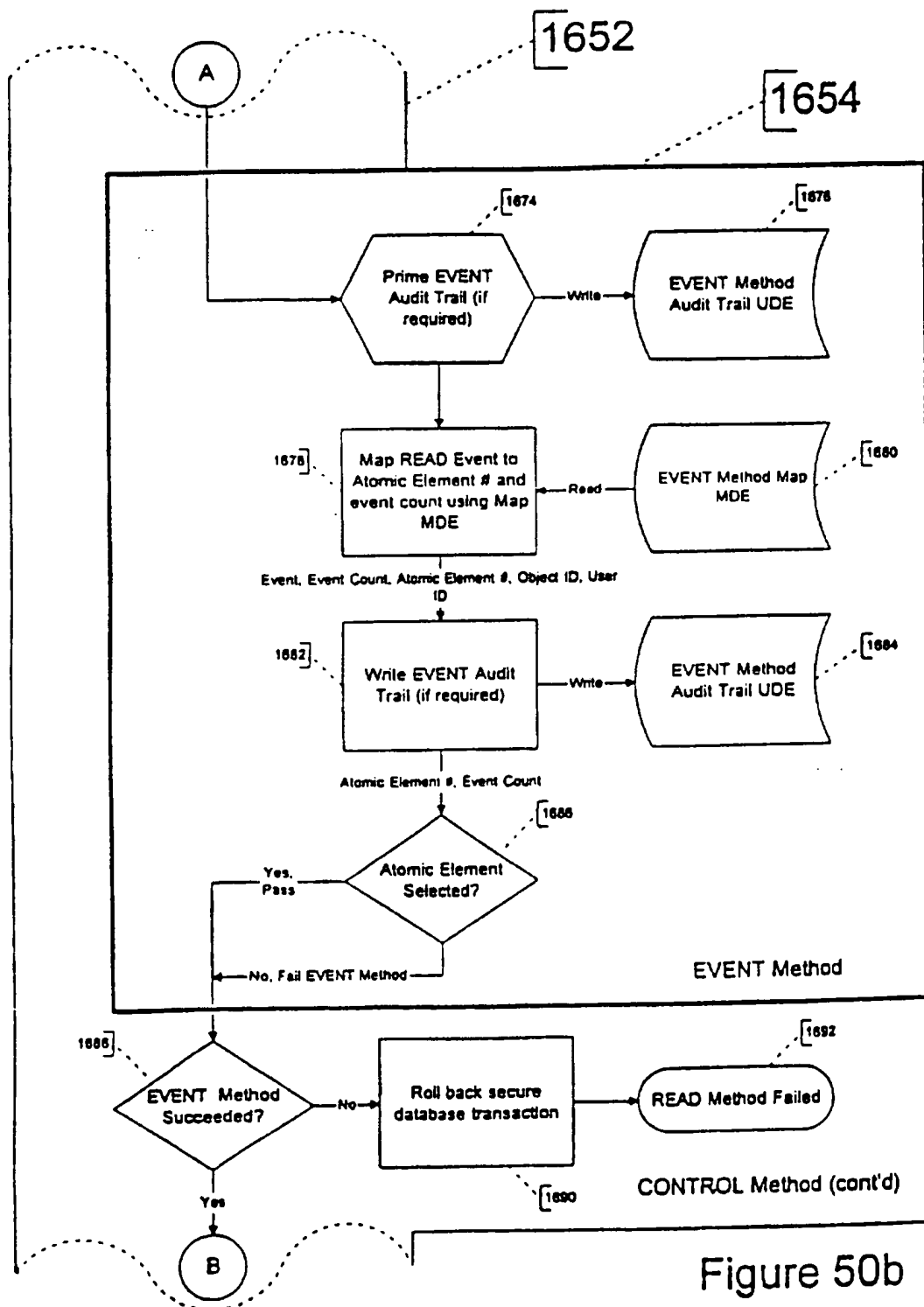


Figure 50b

SUBSTITUTE SHEET (RULE 26)

89/163

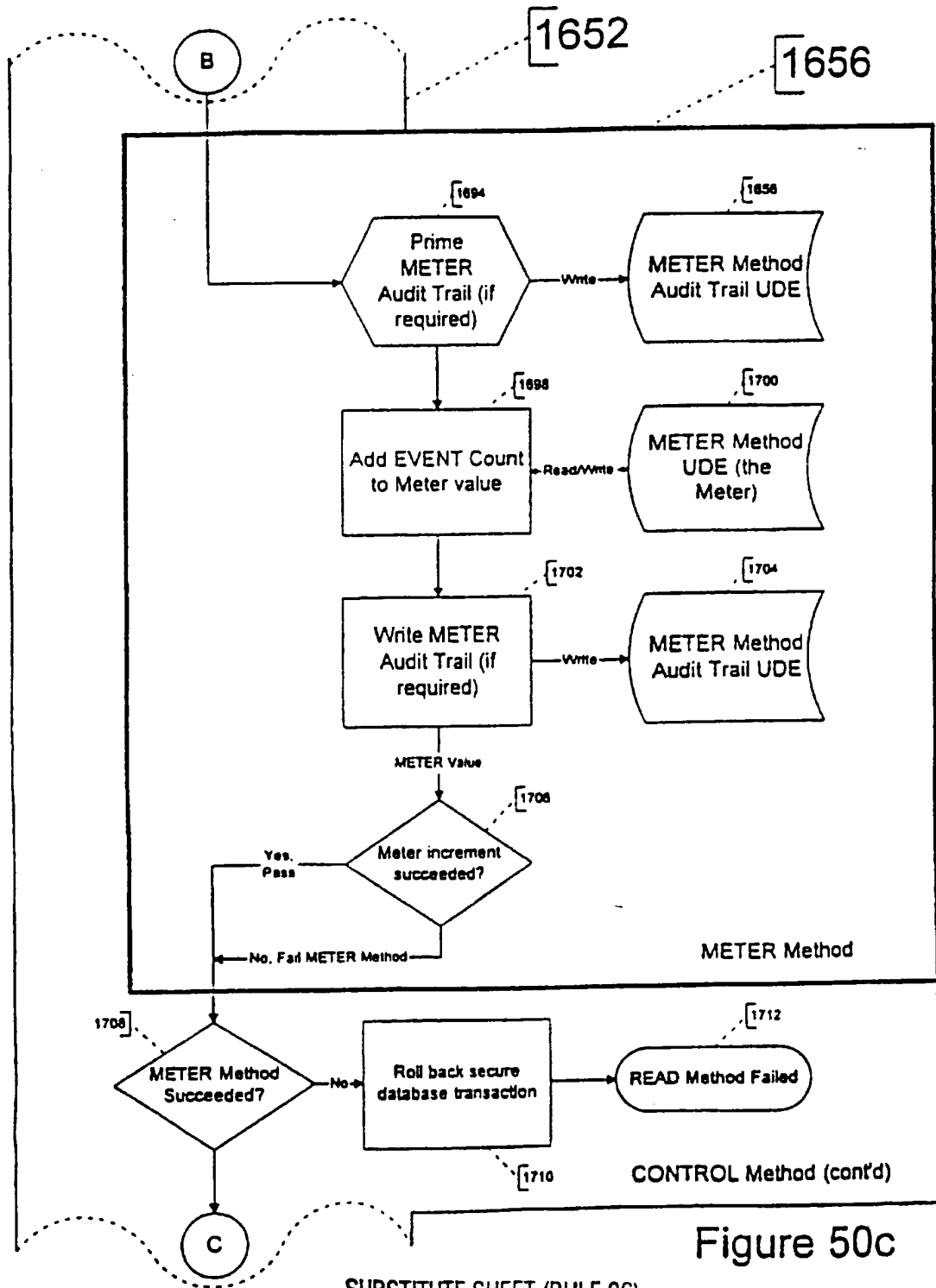


Figure 50c

SUBSTITUTE SHEET (RULE 26)

90/163

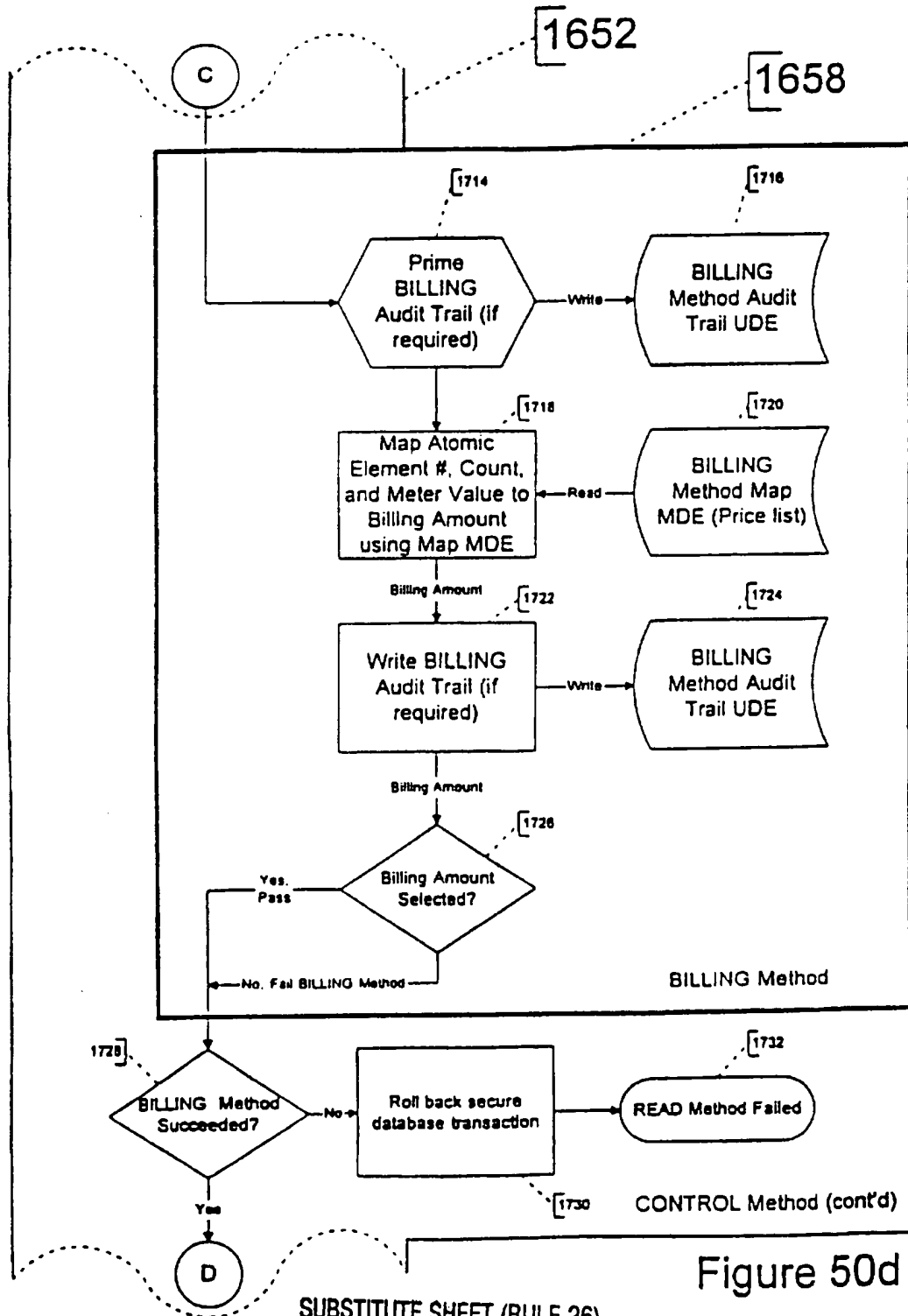


Figure 50d

SUBSTITUTE SHEET (RULE 26)

91/163

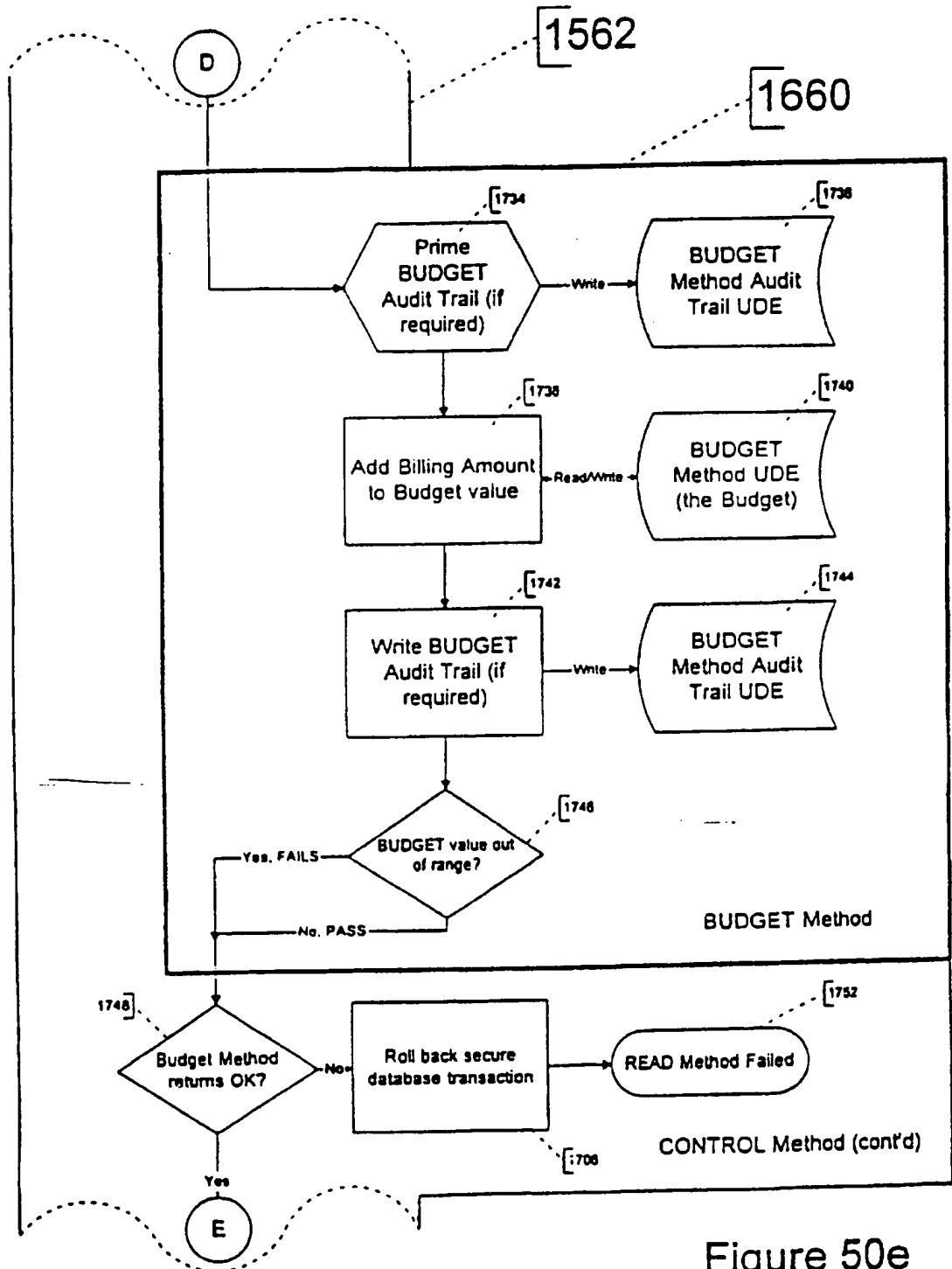
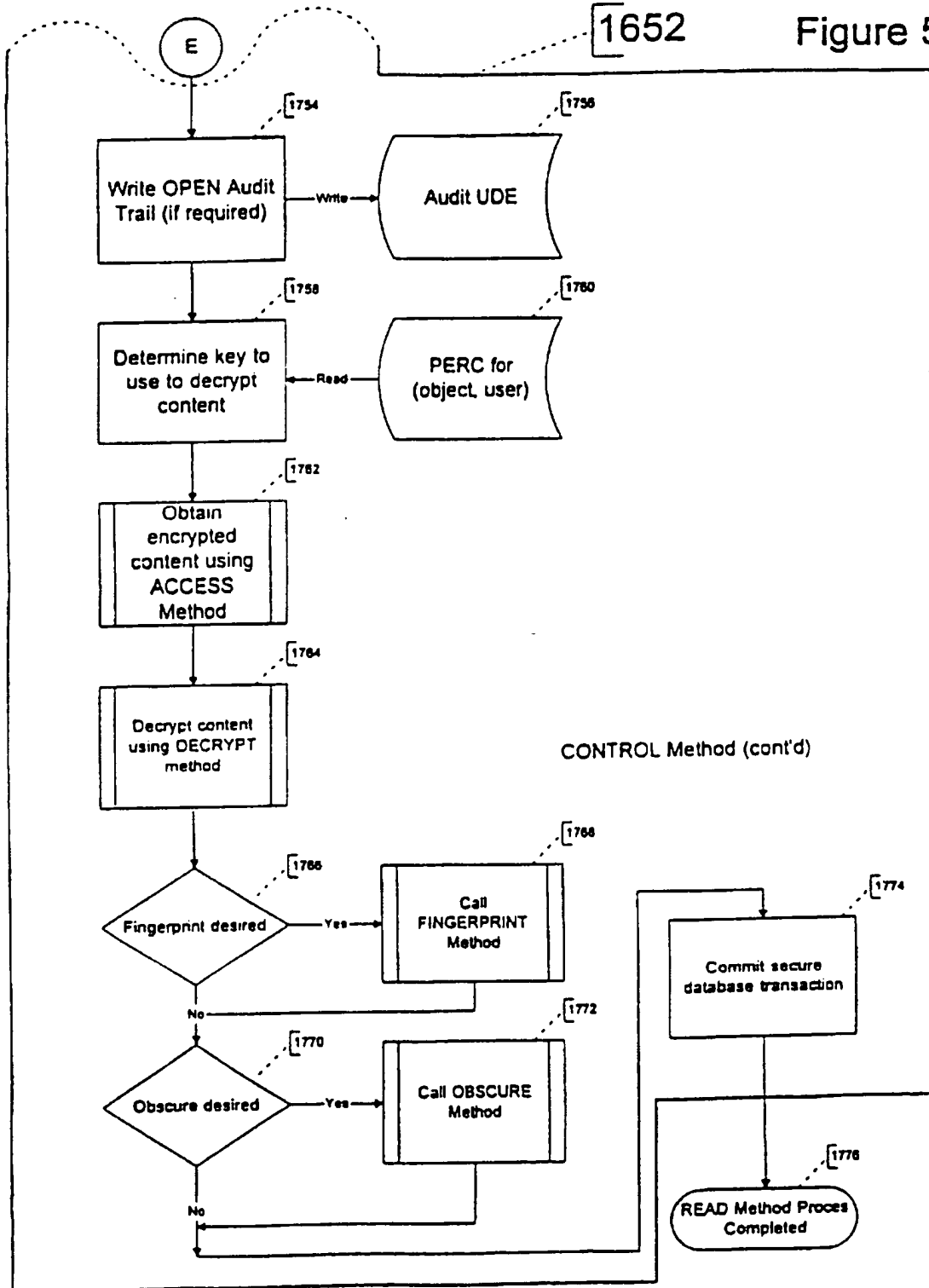


Figure 50e

92/163

1652

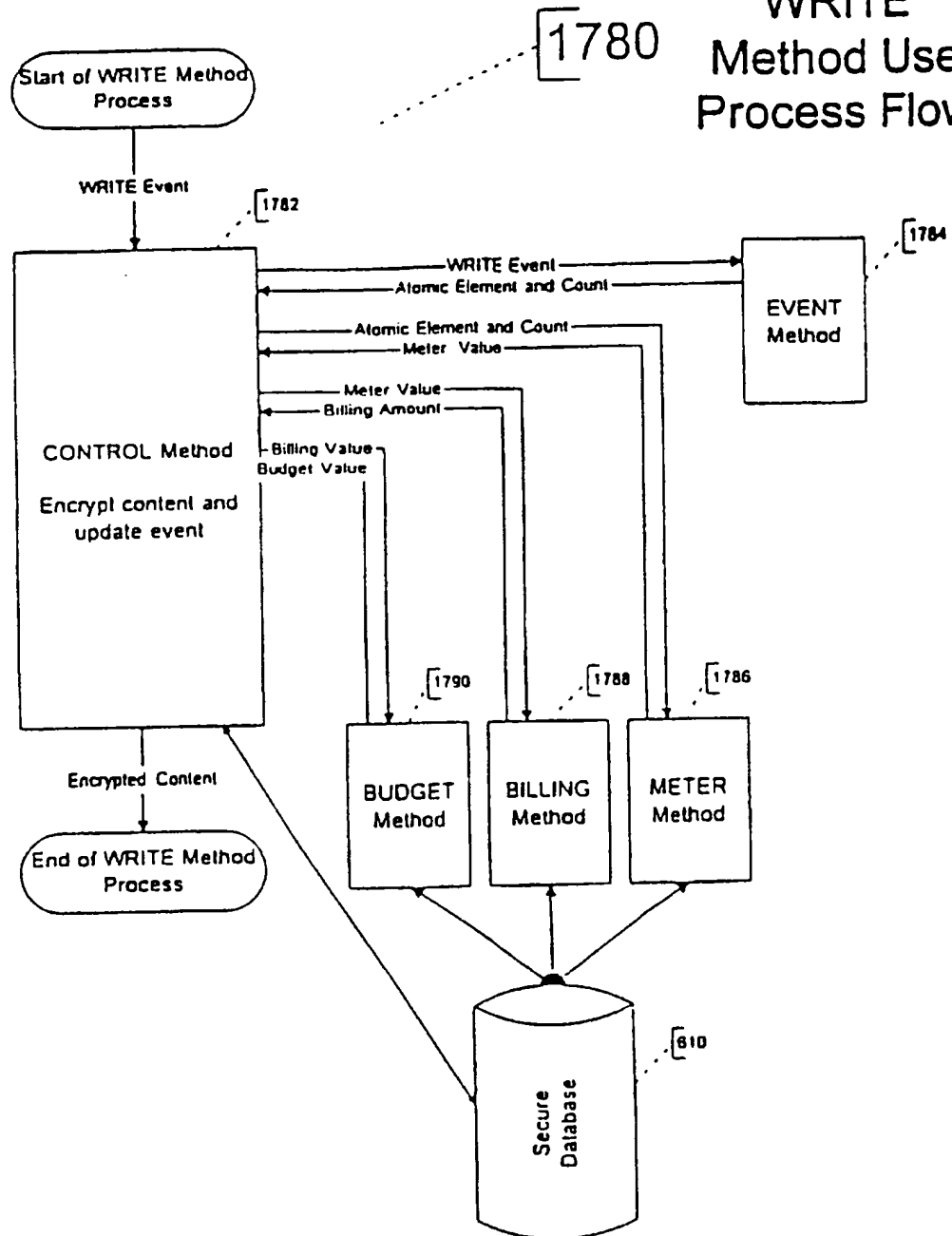
Figure 50f



SUBSTITUTE SHEET (RULE 26)

93/163

WRITE Method Use Process Flow



94/163

1780

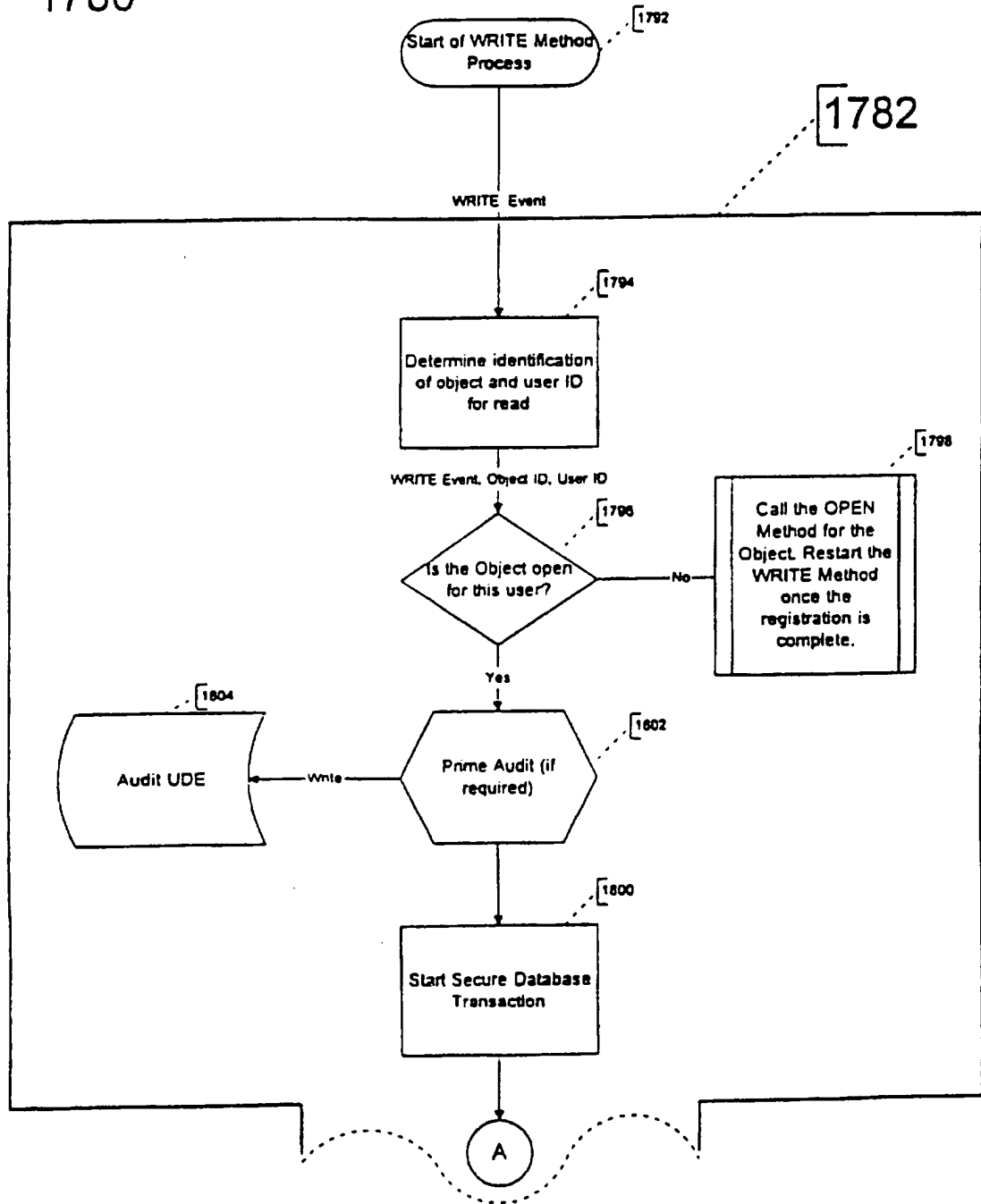


Figure 51a

SUBSTITUTE SHEET (RULE 26)

95/163

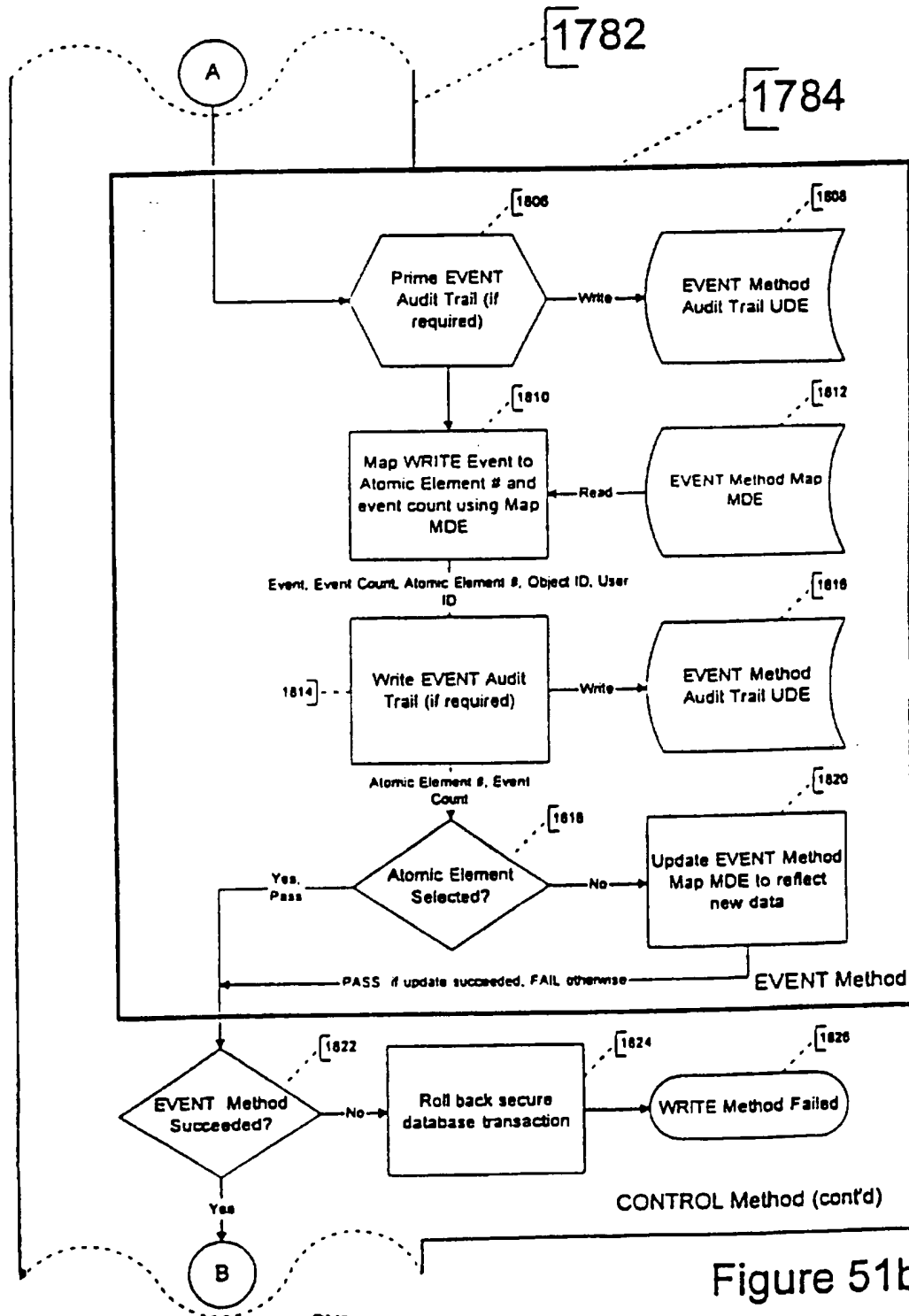


Figure 51b

SUBSTITUTE SHEET (RULE 26)

96/163

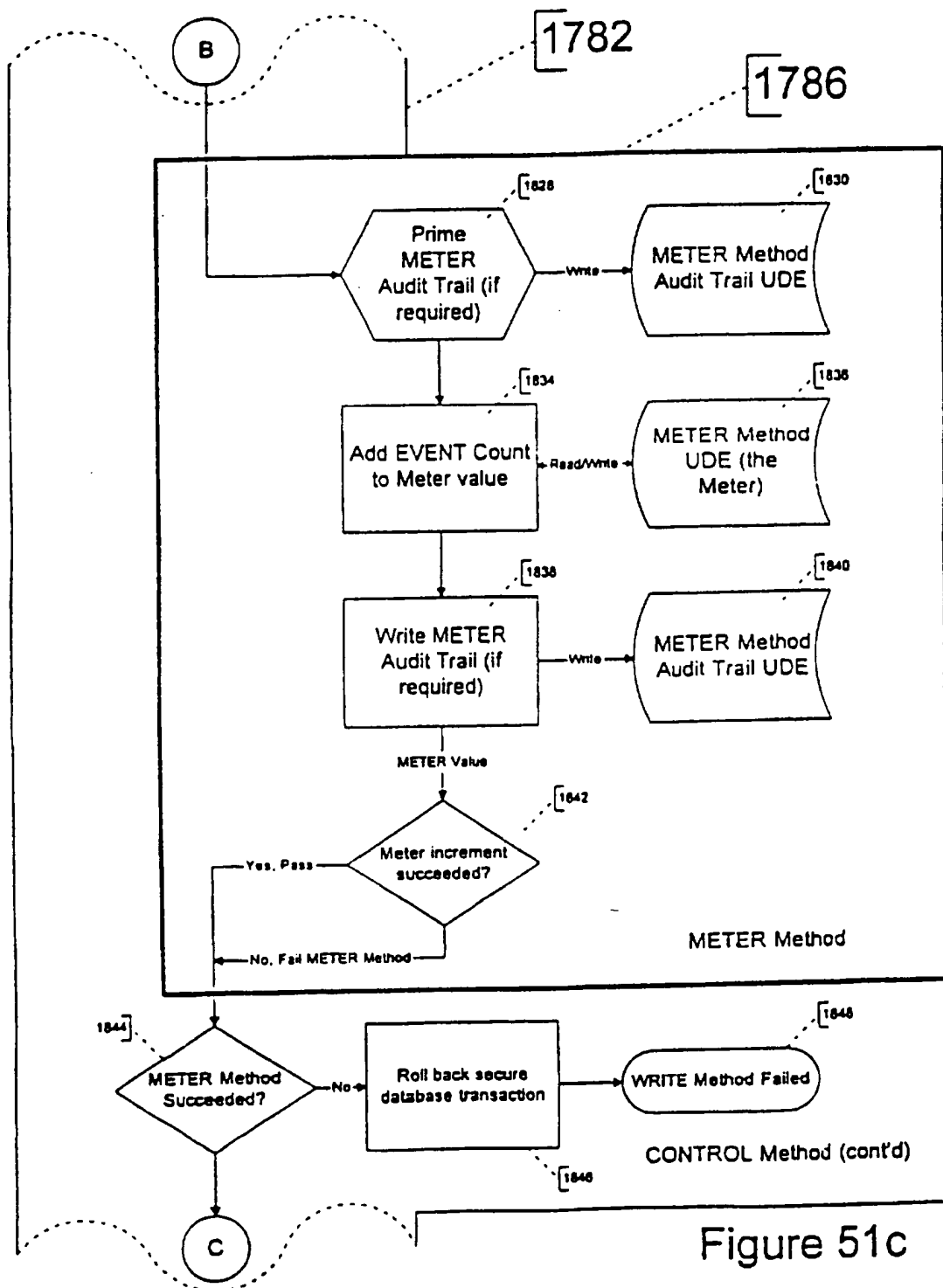


Figure 51c

SUBSTITUTE SHEET (RULE 26)

97/163

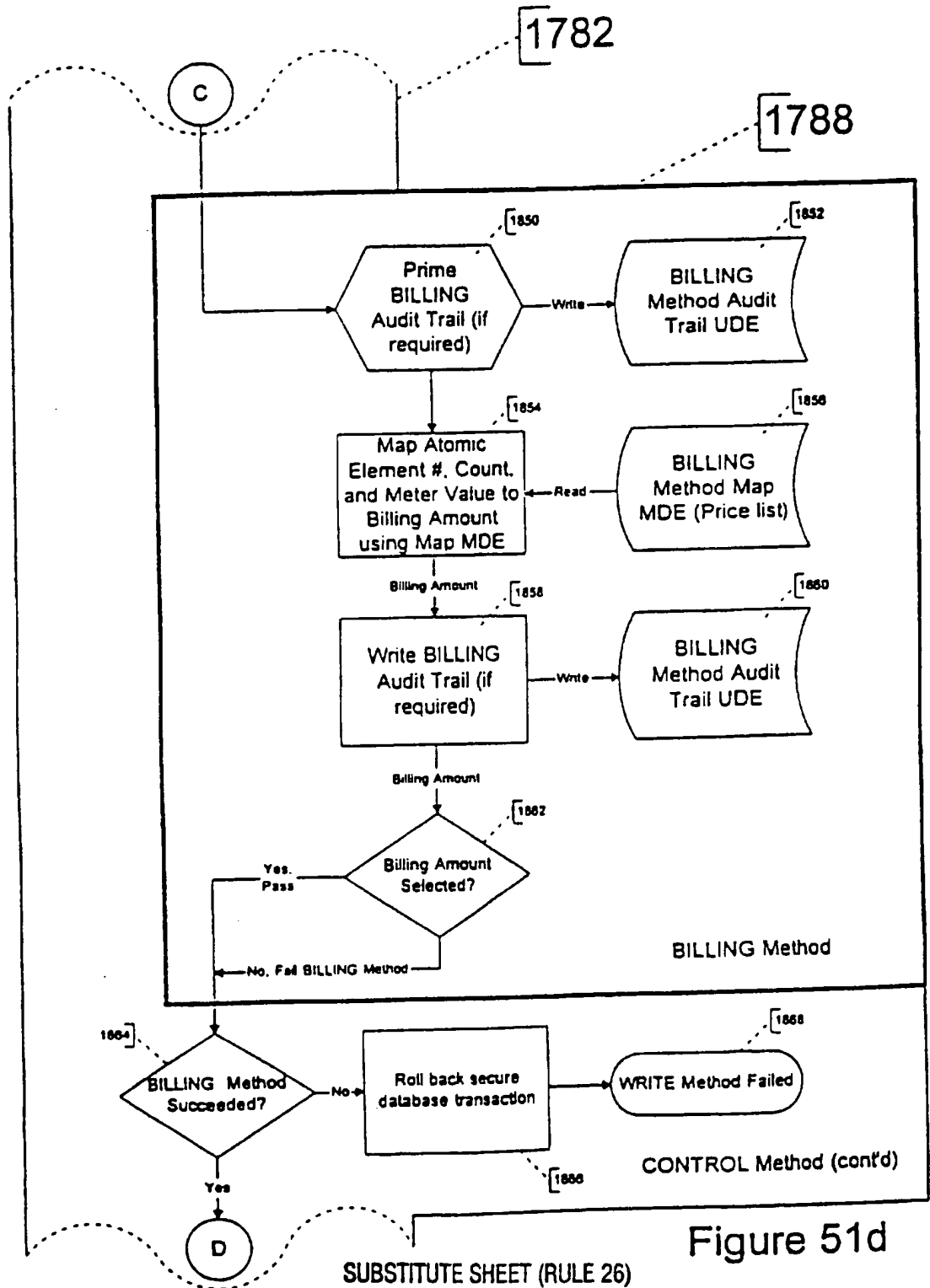


Figure 51d

SUBSTITUTE SHEET (RULE 26)

98/163

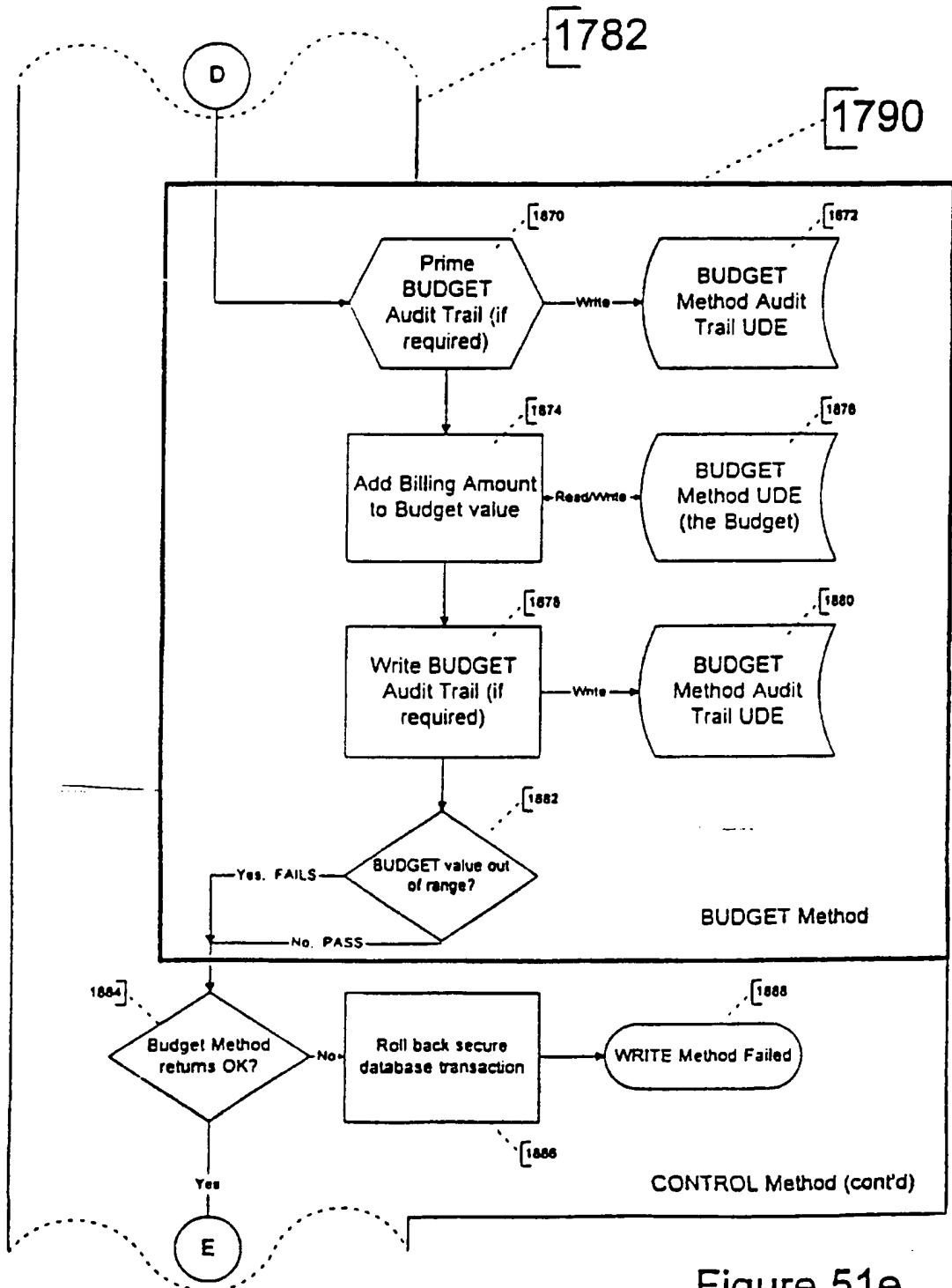


Figure 51e

SUBSTITUTE SHEET (RULE 26)

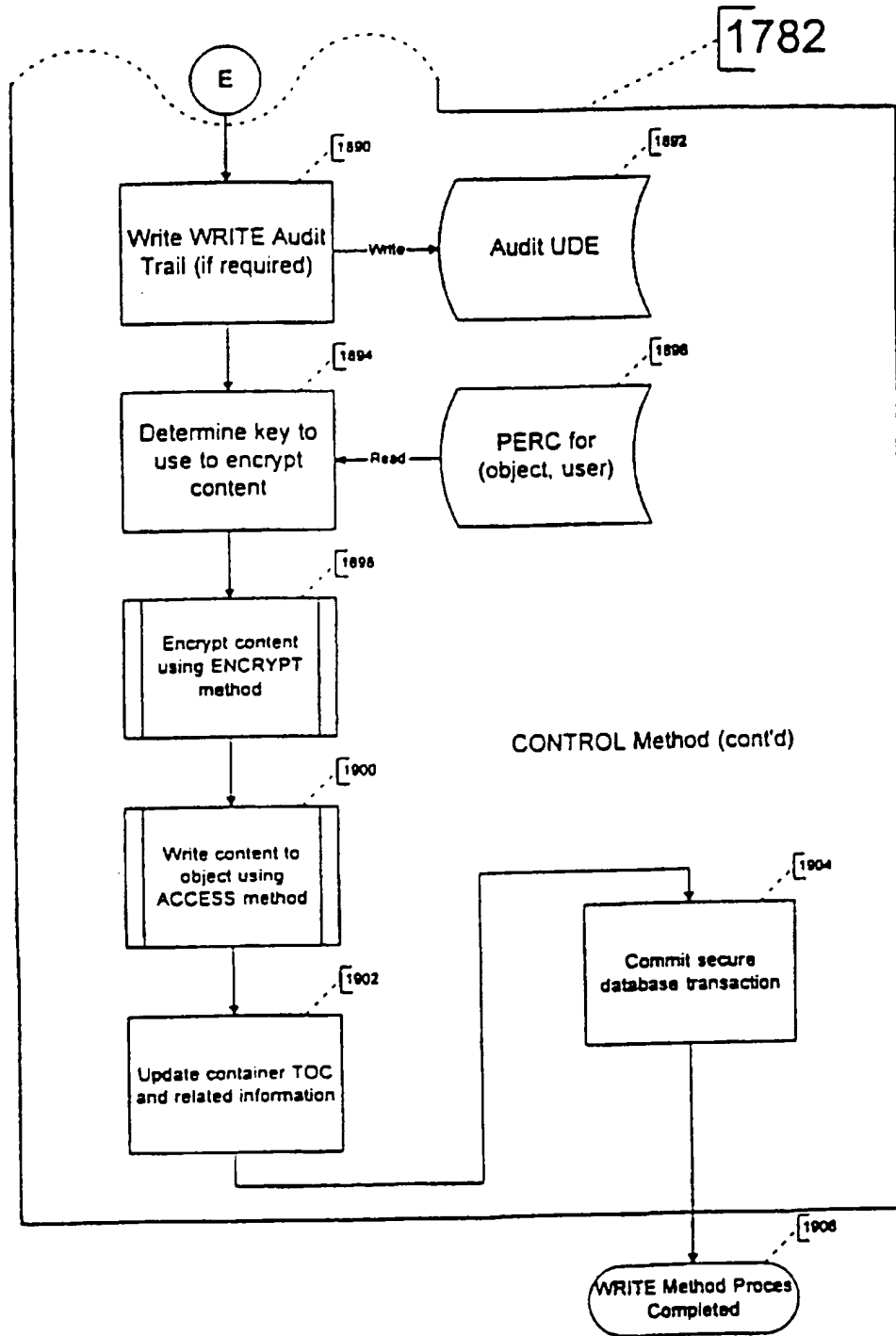
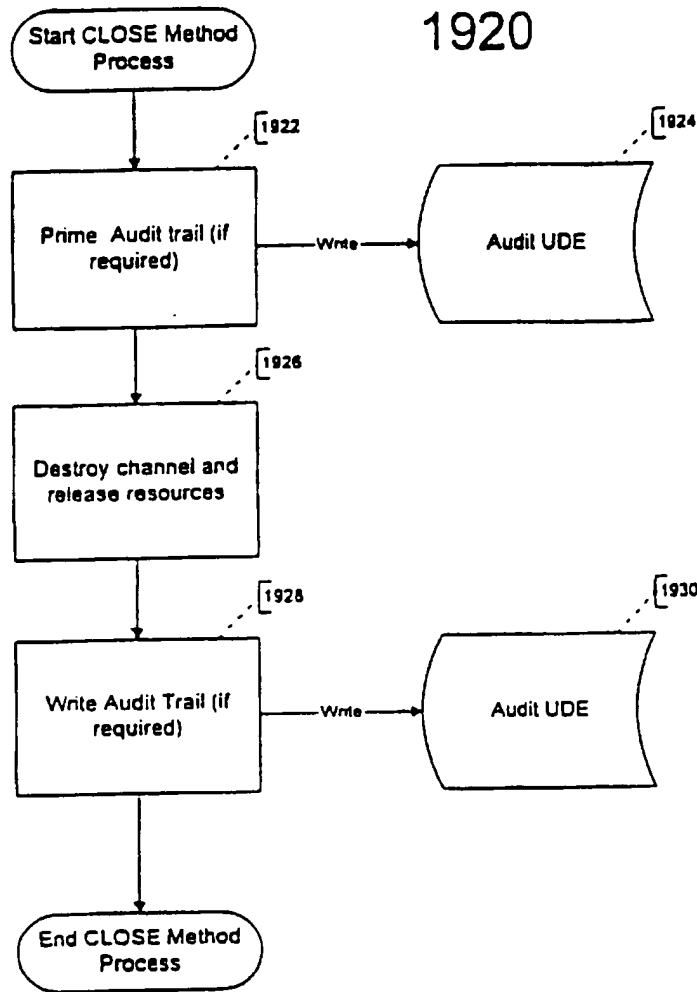


Figure 51f

100/163



CLOSE Method Process Flow

Figure 52

101/163

EVENT Method Process Flows

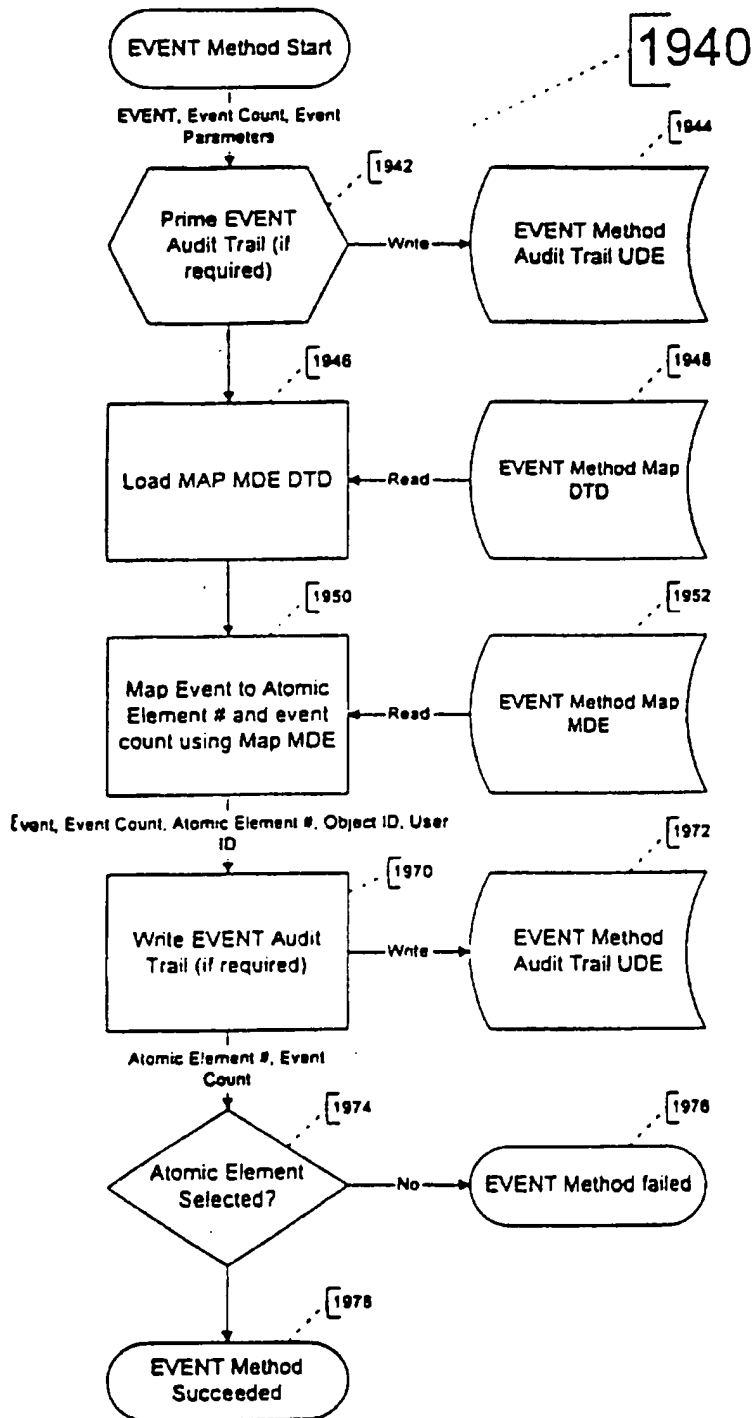


Figure 53a

Sample EVENT Method Mapping Process

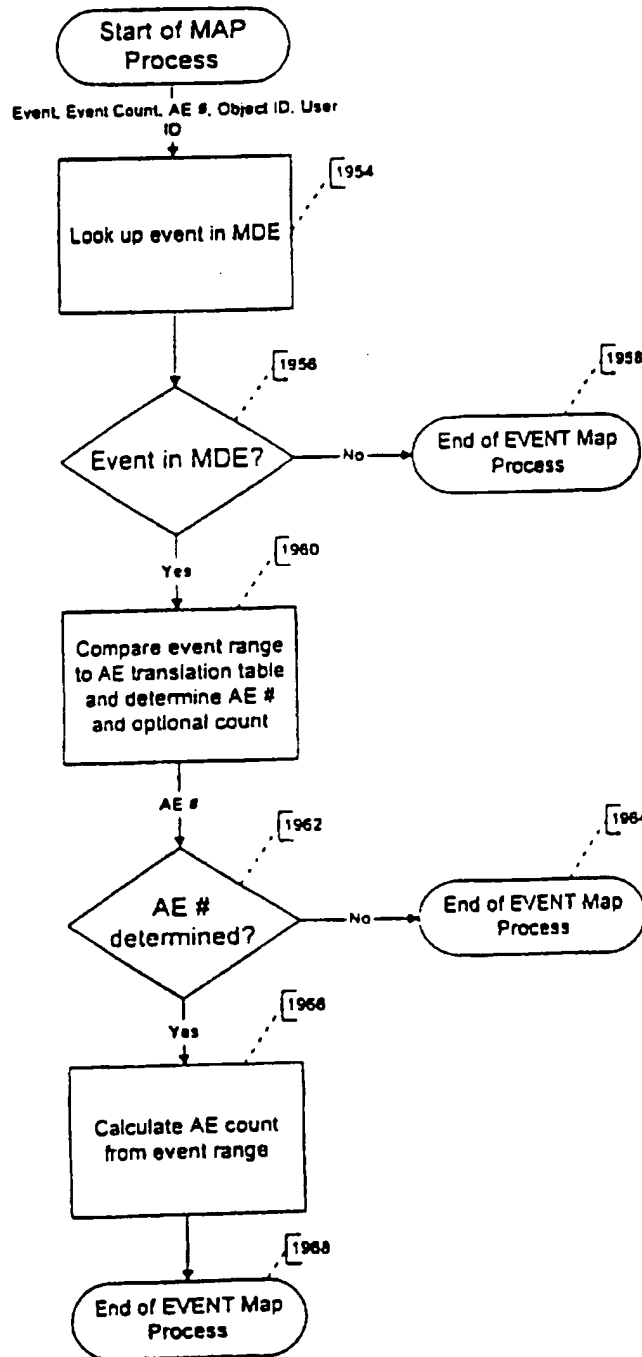
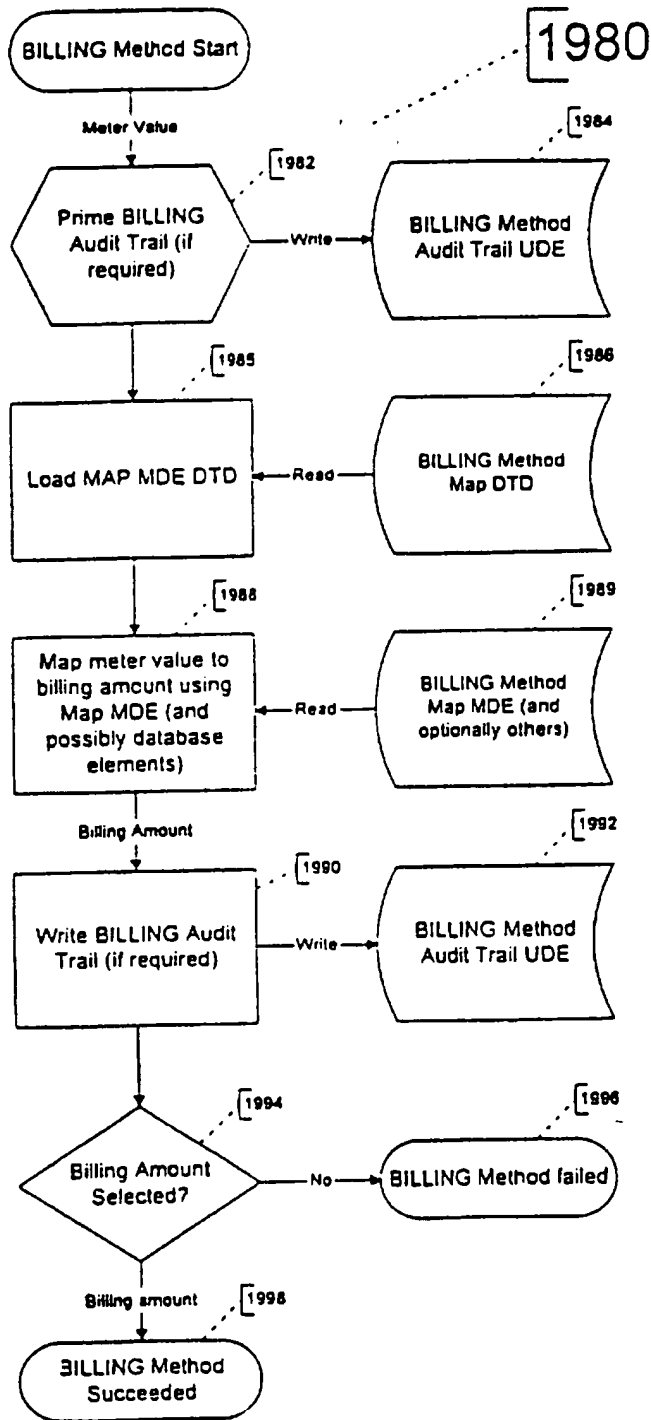


Figure 53b

103/163

BILLING Method Process Flows



104/163

ACCESS Method Process Flow

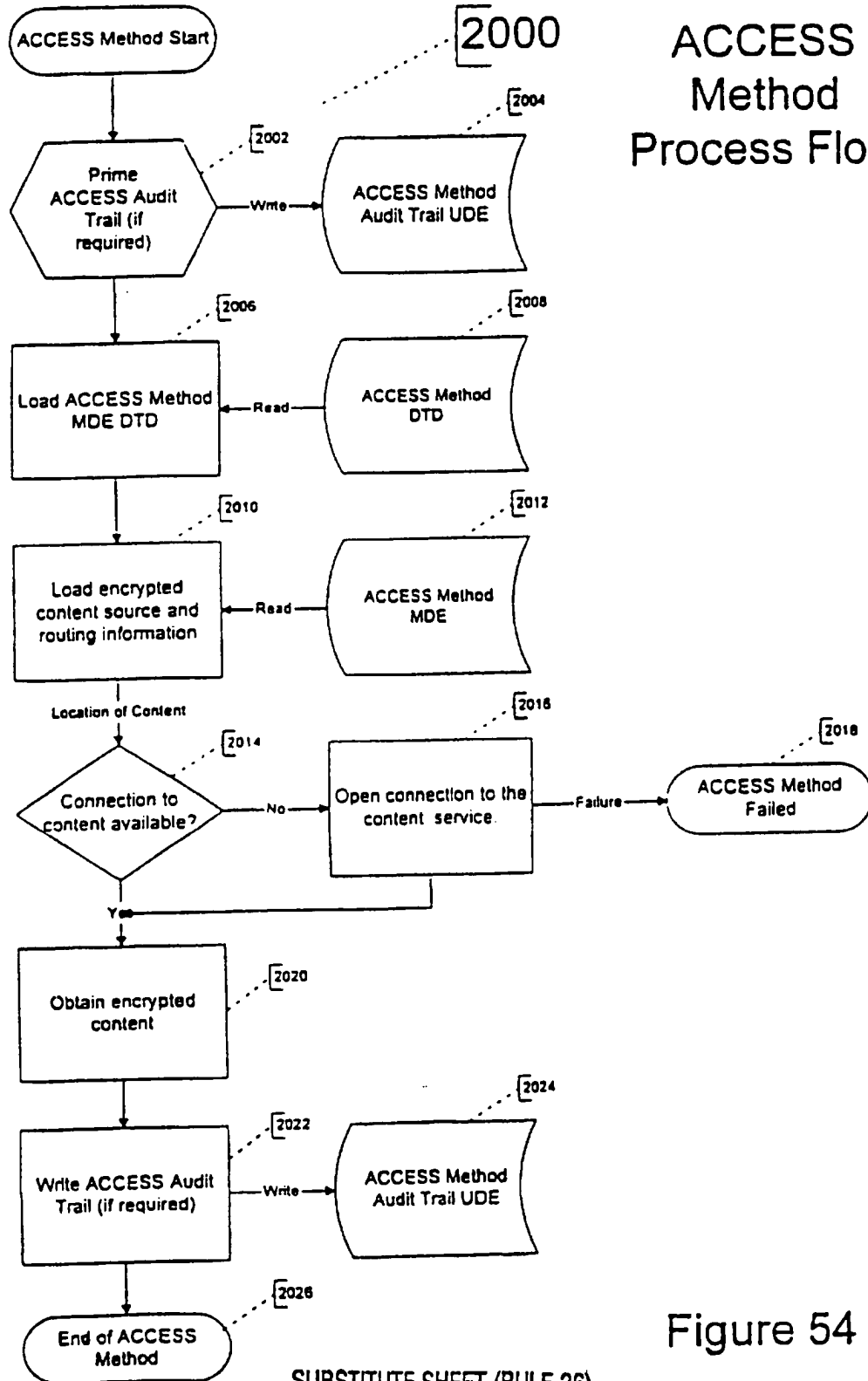


Figure 54

SUBSTITUTE SHEET (RULE 26)

105/163

DECRYPT Method Process Flow

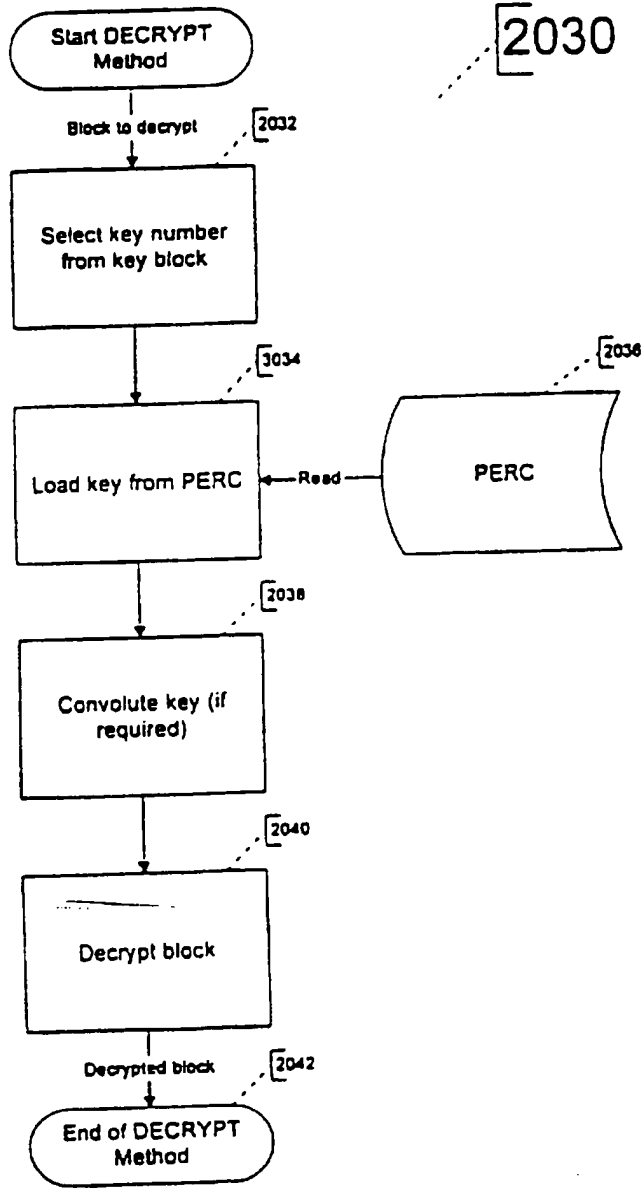
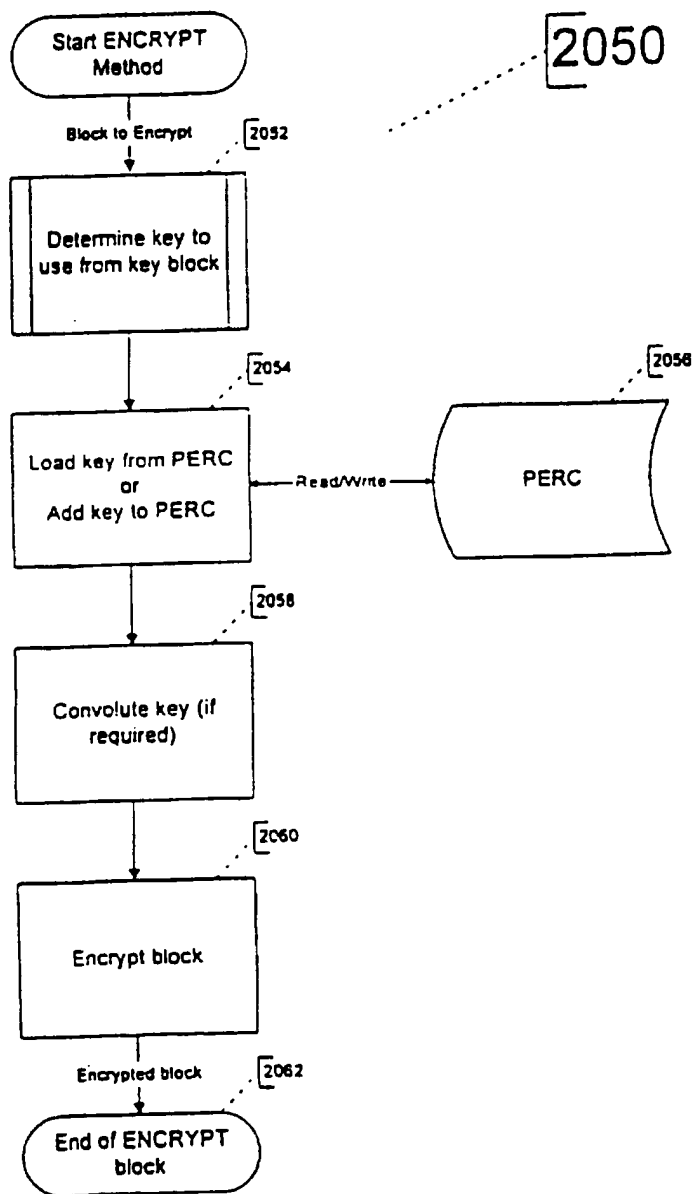


Figure 55a

SUBSTITUTE SHEET (RULE 26)

106/163

ENCRYPT Method Process Flow



107/163

CONTENT Method Process Flow

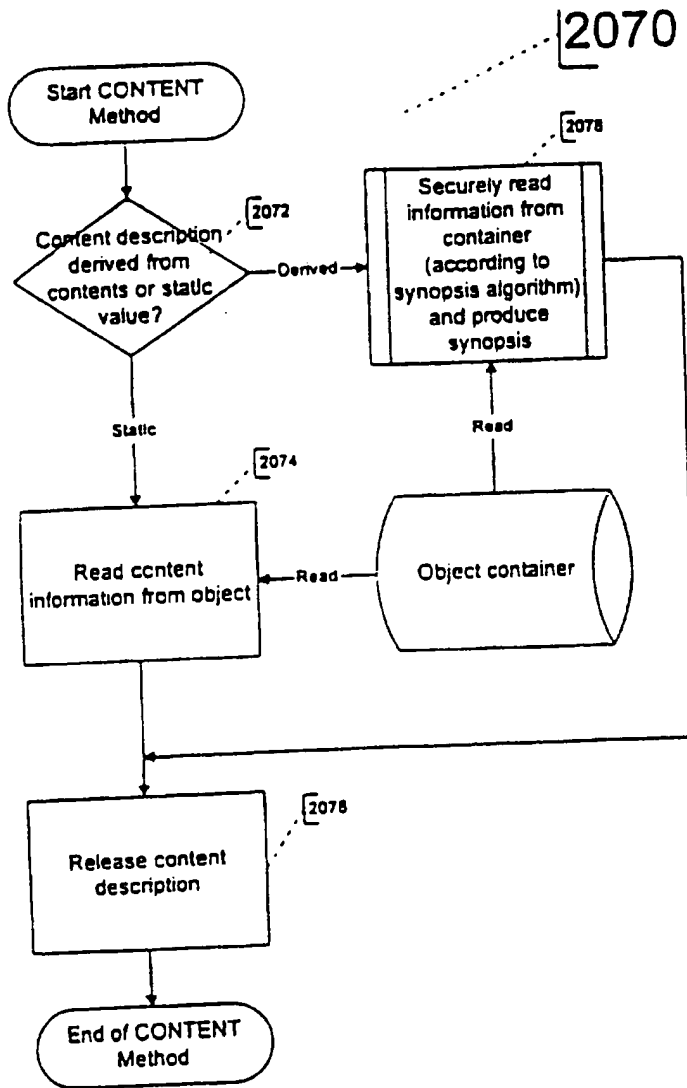
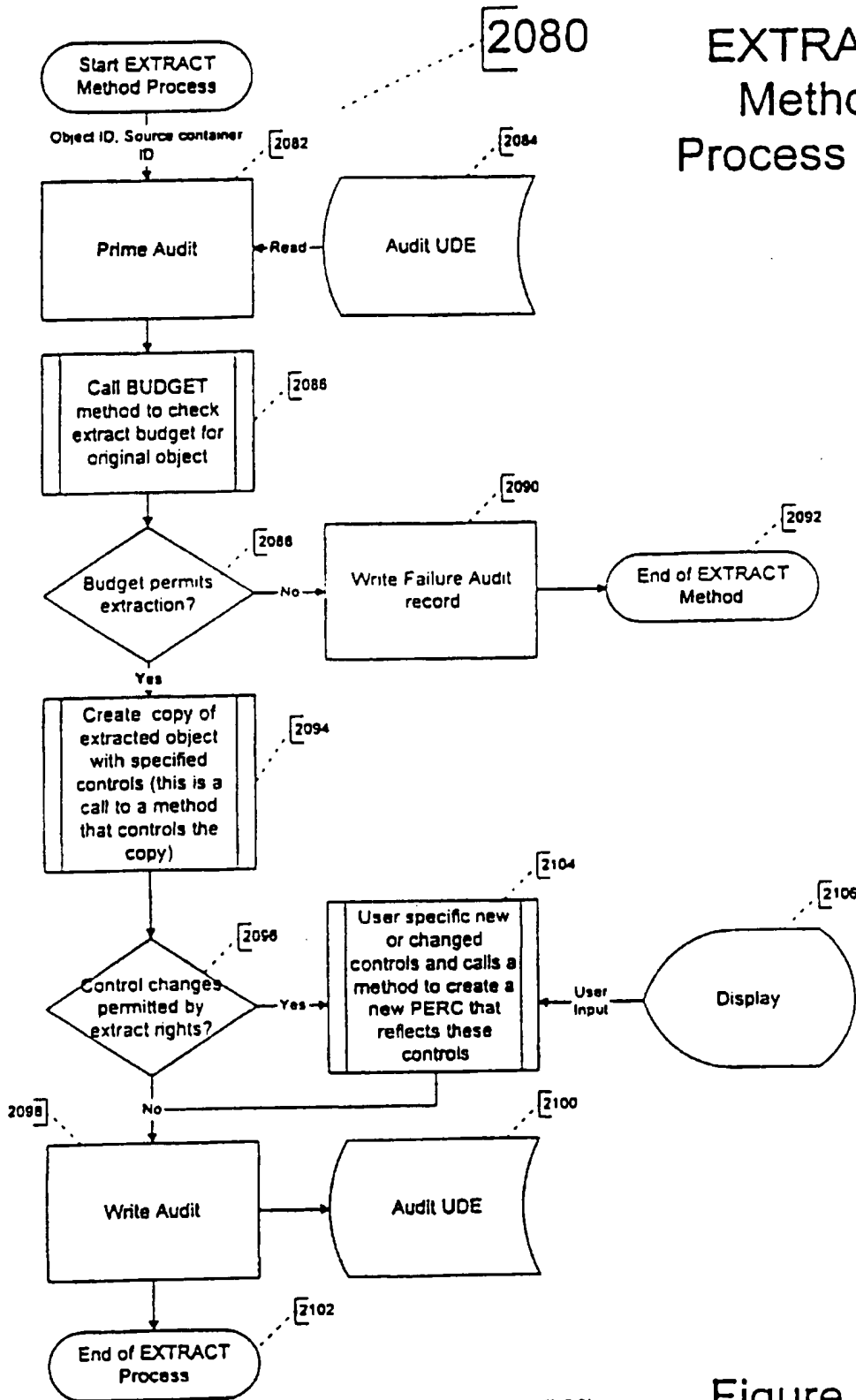


Figure 56

108/163

EXTRACT Method Process Flow



SUBSTITUTE SHEET (RULE 26)

Figure 57a

109/163

EMBED Method Process Flow

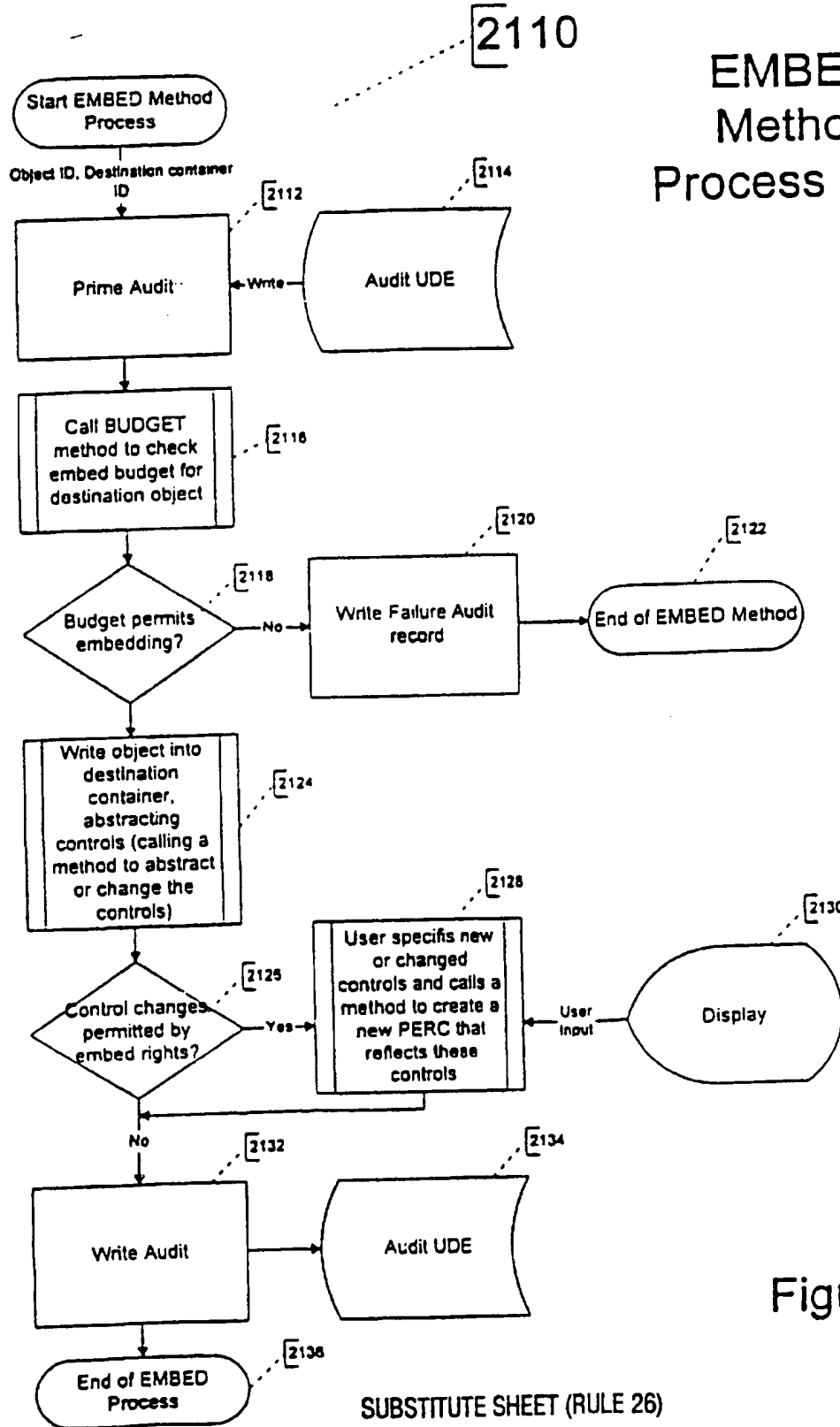
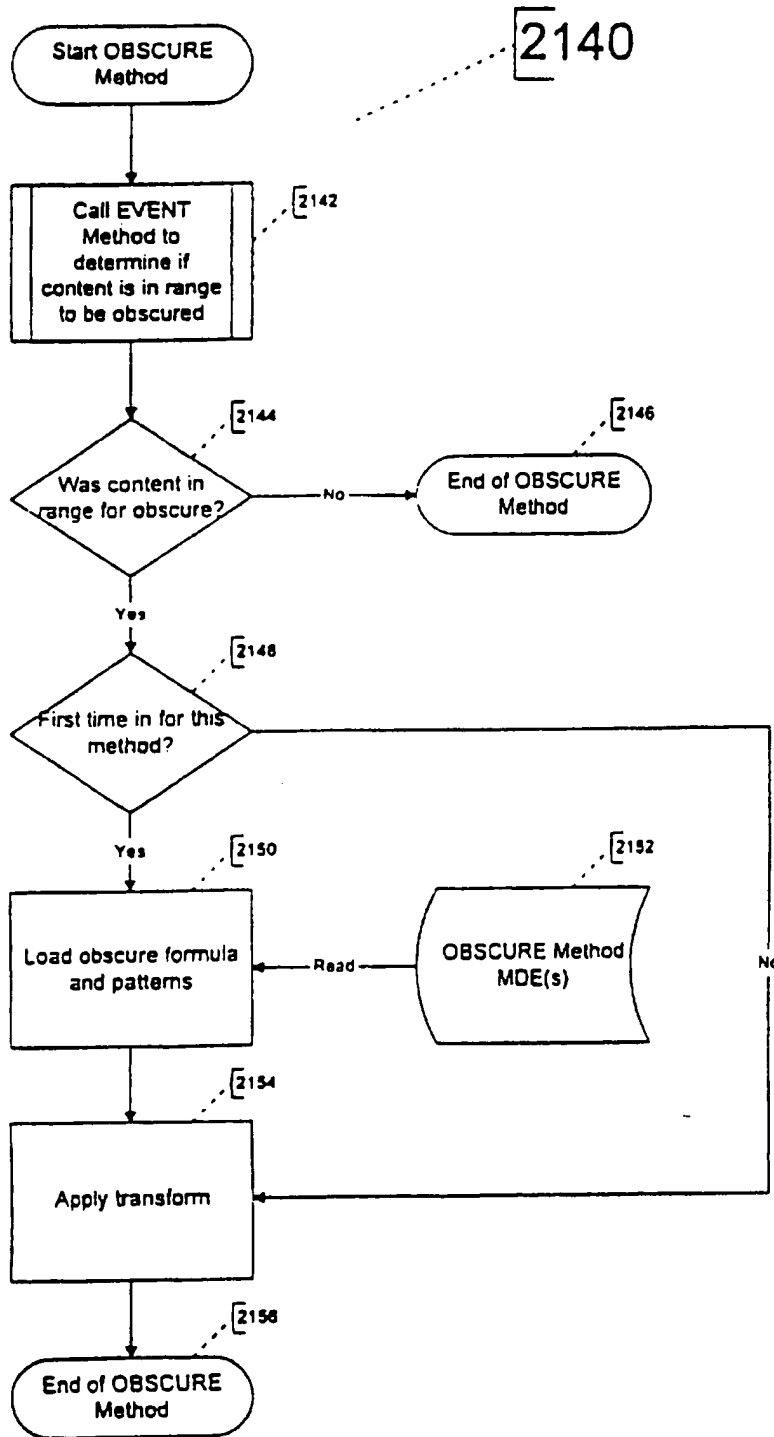


Figure 57b

SUBSTITUTE SHEET (RULE 26)

110/163

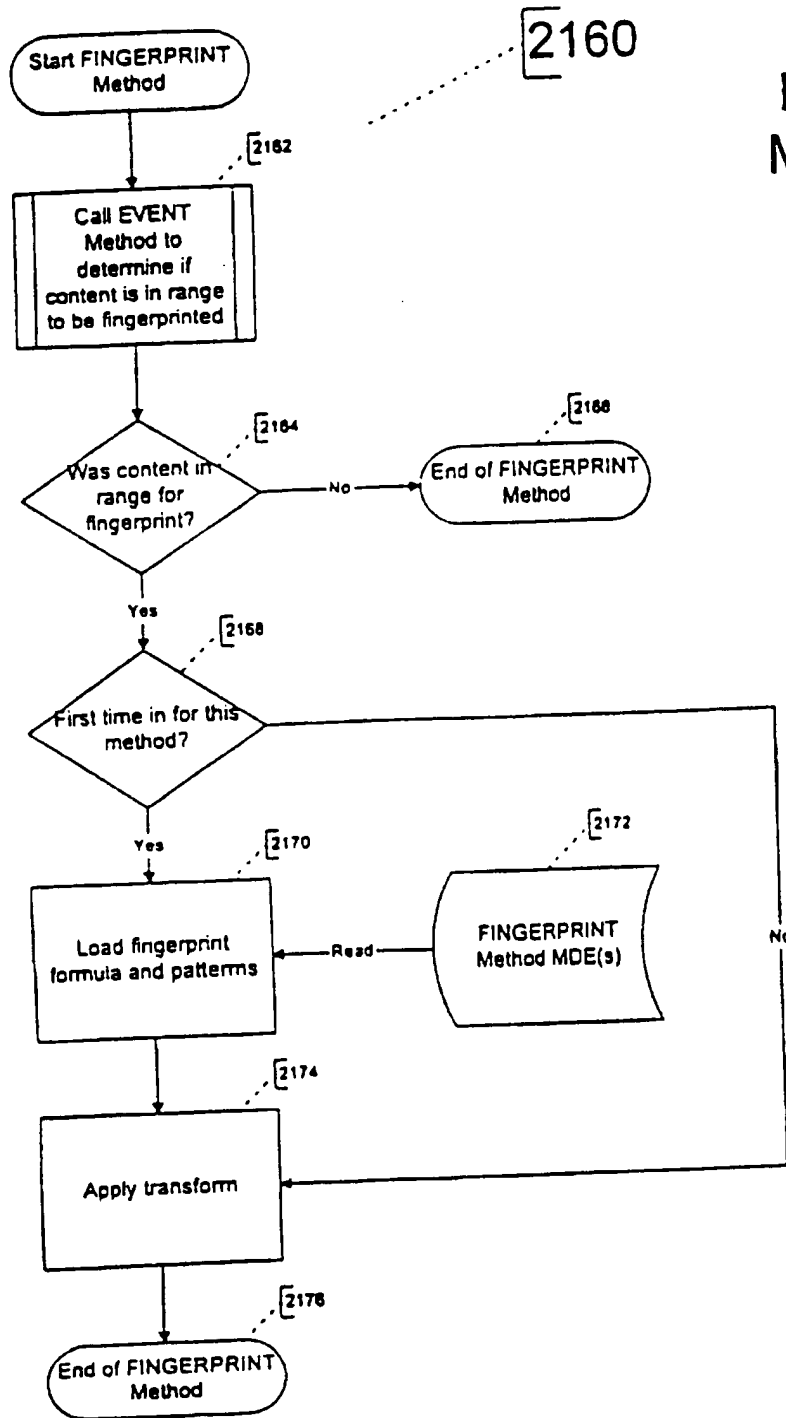


OBSCURE Method Process Flow

Figure 58a

SUBSTITUTE SHEET (RULE 26)

111/163



FINGERPRINT Method Process Flow

Figure 58b

112/163

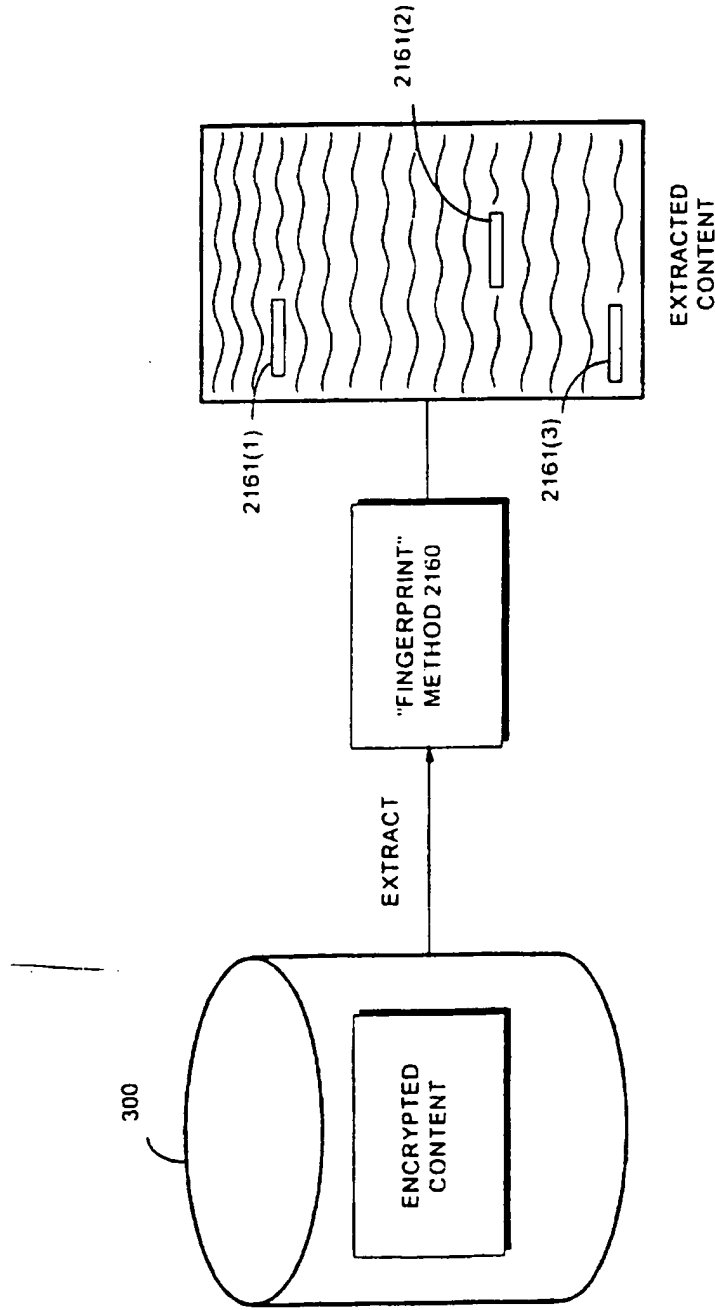


FIG. 58C

DESTROY Method Process Flow

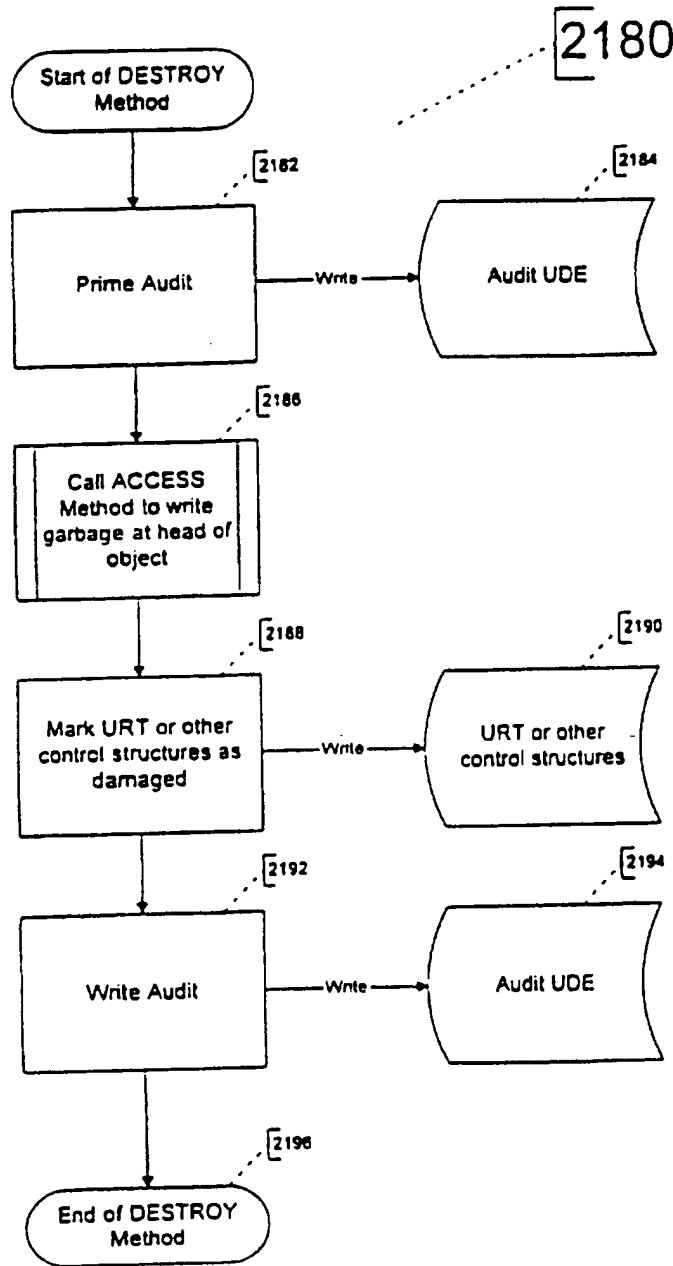
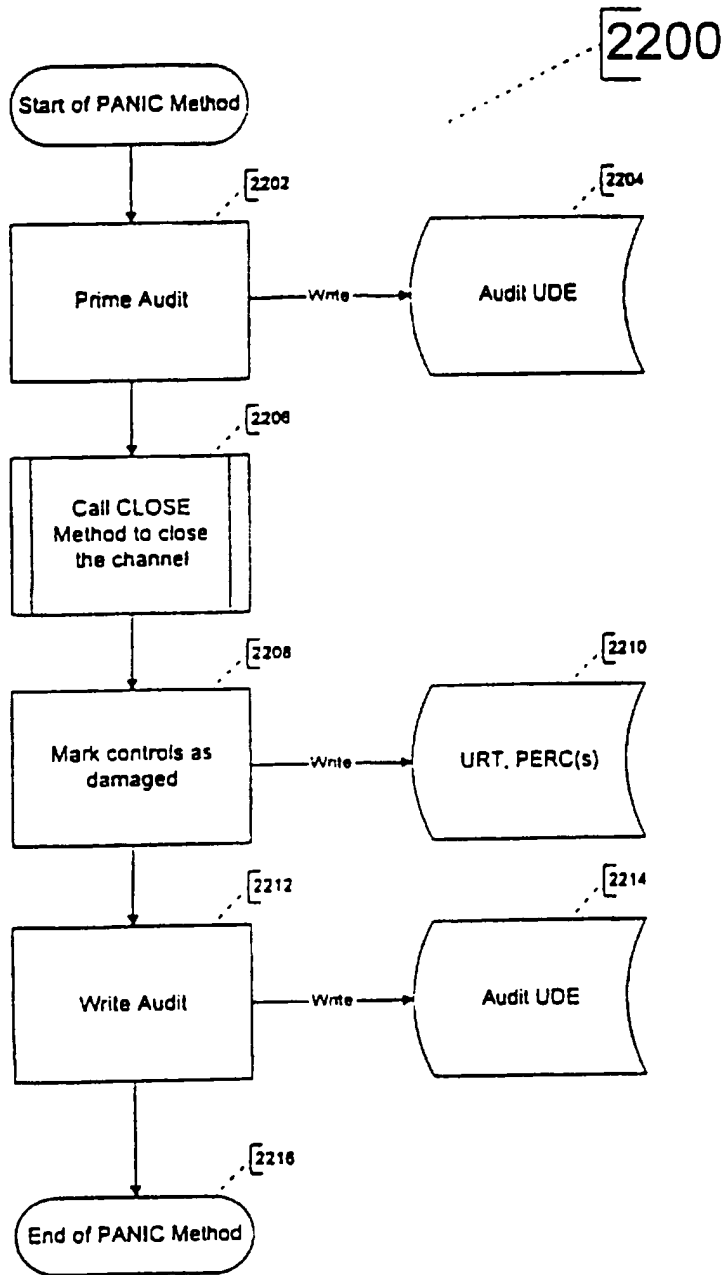


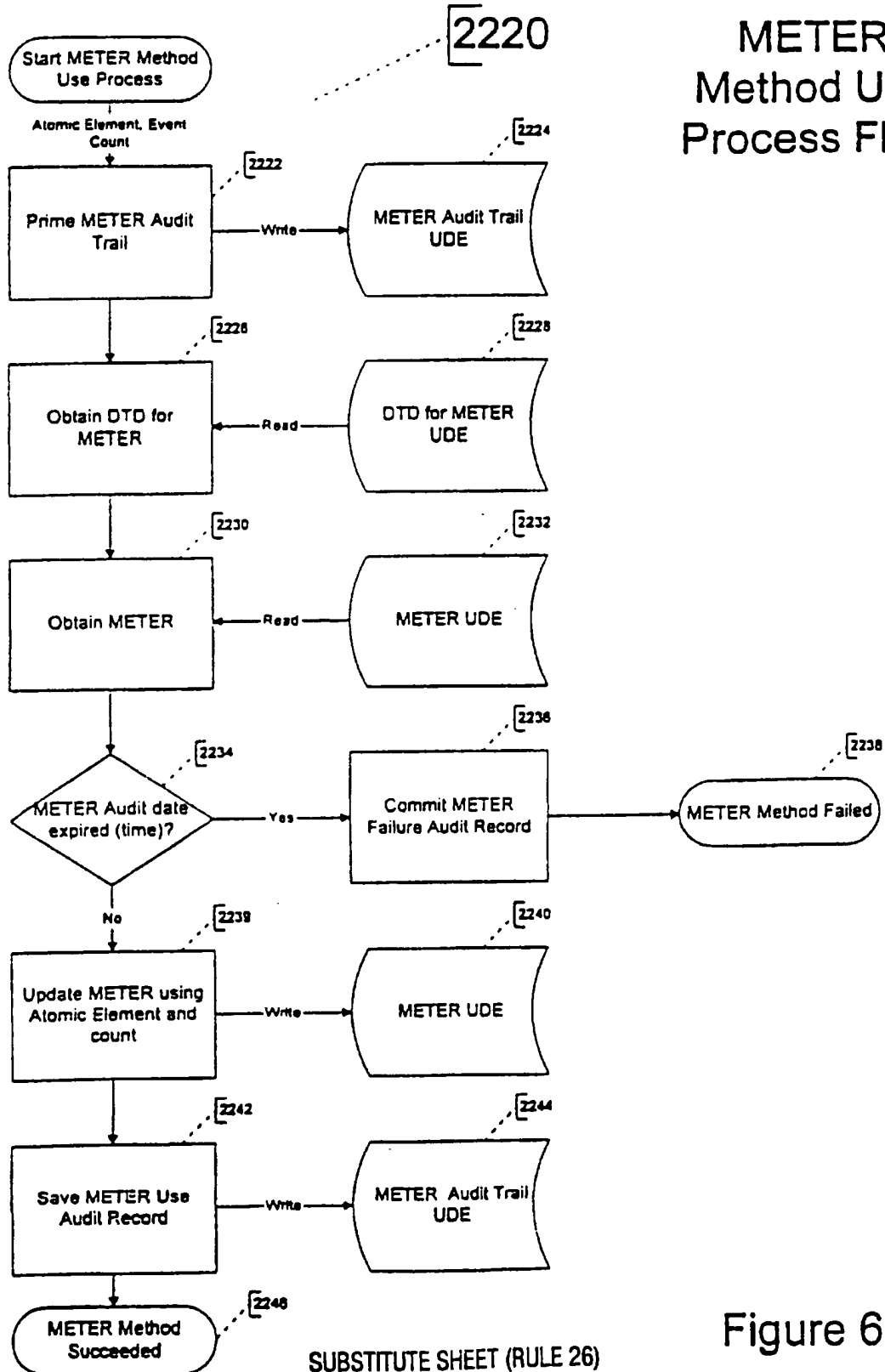
Figure 59

114/163

PANIC Method Process Flow



METER Method Use Process Flow

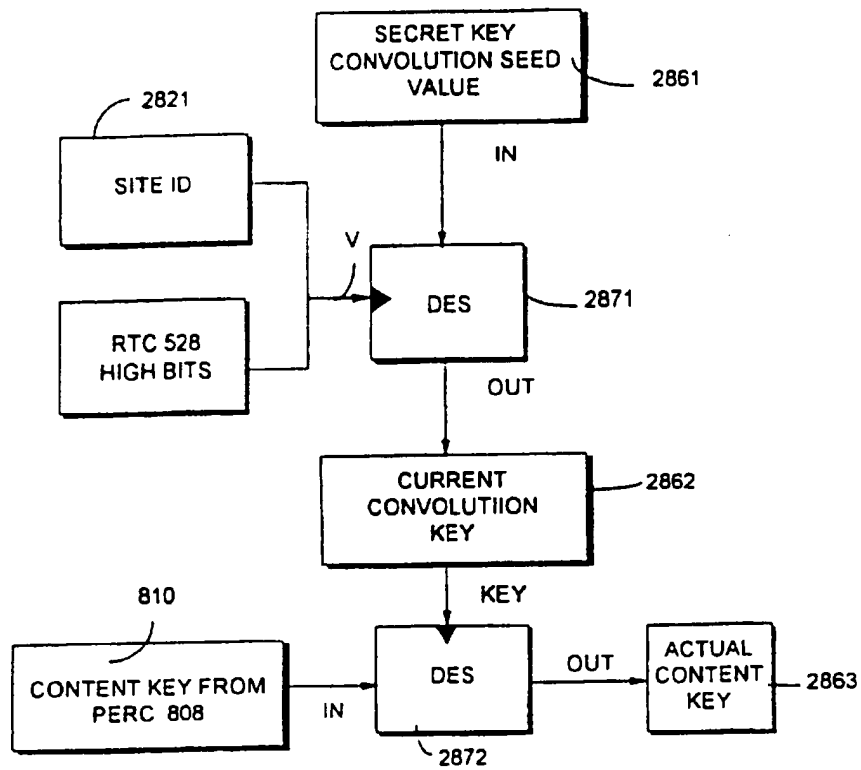


SUBSTITUTE SHEET (RULE 26)

Figure 61

116/163

FIG. 62
KEY CONVOLUTION PROCESS



SUBSTITUTE SHEET (RULE 26)

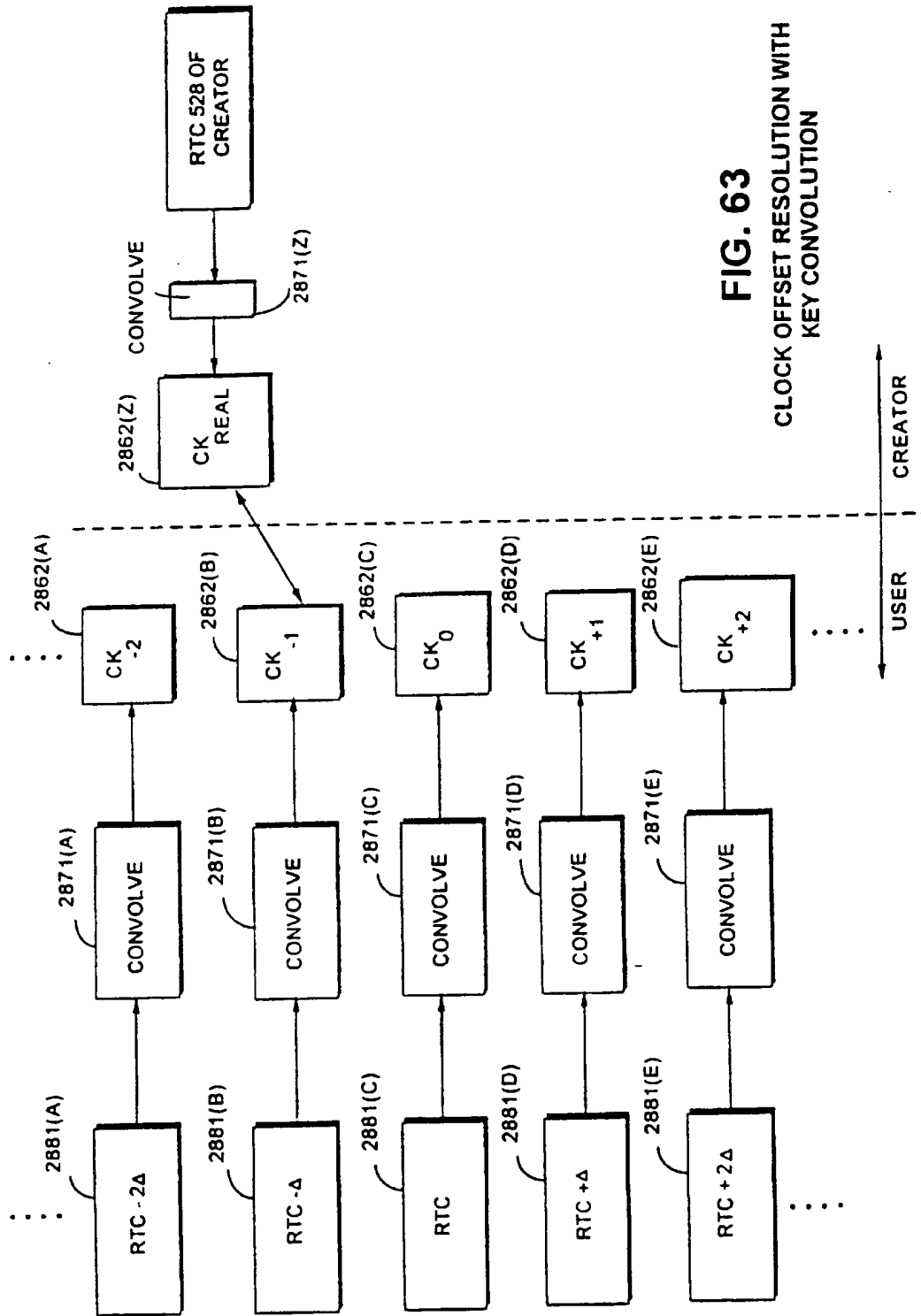
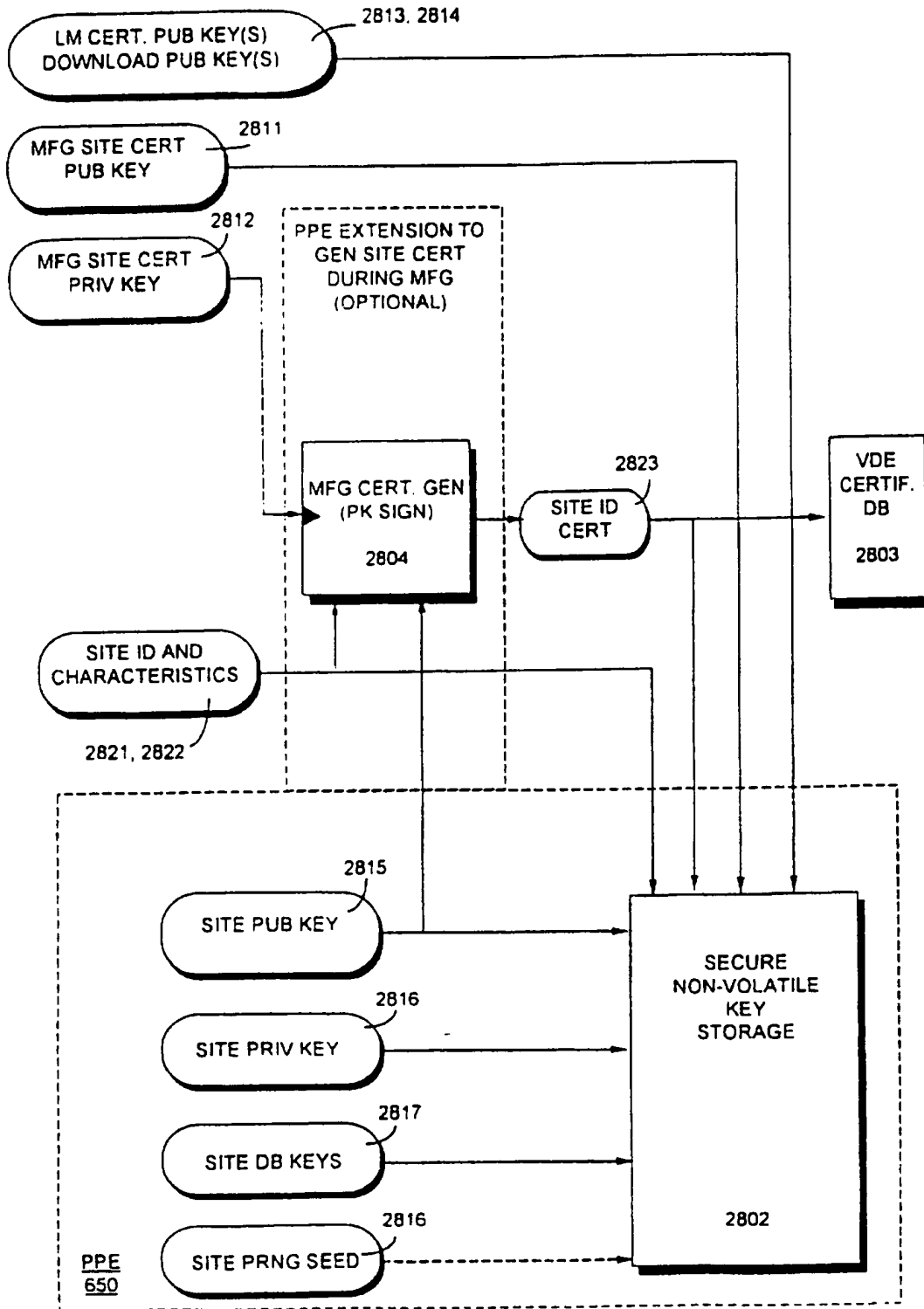


FIG. 63

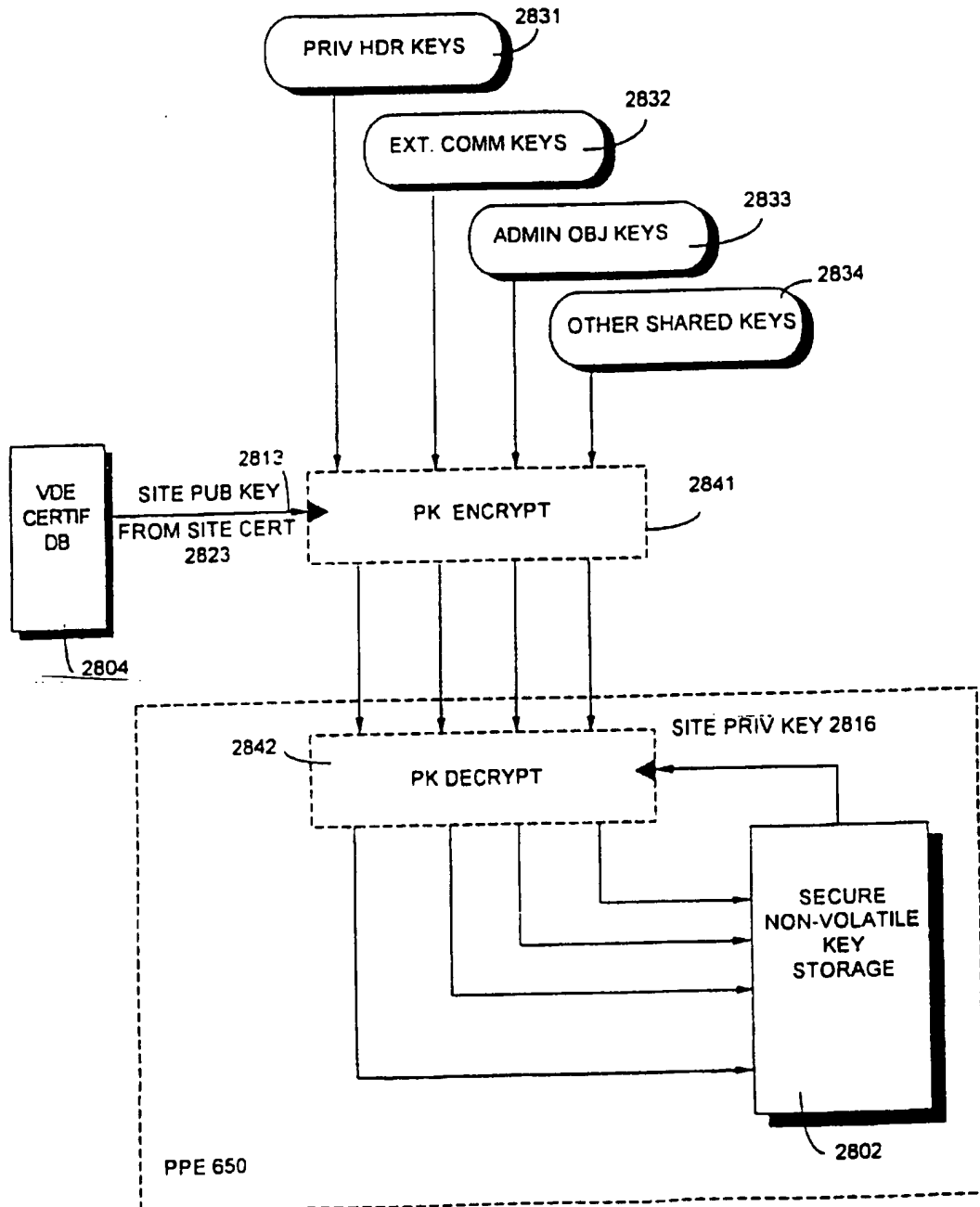
CLOCK OFFSET RESOLUTION WITH KEY CONVOLUTION

FIG. 64 SPU KEY INITIALIZATION/INSTALLATION



SUBSTITUTE SHEET (RULE 26)

FIG. 65 KEY INSTALLATION & UPDATE



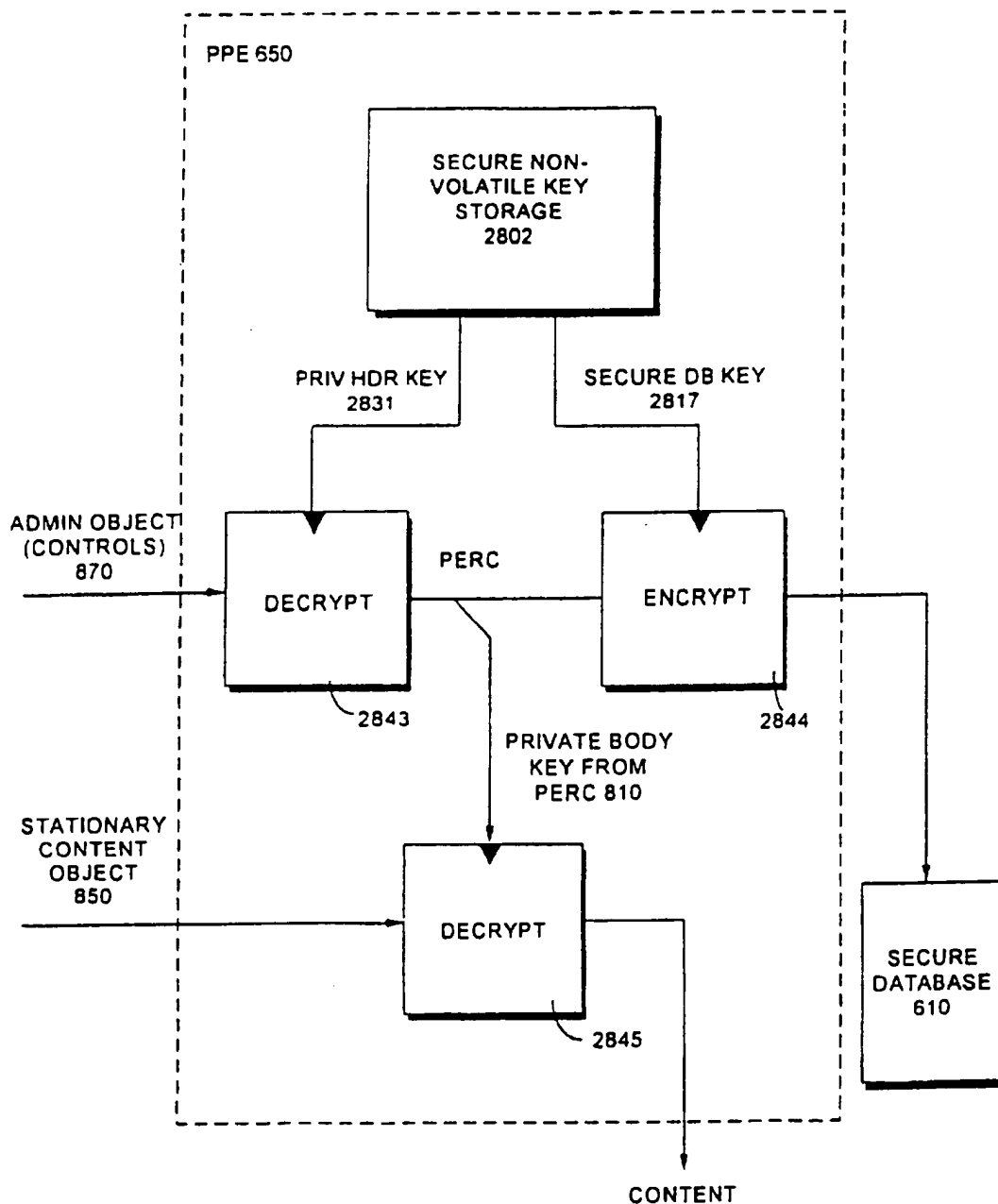


FIG. 66 STATIONARY OBJECT DECRYPTION

SUBSTITUTE SHEET (RULE 26)

121/163

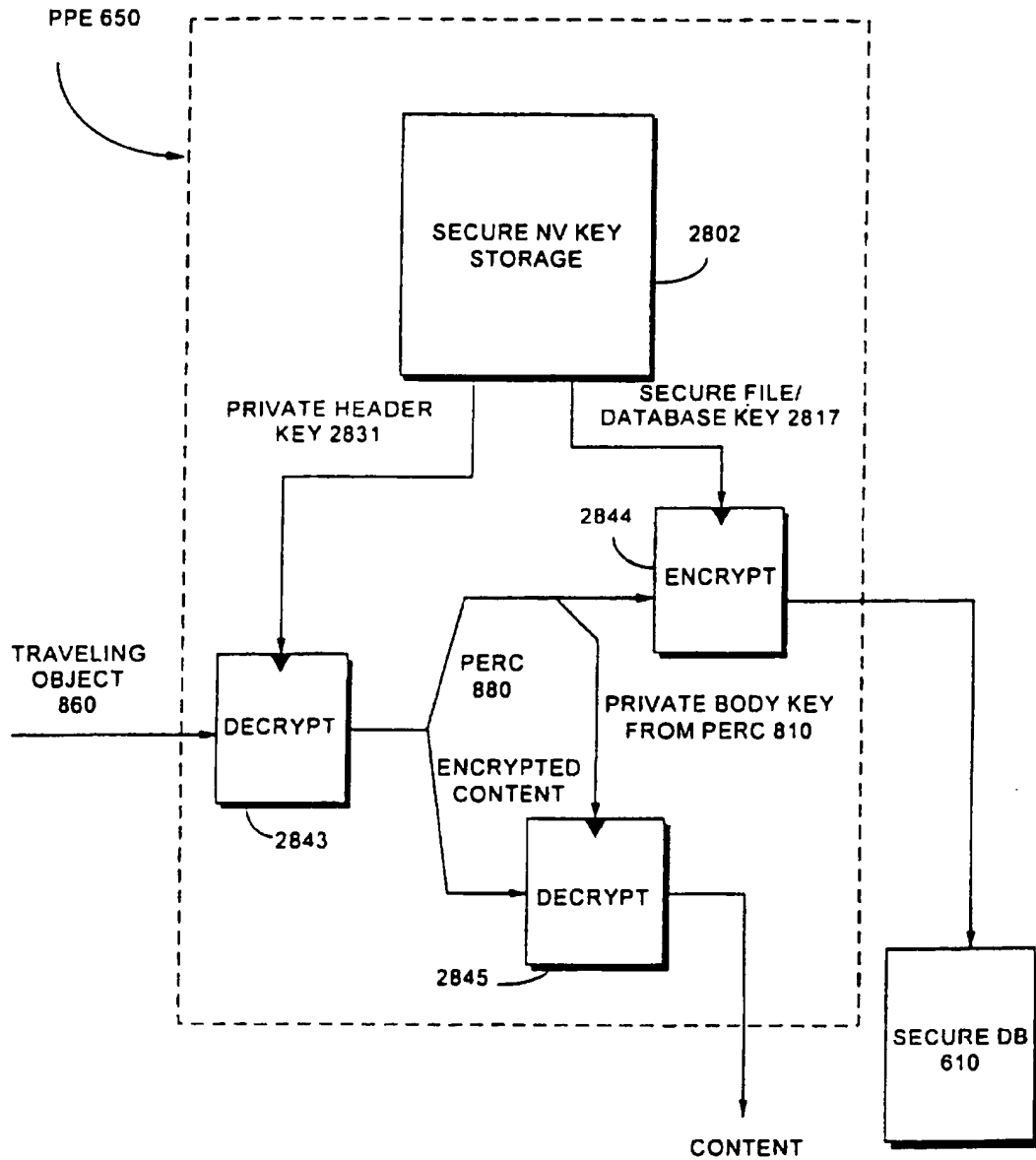


FIG. 67 TRAVELING OBJECT DECRYPTION

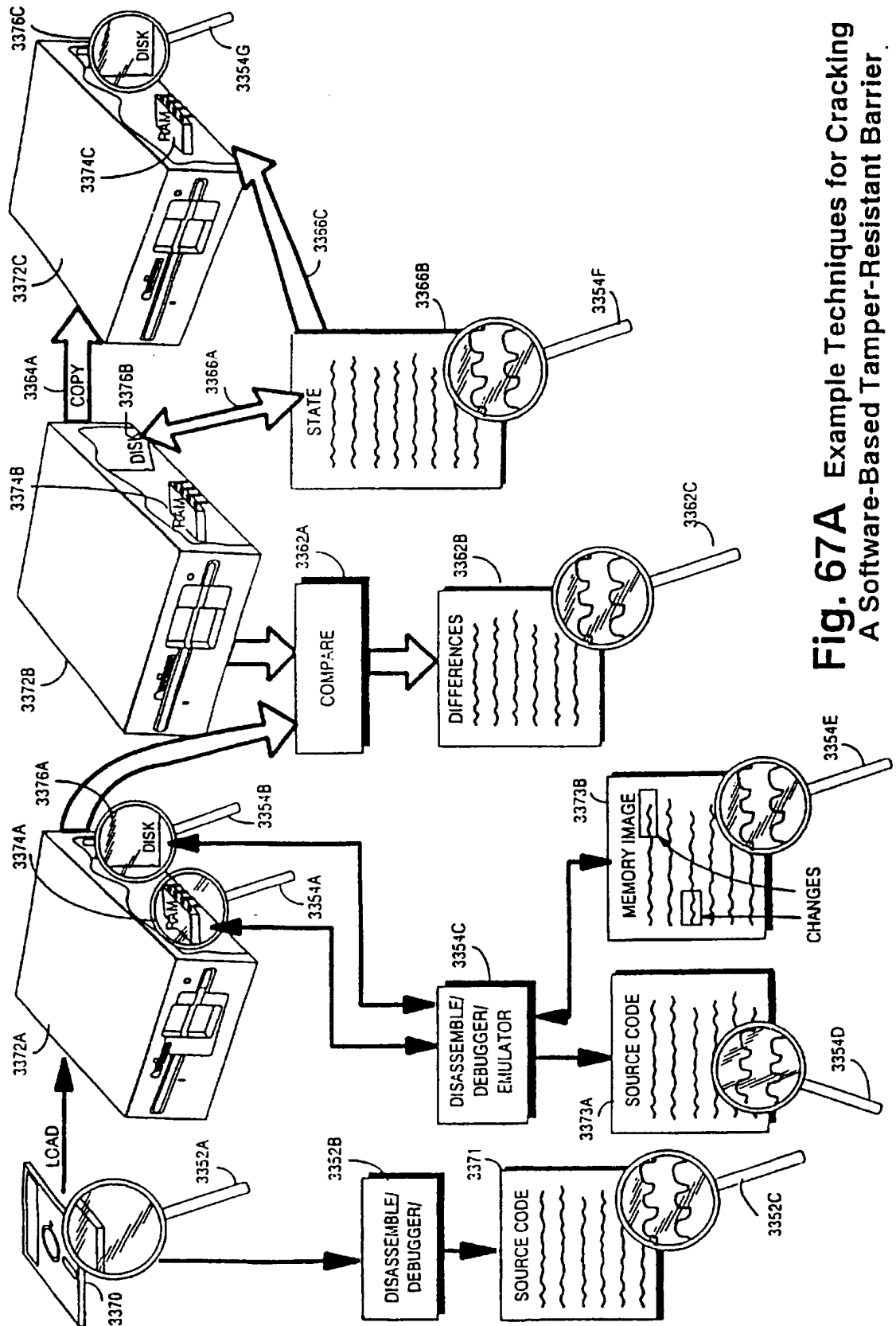
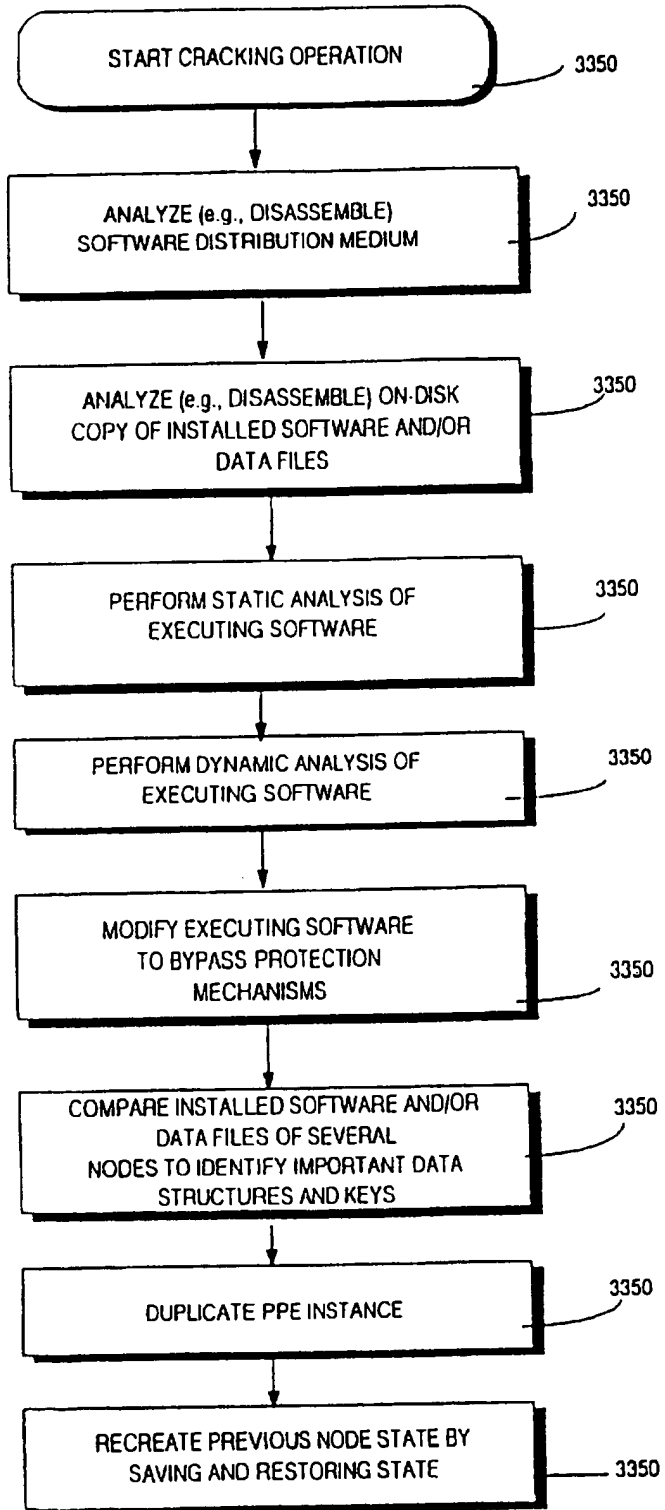


Fig. 67A Example Techniques for Cracking A Software-Based Tamper-Resistant Barrier

123/163

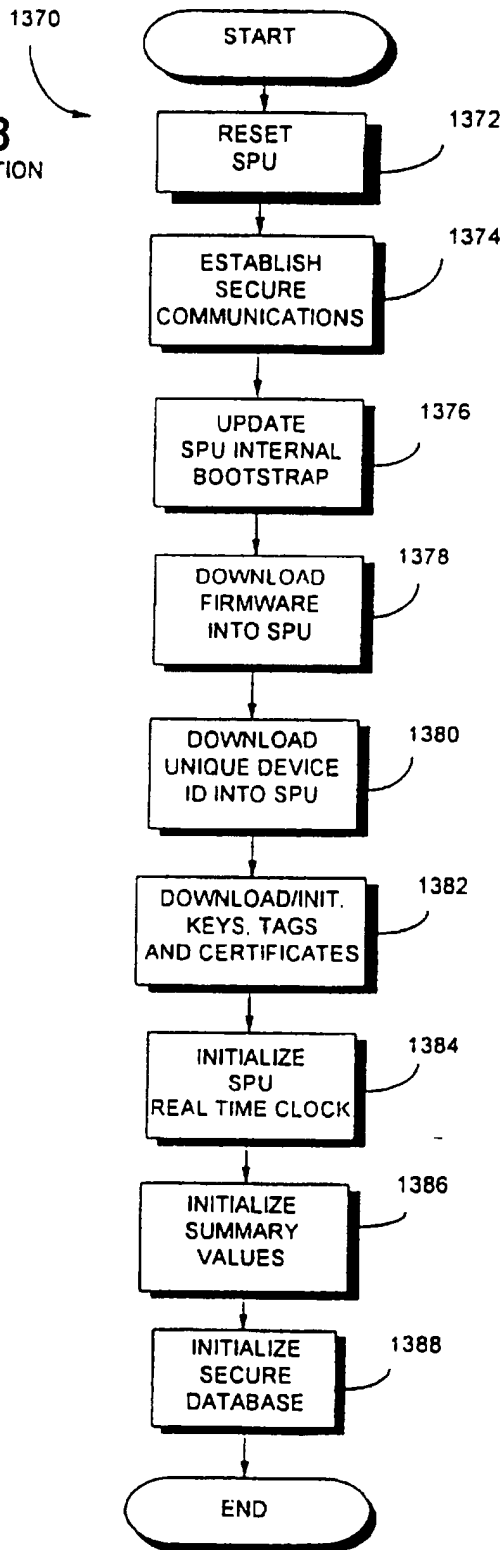
Fig. 67B
Example Cracking
Operation



SUBSTITUTE SHEET (RULE 26)

124/163

FIG. 68
SPU INITIALIZATION



SUBSTITUTE SHEET (RULE 26)

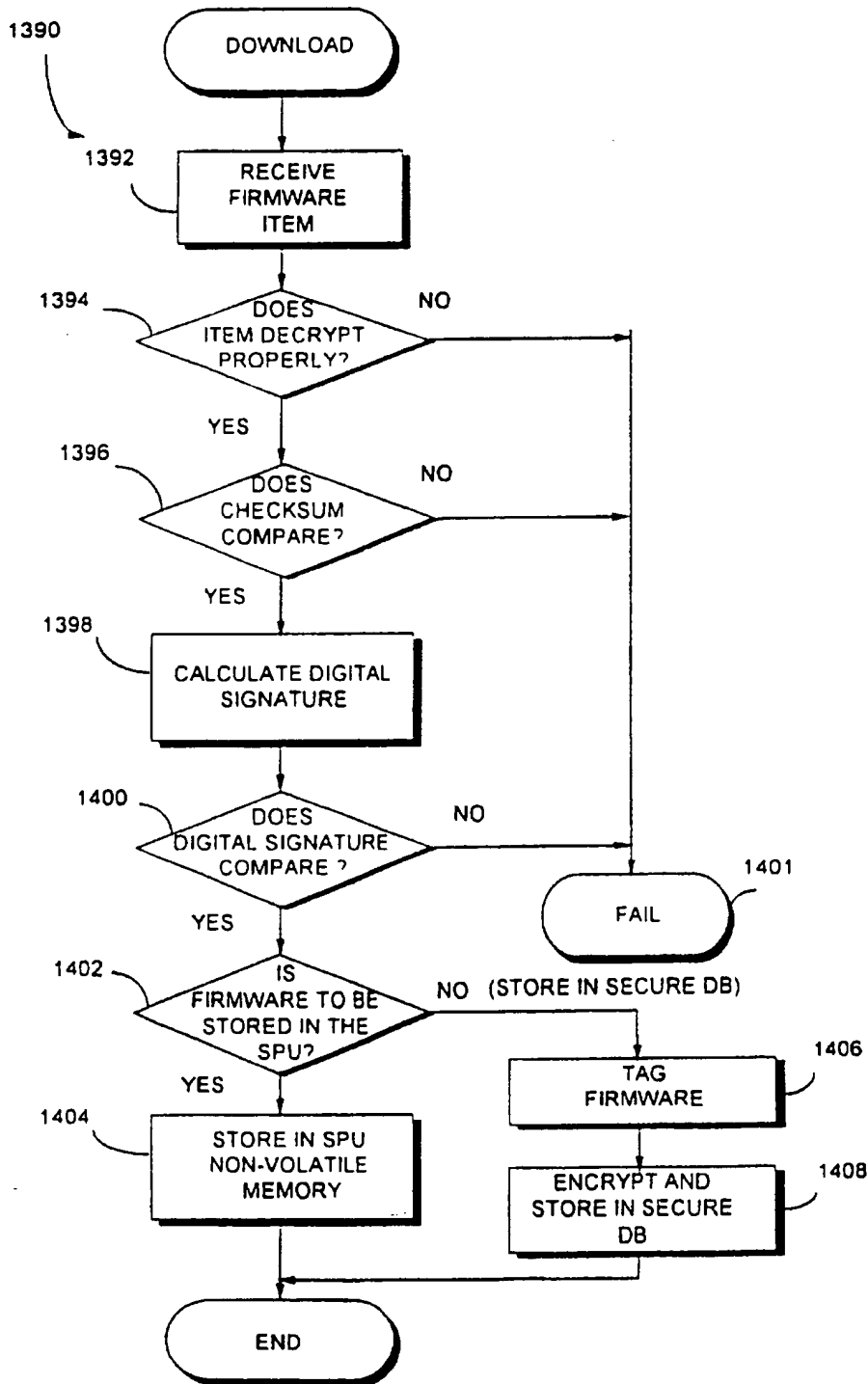
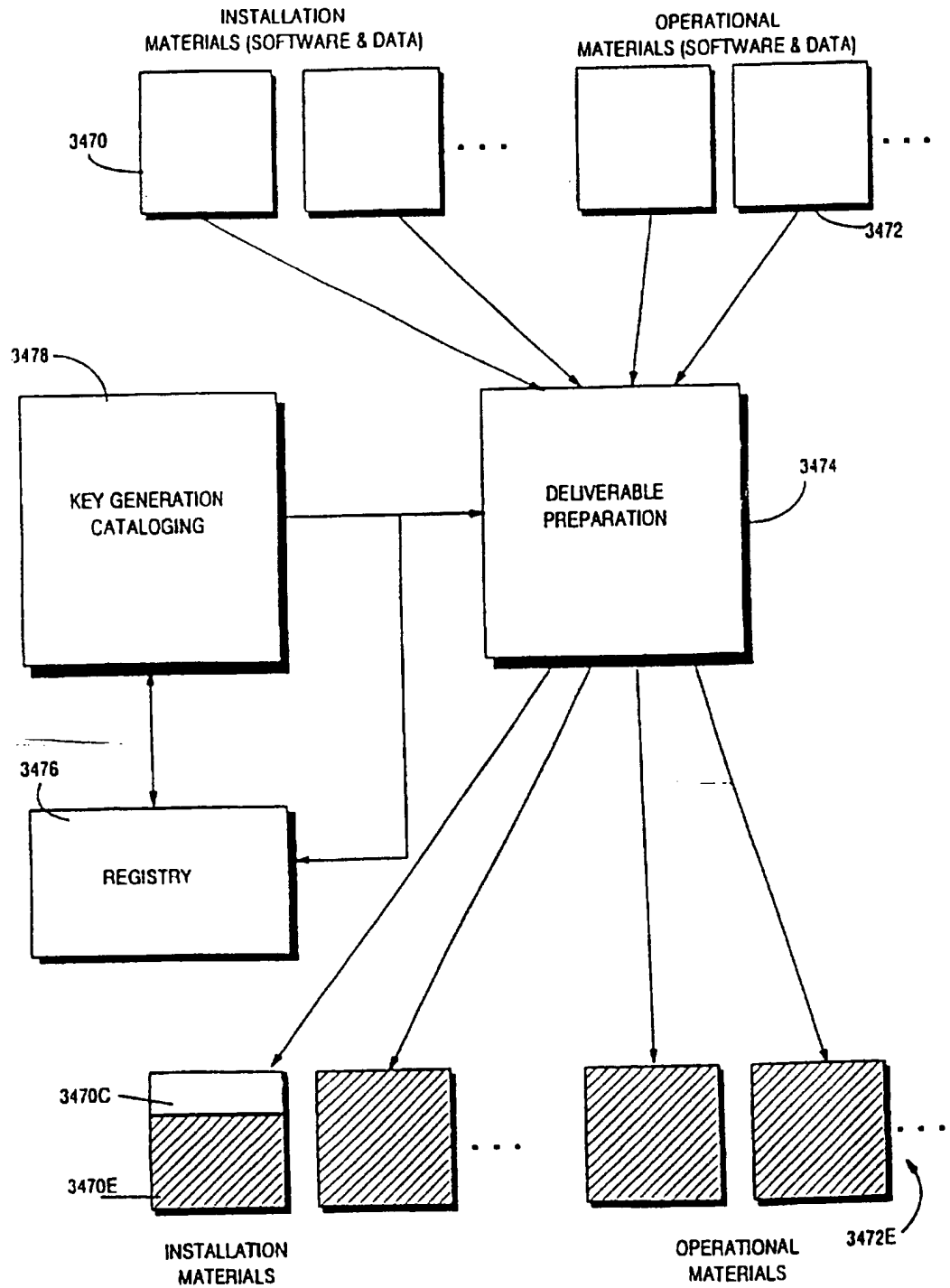


FIG. 69

SPU FIRMWARE
DOWNLOAD

SUBSTITUTE SHEET (RULE 26)

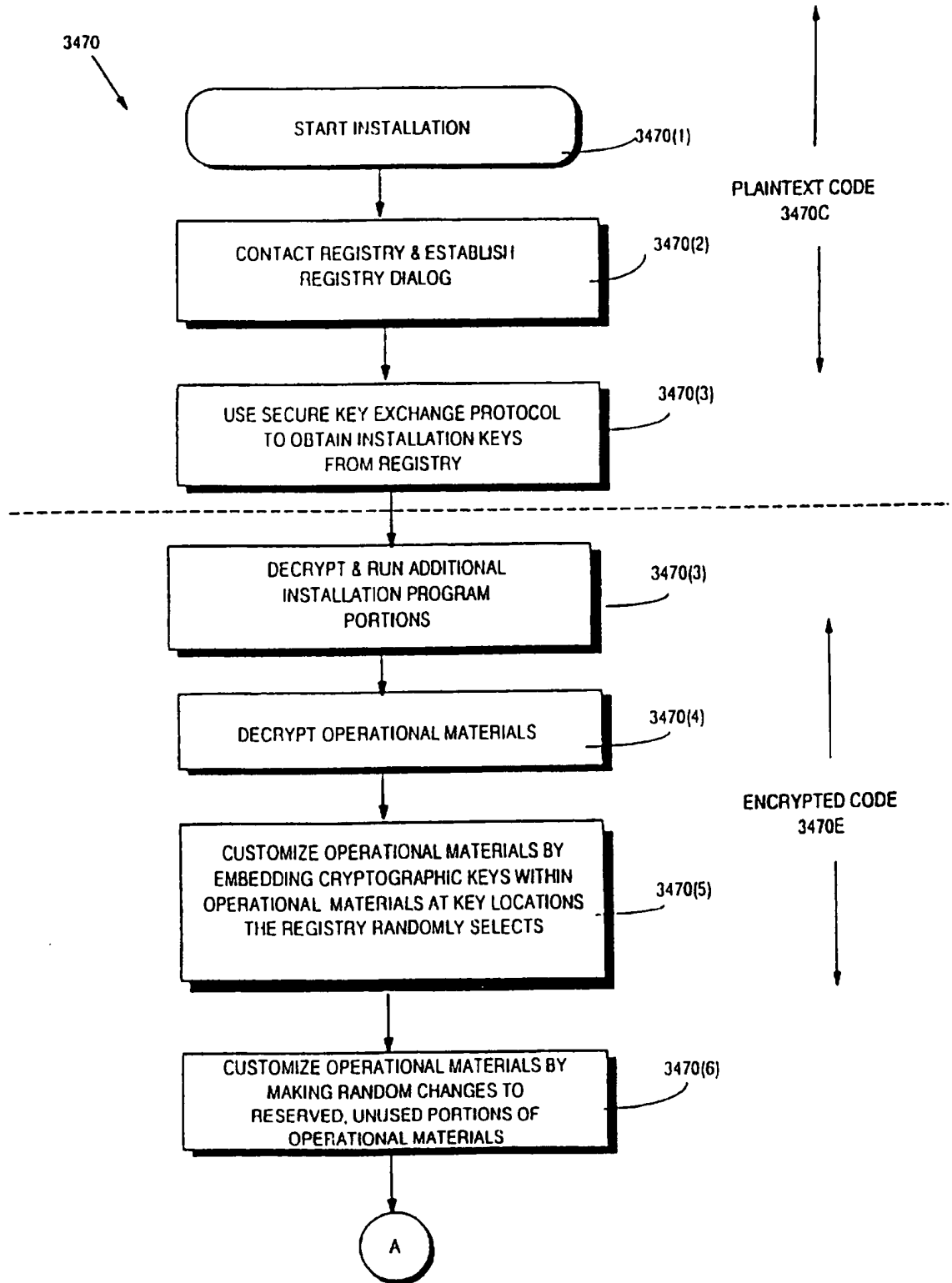
Fig. 69A Software Distributed In Encrypted Form



SUBSTITUTE SHEET (RULE 26)

127/163

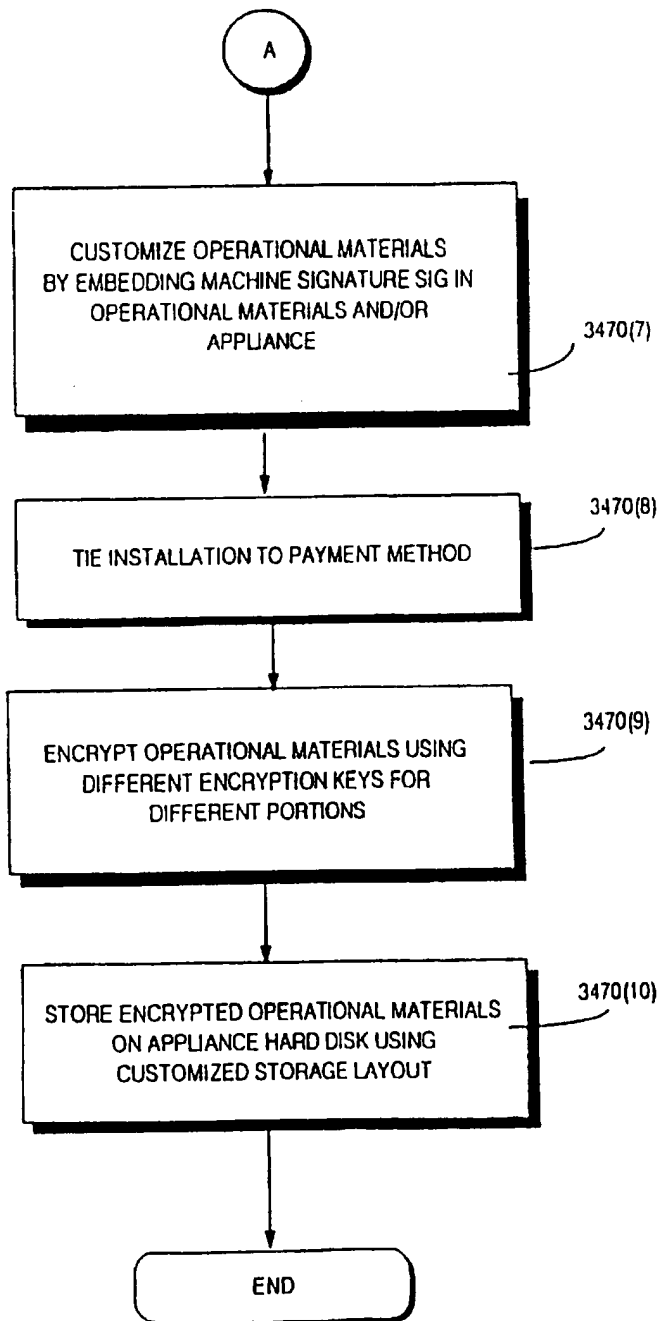
Fig. 69B Example Installation Routine



SUBSTITUTE SHEET (RULE 26)

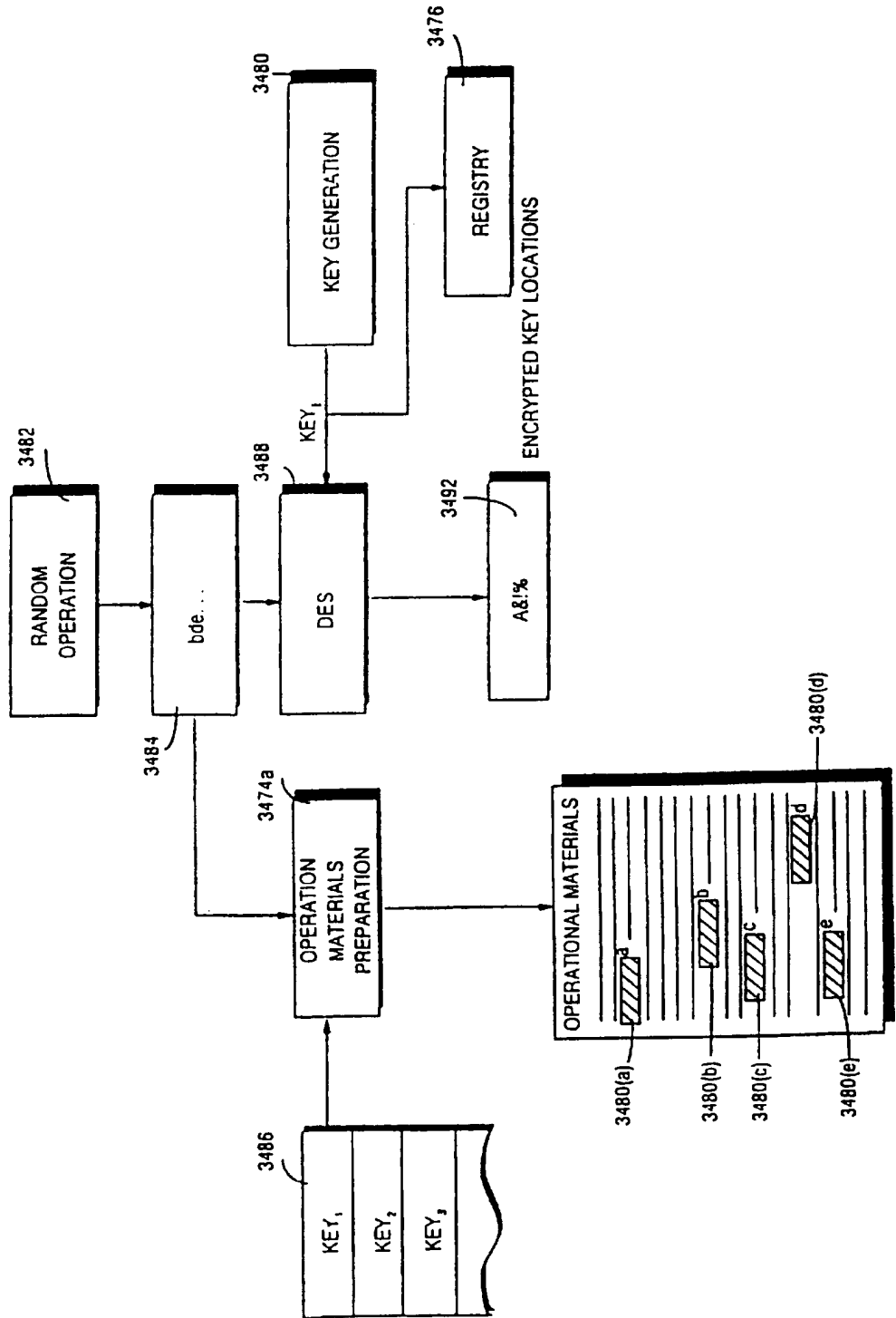
128/163

Fig. 69C Example Installation Routine



SUBSTITUTE SHEET (RULE 26)

Fig. 69D Embedding Keys At Different Locations With Operational Materials



SUBSTITUTE SHEET (RULE 26)

130/163

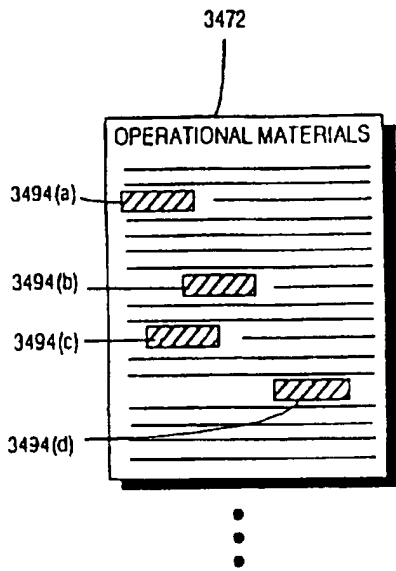
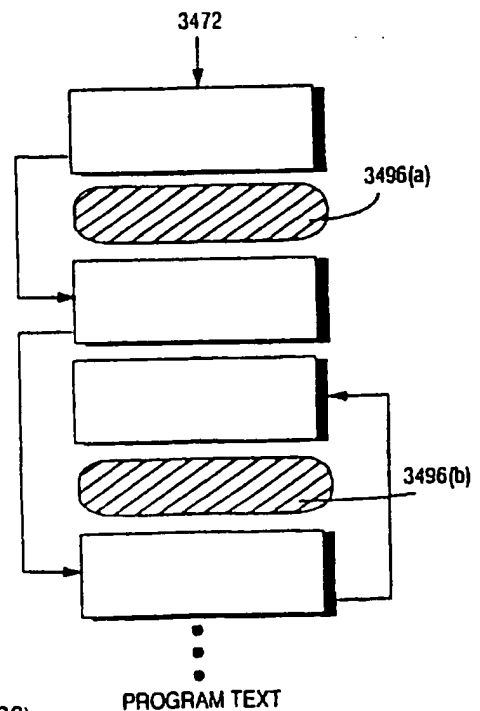


Fig. 69E
Example Locations For Random
Modifications and Fingerprints

Fig. 69F
Example Customized Static
Storage Layout



SUBSTITUTE SHEET (RULE 26)

Fig. 69G
Example Machine Signature
Locations

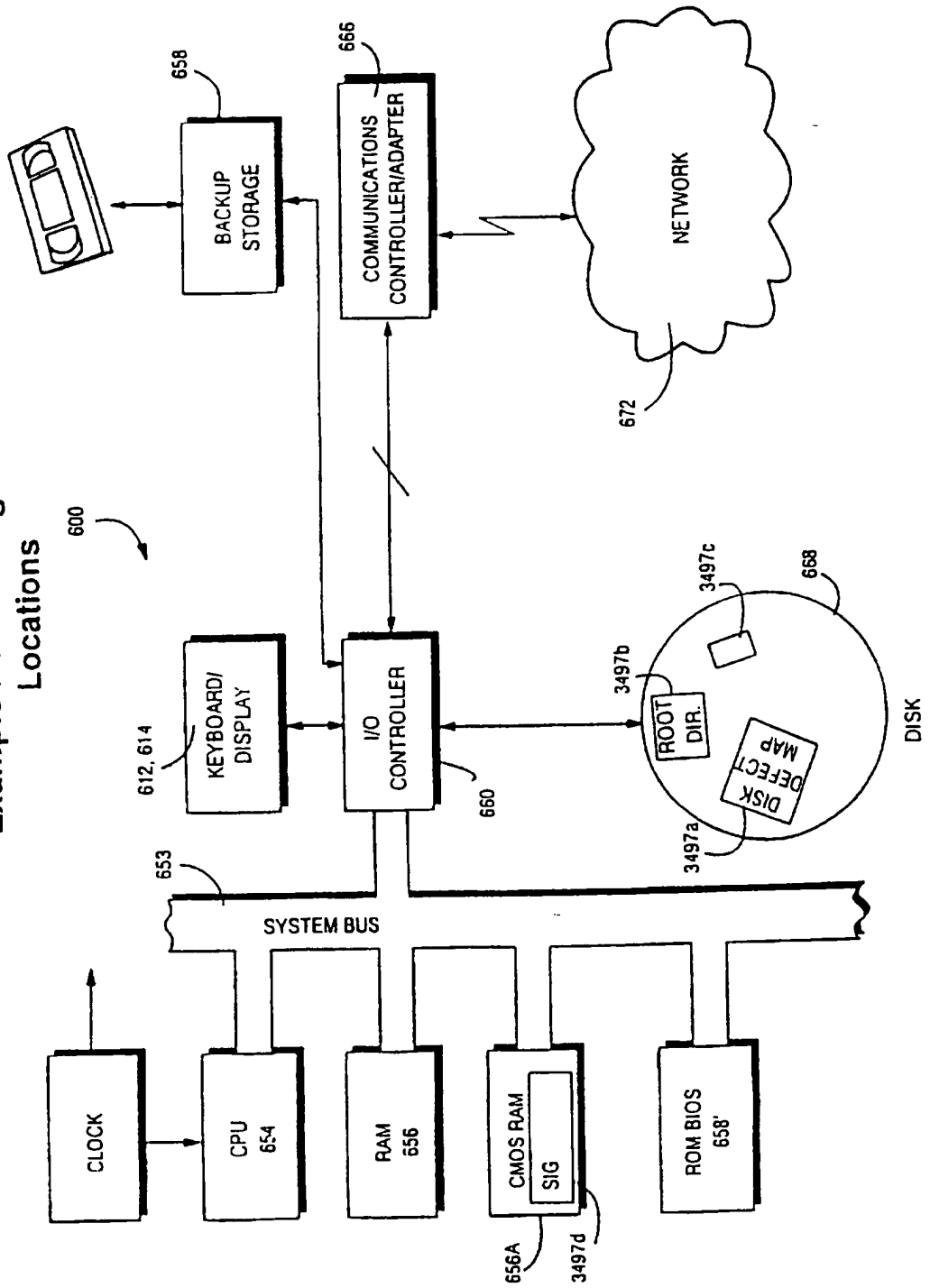


Fig. 69H
Sequence Dependent and Independent Processes

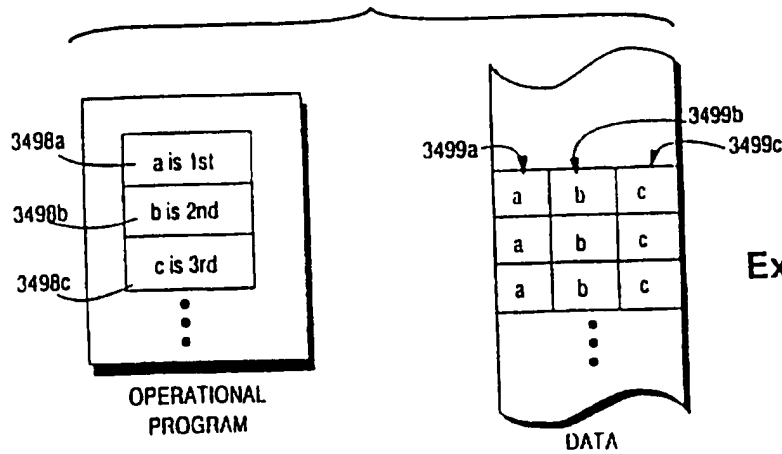
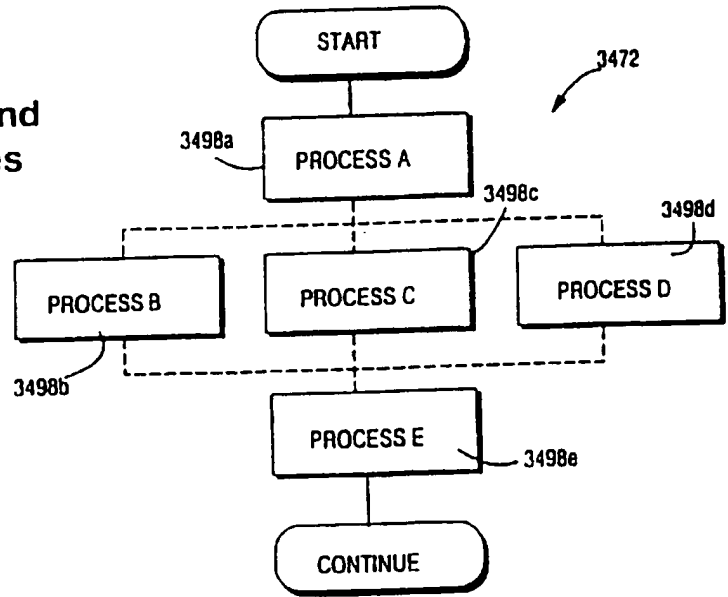


Fig. 69I
Example First Order

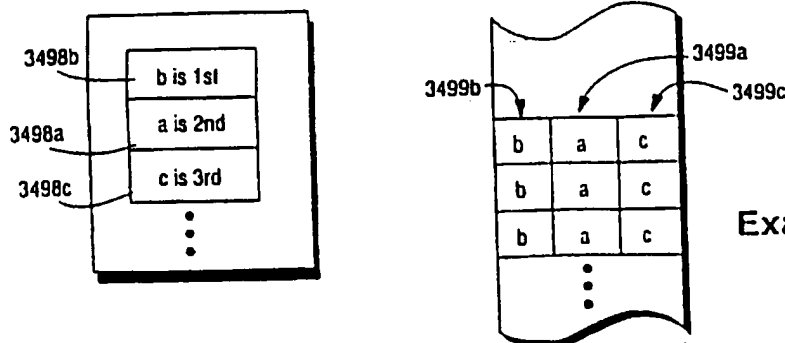


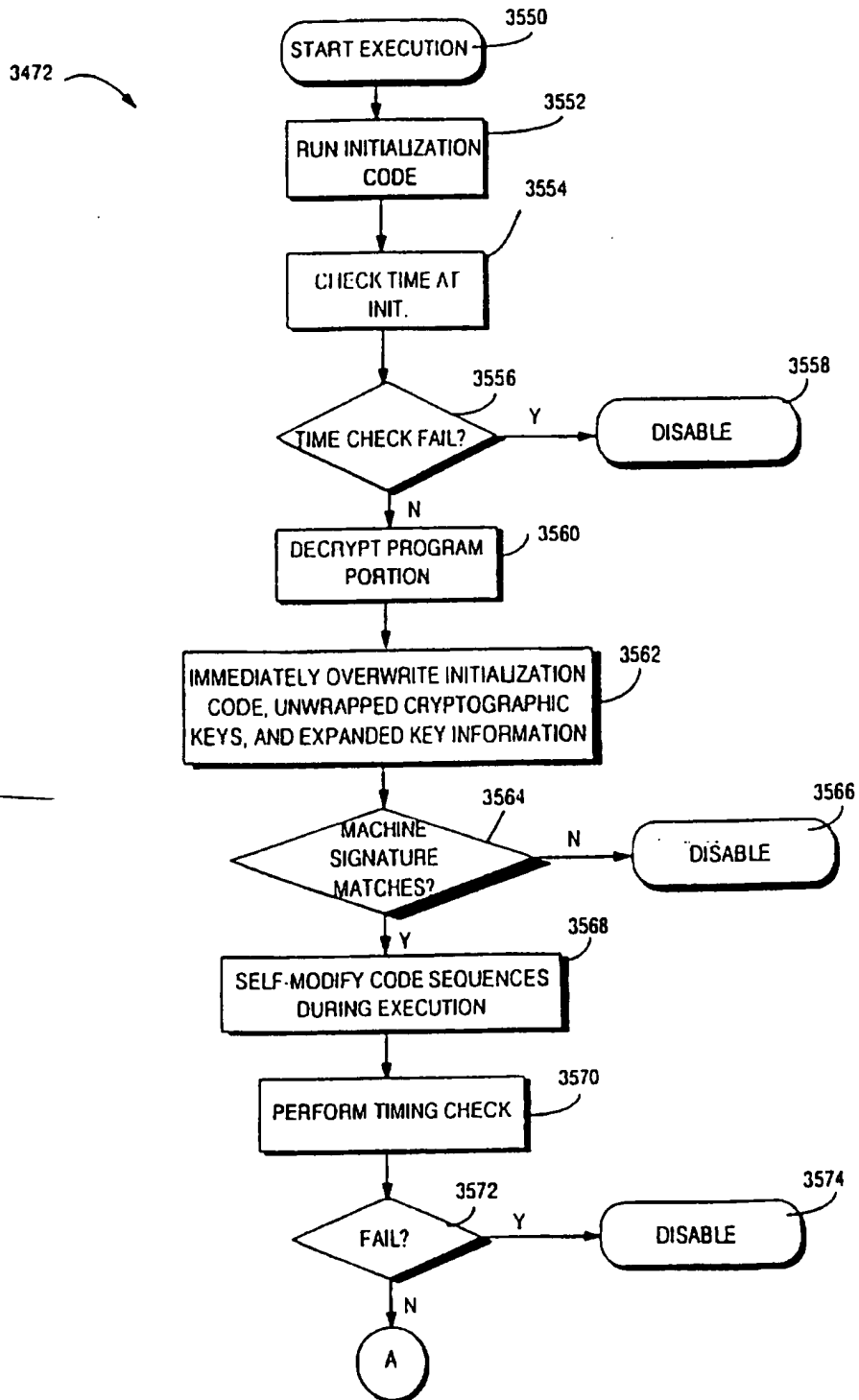
Fig. 69J
Example Second Order

SUBSTITUTE SHEET (RULE 26)

133/163

Fig. 69K

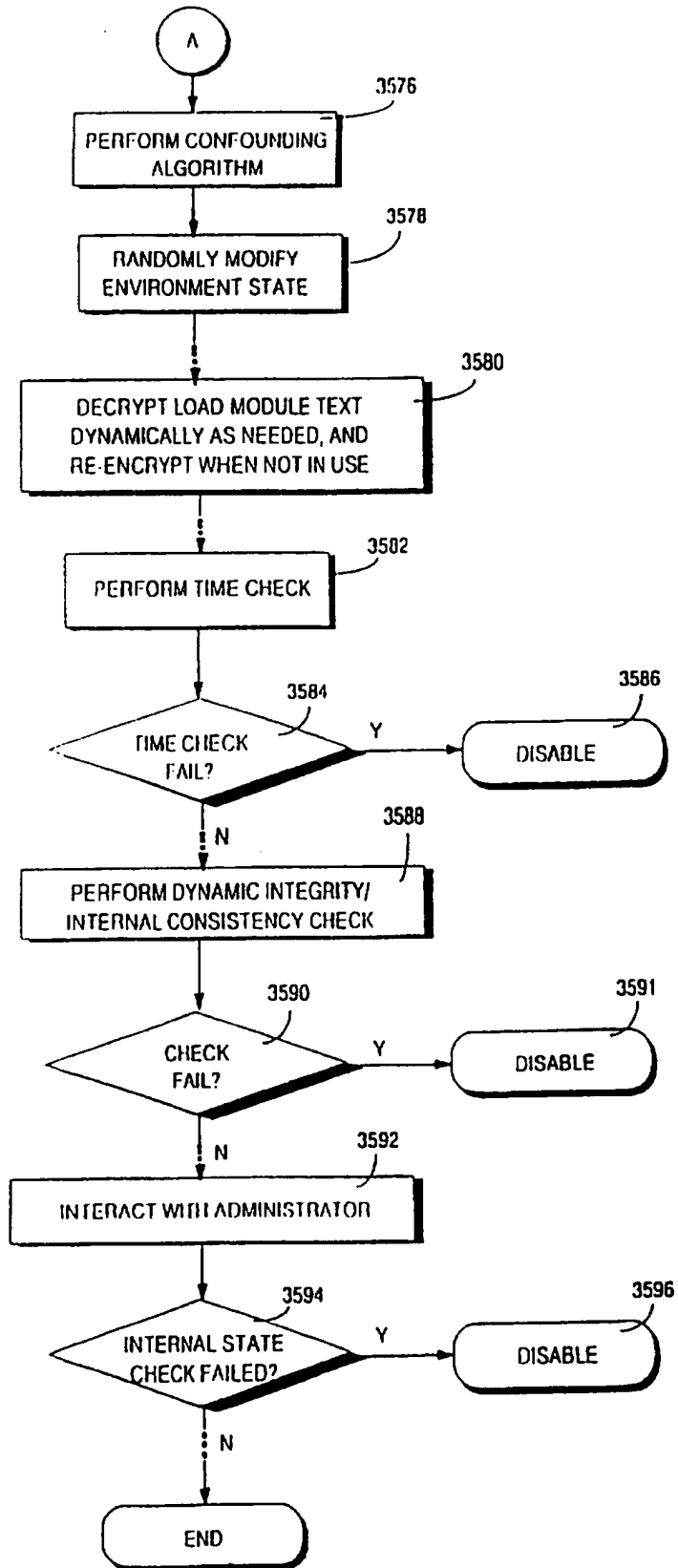
Example Dynamic Protection Mechanisms



SUBSTITUTE SHEET (RULE 26)

134/163

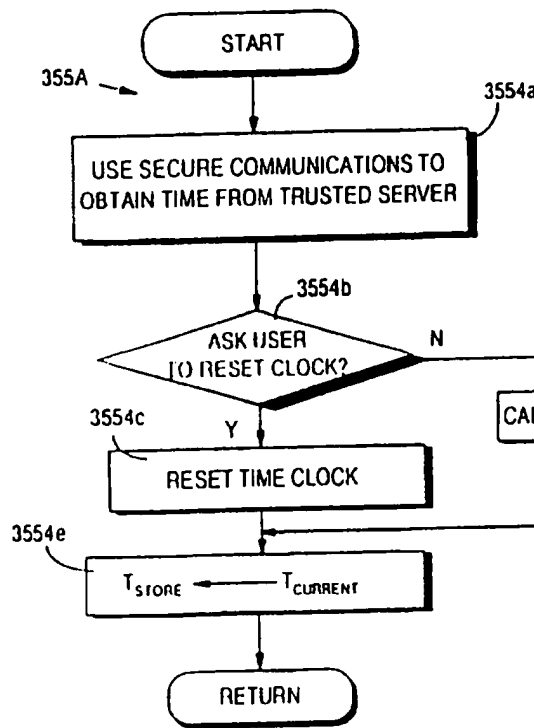
Fig. 69L
Example Dynamic
Protection
Mechanisms



SUBSTITUTE SHEET (RULE 26)

135/163

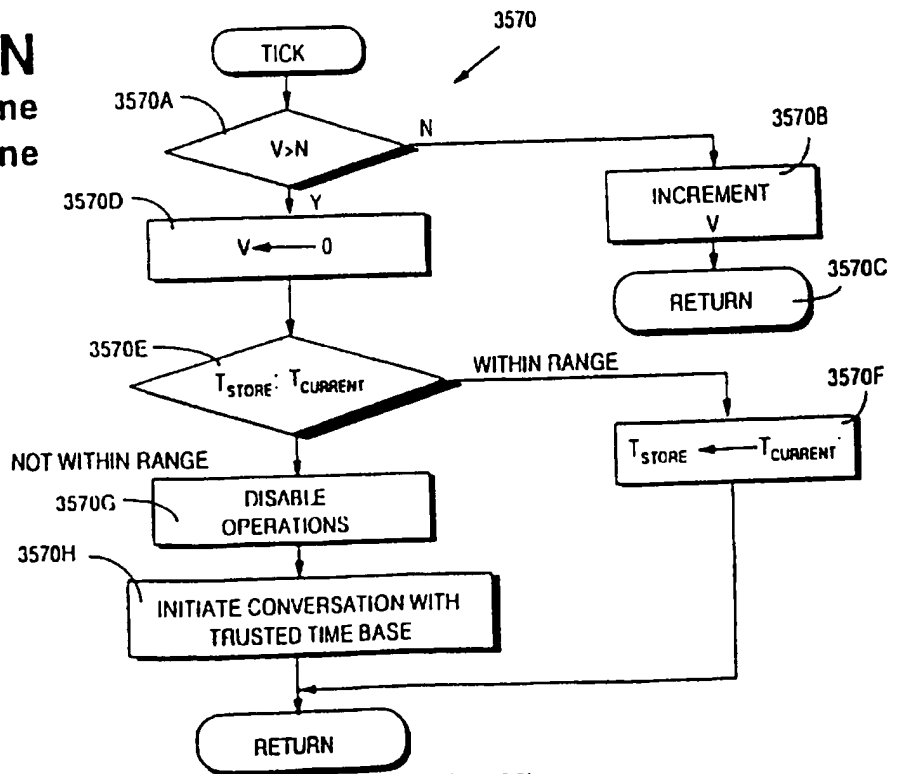
Fig. 69M
Example Time
Check At
Initialization
Routine



DRAFT BUDGET
AMT DRAFTED
ΔT
T _{STORE}

Fig. 69O

Fig. 69N
Example Time
Check Routine



SUBSTITUTE SHEET (RULE 26)

136/163

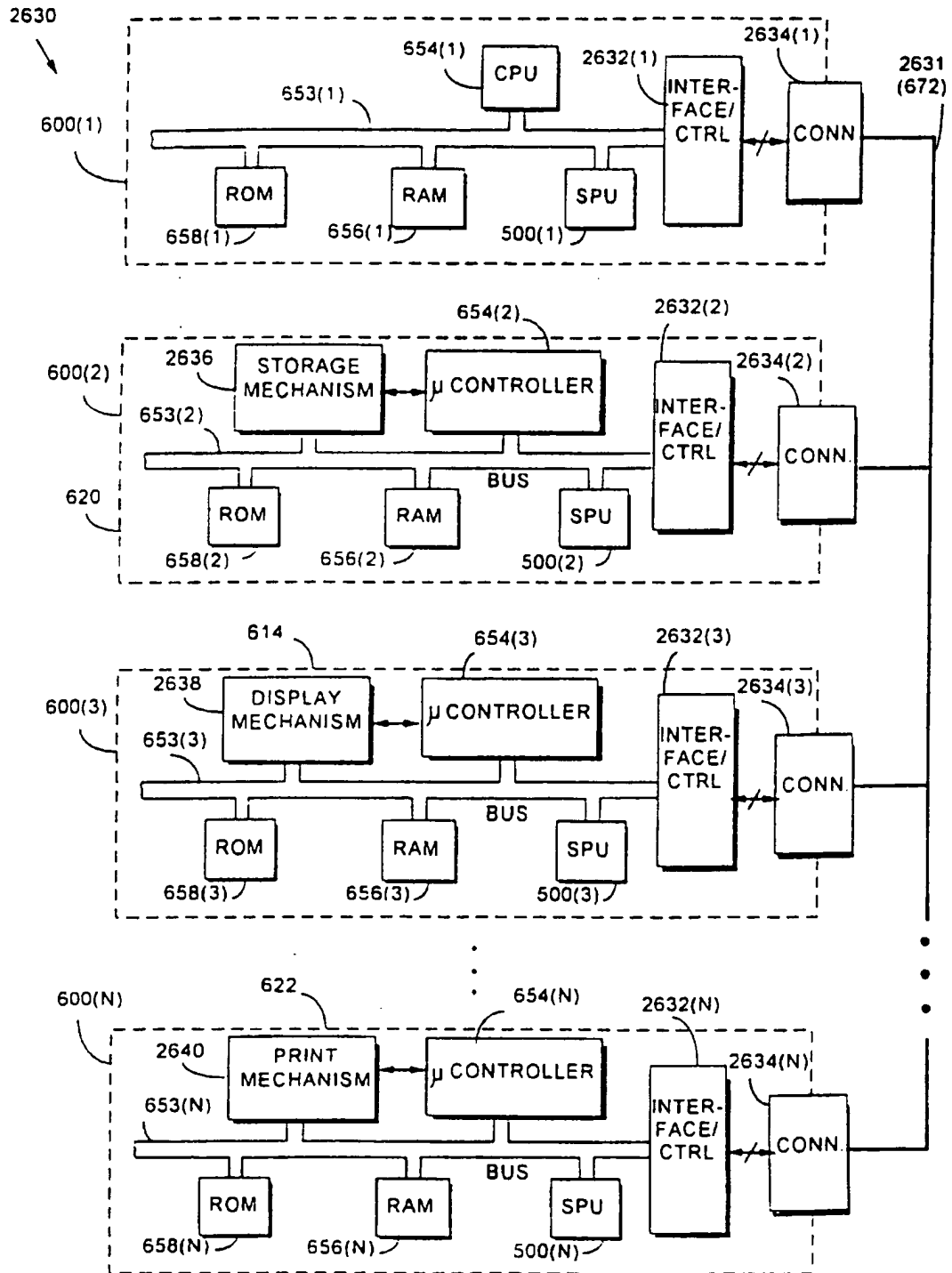
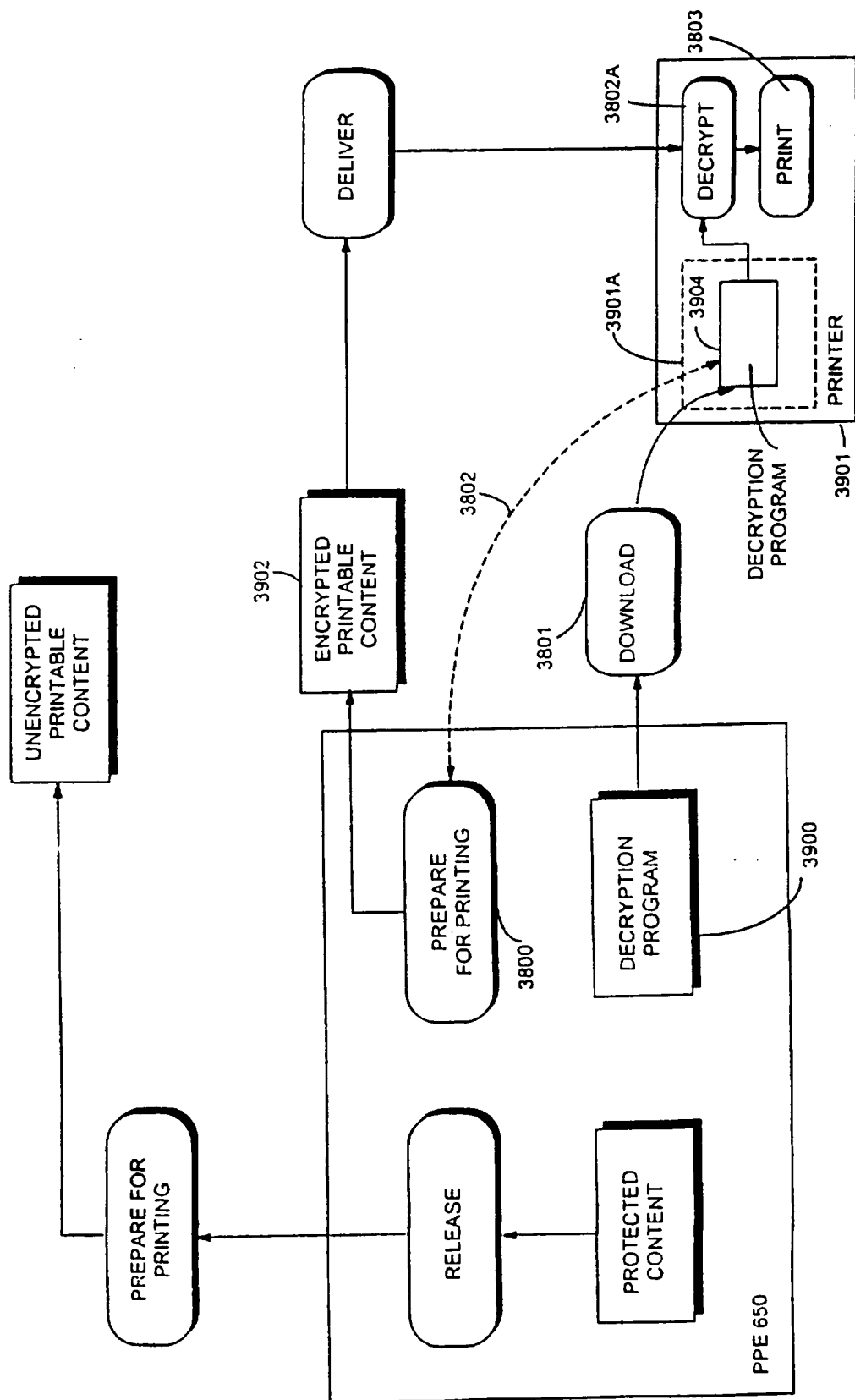


FIG. 70

SUBSTITUTE SHEET (RULE 26)

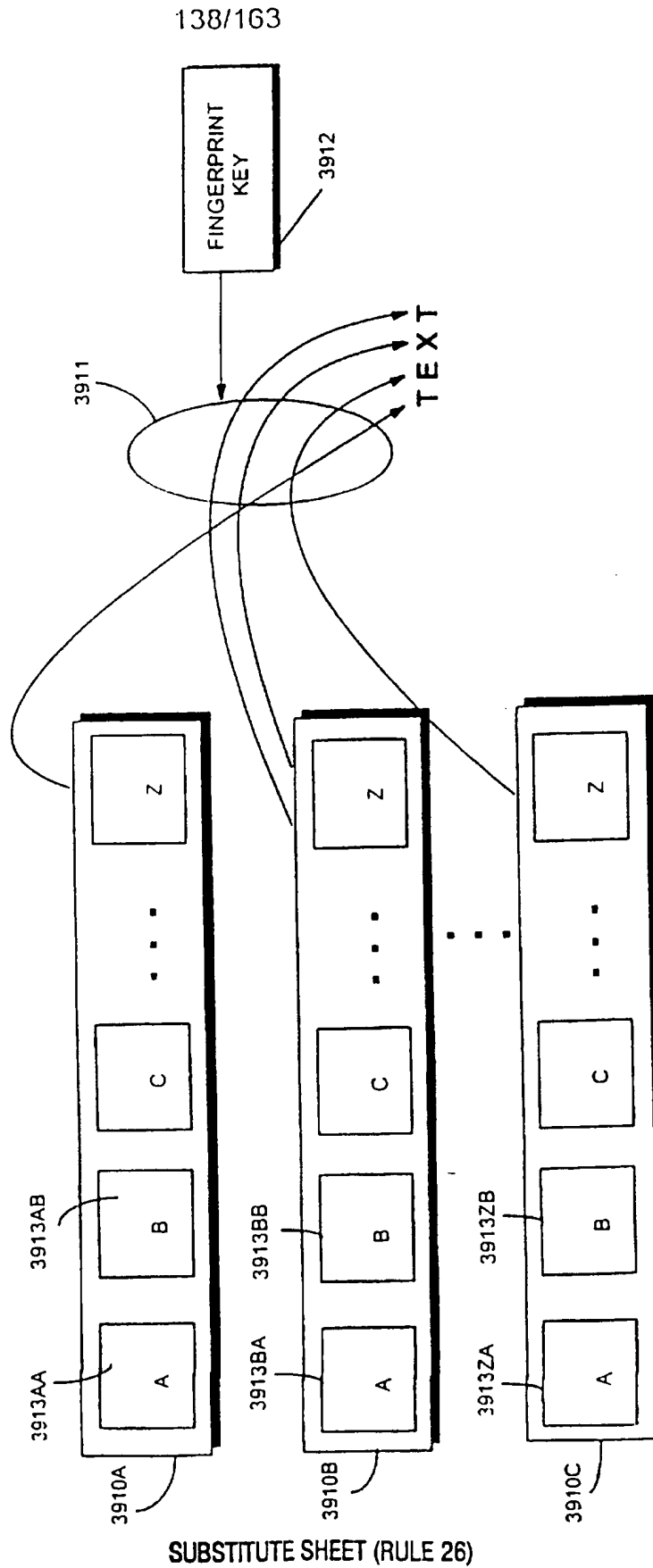
137/163

FIG. 70A



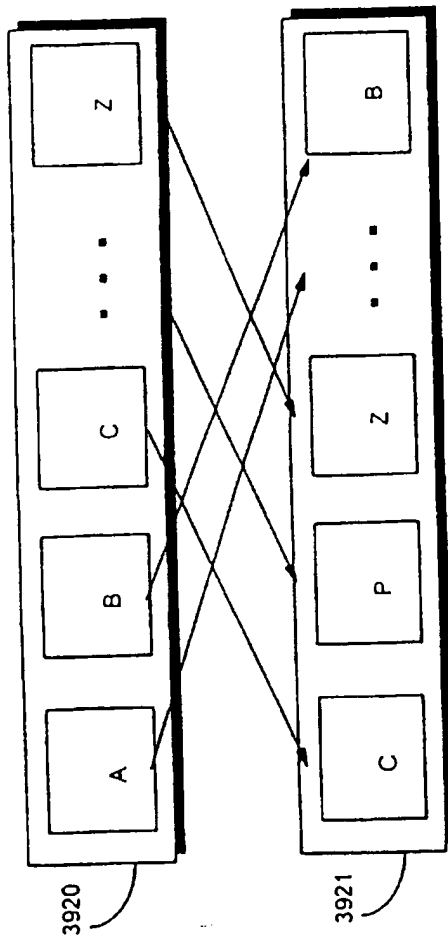
SUBSTITUTE SHEET (RULE 26)

FIG. 70B



139/163

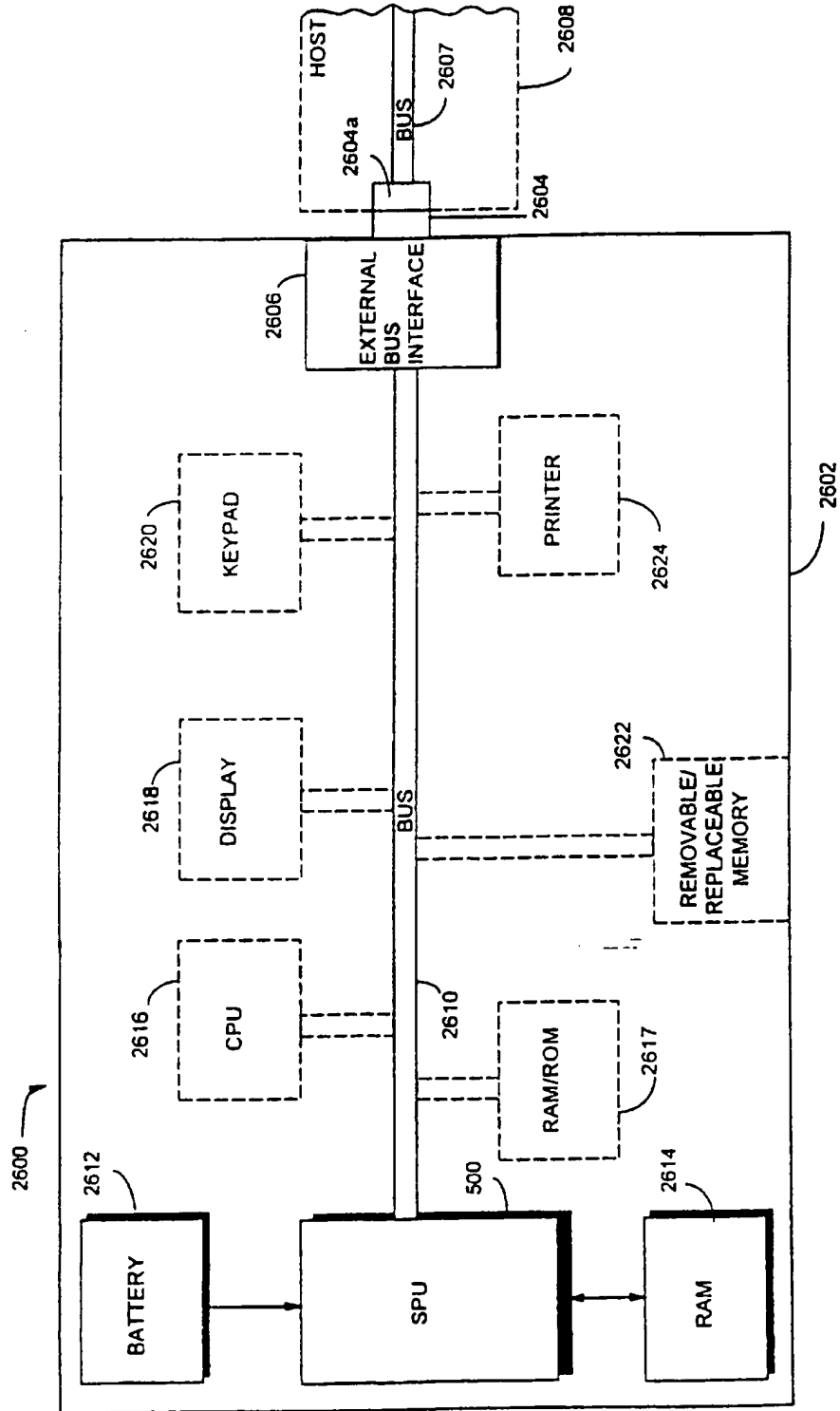
FIG. 70C



SUBSTITUTE SHEET (RULE 26)

140/163

FIG. 71
PORTABLE APPLIANCE



SUBSTITUTE SHEET (RULE 26)

141/163



LOG IN USER INTERFACE 182

USER NAME:	<input type="text" value="SHEAR. V."/>	<input type="button" value="LOGIN"/>
PASSWORD:	<input type="password" value="*****"/>	<input type="button" value="CANCEL"/>
<input type="checkbox"/> LOGIN AT STARTUP		<input type="button" value="HELP"/>

FIG. 72A

FIG. 72B

2660

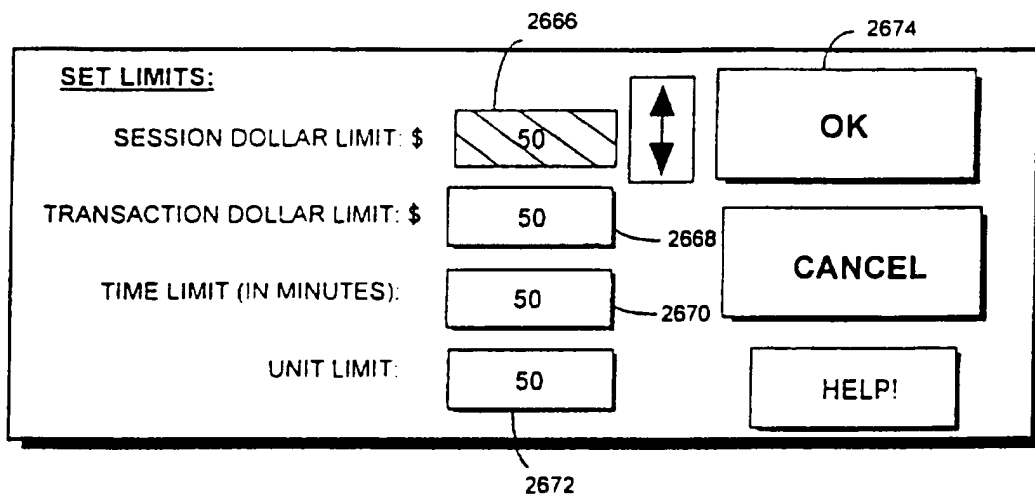
	YOU HAVE REQUESTED THESE PROPERTIES:	<input type="button" value="CANCEL"/>
<u>LOONEY TUNES NEWS!</u>	<input type="button" value="APPROVE"/> 2662	<input type="button" value="SUSPEND"/>
<input type="button" value="PROPERTY INFO"/>	Your Cost: \$7.50	MORE OPTIONS 

2664

SUBSTITUTE SHEET (RULE 26)

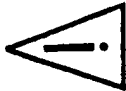
142/163

FIG. 72C



SUBSTITUTE SHEET (RULE 26)

FIG. 72D



YOU HAVE REQUESTED THESE PROPERTIES:

LOONEY TUNE NEWS!

CANCEL

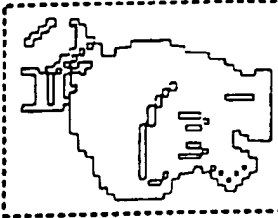
APPROVE

SUSPEND

PROPERTY INFO

YOUR COST : \$7.50

More Options



Show Thumbnail

PROPERTY:	SIZE:	PUBLISHER:	AMOUNT:	UNITS:	COST/UNIT:	TYPE:	USE?:	LINKS:	HIST:
CHUCK JONES BIOGRA...	256KB	WARNER NEW MEDIA	64	KBYTE	\$1.25	PREVIEW	<input checked="" type="checkbox"/>		<input type="checkbox"/>
▼ BUGS BUNNY.JPE...	1MB	WARNER NEW MEDIA	1	RECORD	\$5.00	DISPLAY	<input checked="" type="checkbox"/>		<input type="checkbox"/>
BUGS BUNNY.JPEG...	1MB	WARNER NEW MEDIA	10	RECORD	\$3.50	DISPLAY	<input type="checkbox"/>		<input type="checkbox"/>
BUGS BUNNY.JPEG...	1MB	WARNER NEW MEDIA	25	RECORD	\$2.50	DISPLAY	<input type="checkbox"/>		<input type="checkbox"/>
FRIZ FRELENG BIOGRA...	256KB	WARNER NEW MEDIA	120	SECTOR	\$5.00	PRINT	<input type="checkbox"/>		<input type="checkbox"/>
TEX AVERY BIOGRAP...	256KB	WARNER NEW MEDIA	50	PERCENT	\$2.50	COPY	<input type="checkbox"/>		<input type="checkbox"/>
▶ DUCKI RABBITI DU...	64MB	WARNER NEW MEDIA	7.0	MINUTE	\$7.50	COPY-PRO	<input type="checkbox"/>		<input type="checkbox"/>
MEL BLANC BIOGRAPH...	256KB	WARNER NEW MEDIA	1	SPECIAL	\$25.25	INSTALL	<input type="checkbox"/>		<input type="checkbox"/>
LOONEY TUNES DATA...	600MB	WARNER NEW MEDIA	1	OBJECT	\$2000.00	ALL	<input type="checkbox"/>		<input type="checkbox"/>

SET LIMITS...

SHOW BUDGETS

ACQUIRE BUDGET...

HISTORY...

TRANSFER...

PREFERENCES...

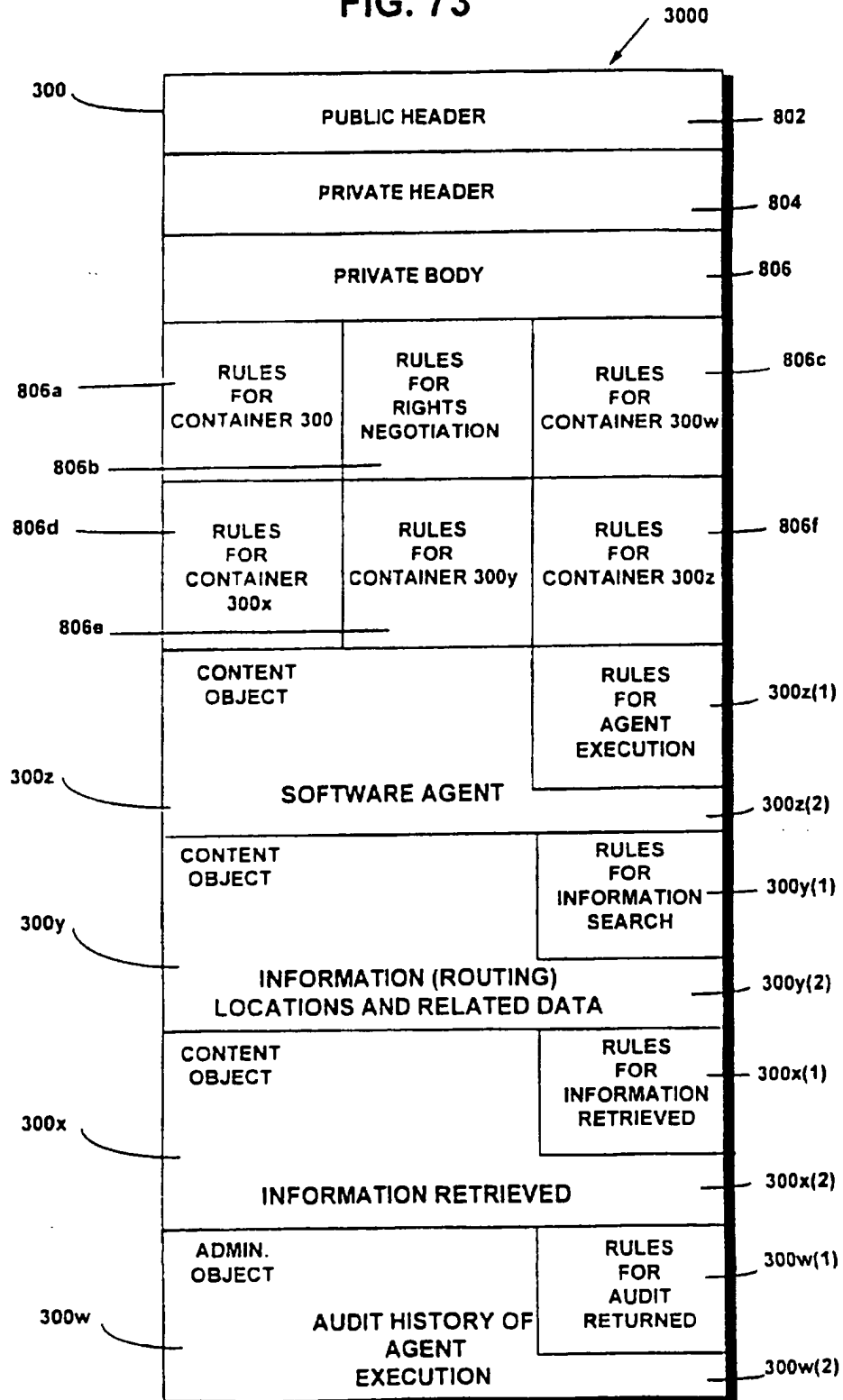
FEEDBACK...

HELP!

SUBSTITUTE SHEET (RULE 26)

144/163

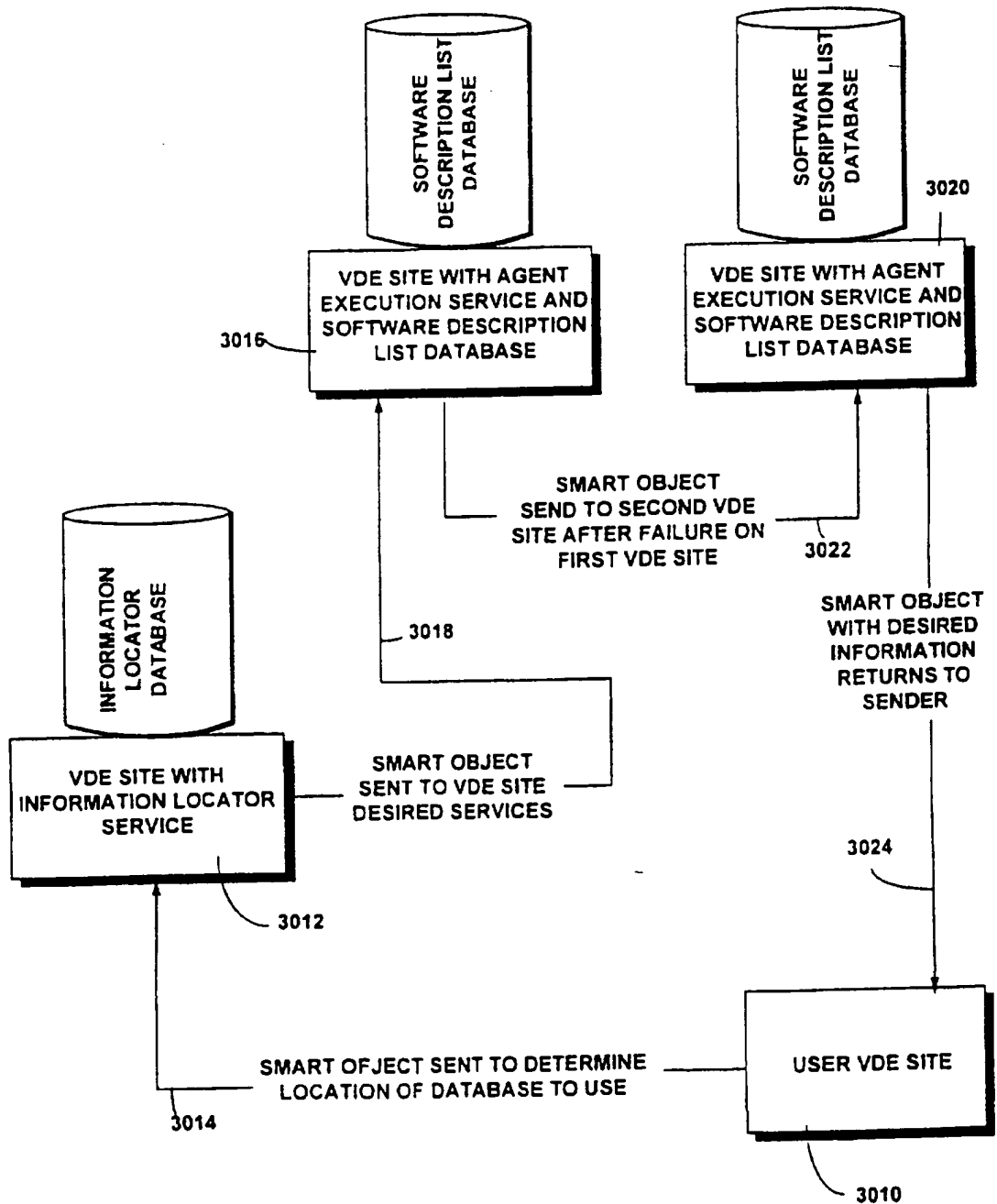
FIG. 73



SUBSTITUTE SHEET (RULE 26)

145/163

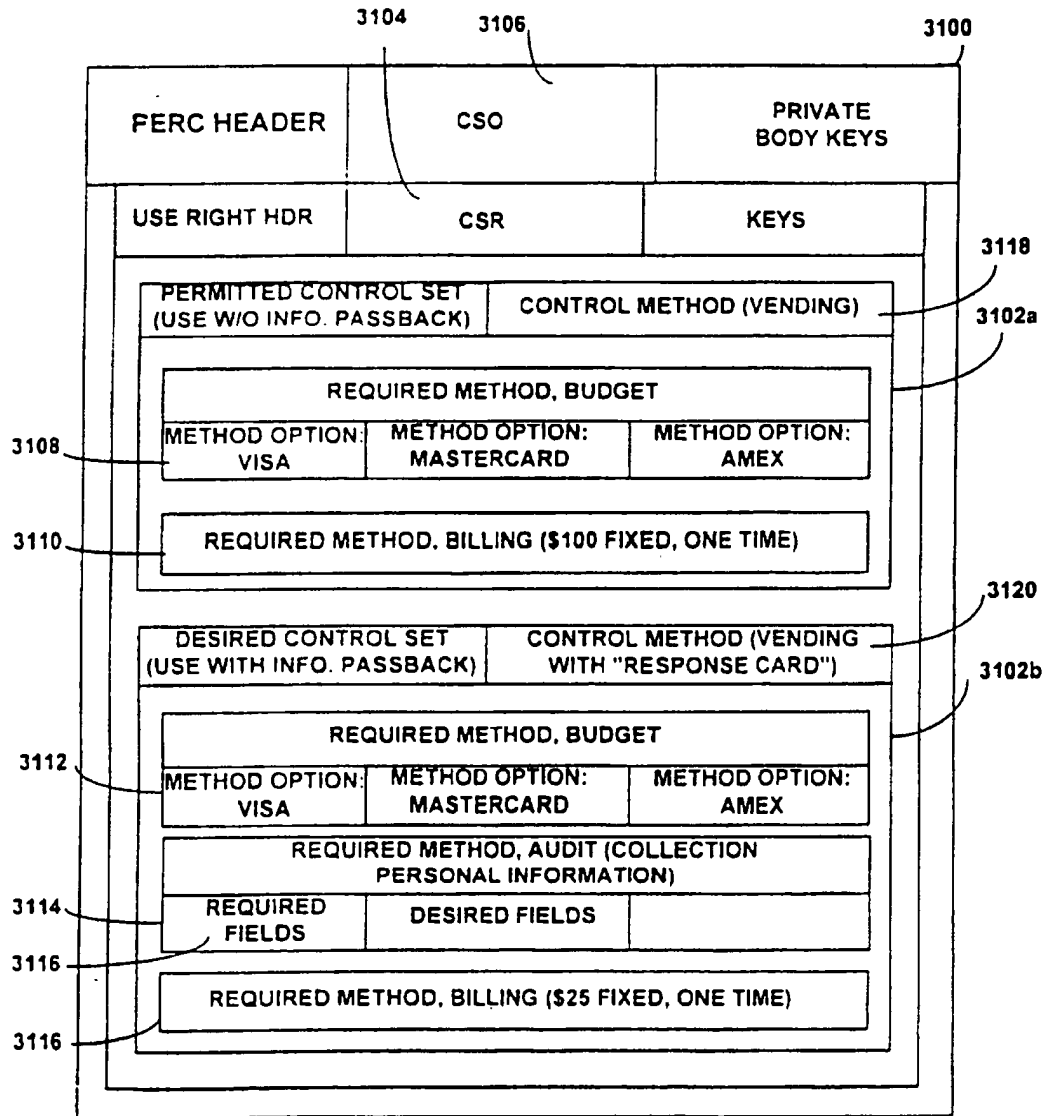
FIG. 74



SUBSTITUTE SHEET (RULE 28)

146/163

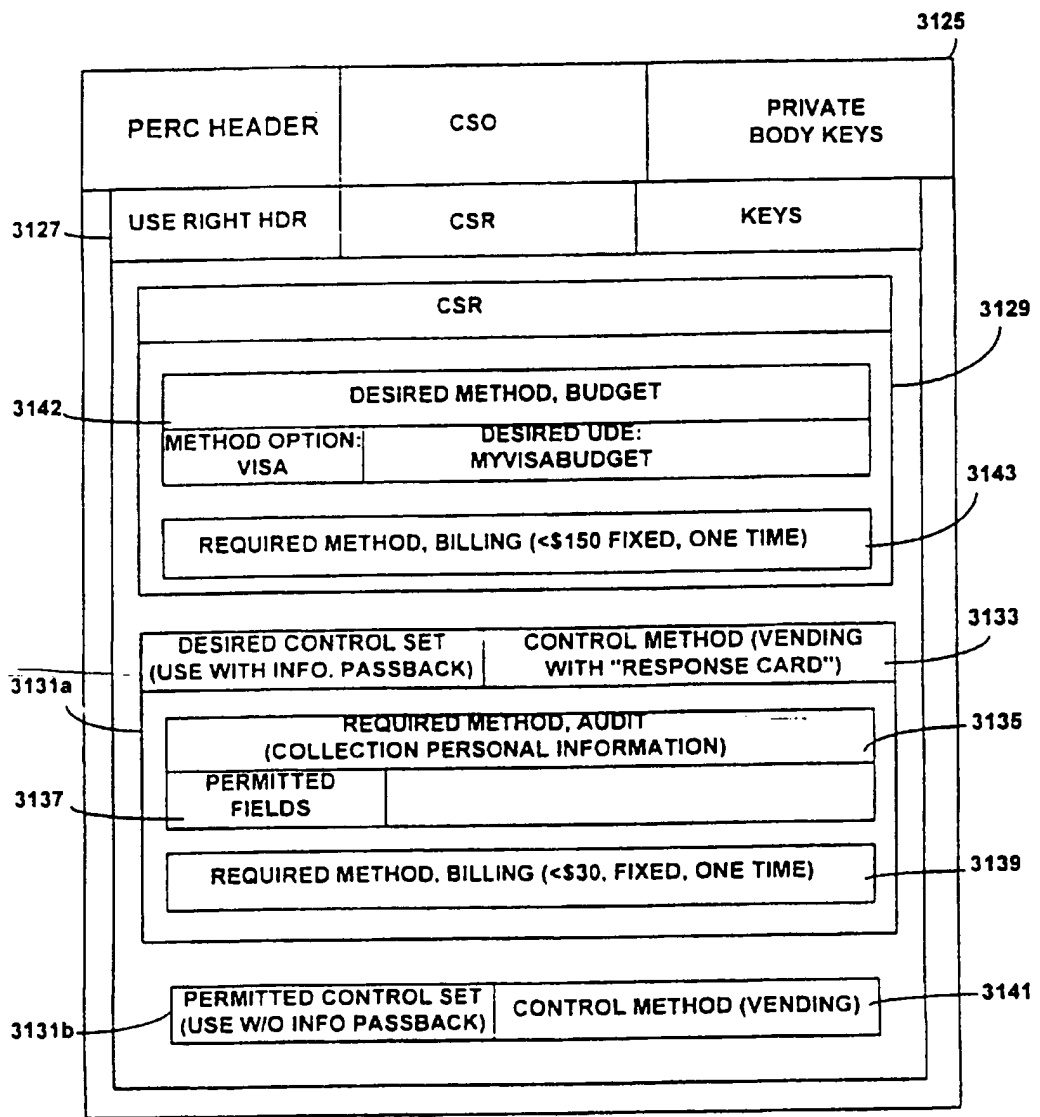
FIG. 75A



SUBSTITUTE SHEET (RULE 26)

147/163

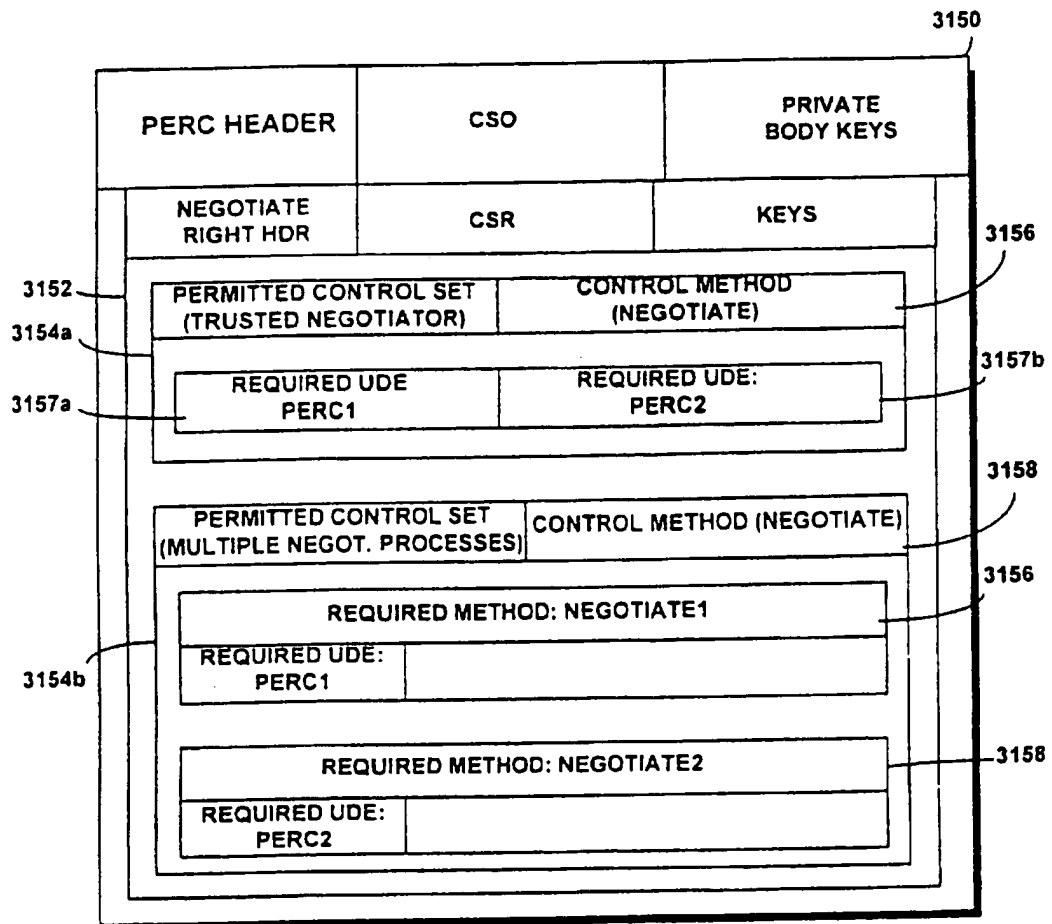
FIG. 75B



SUBSTITUTE SHEET (RULE 26)

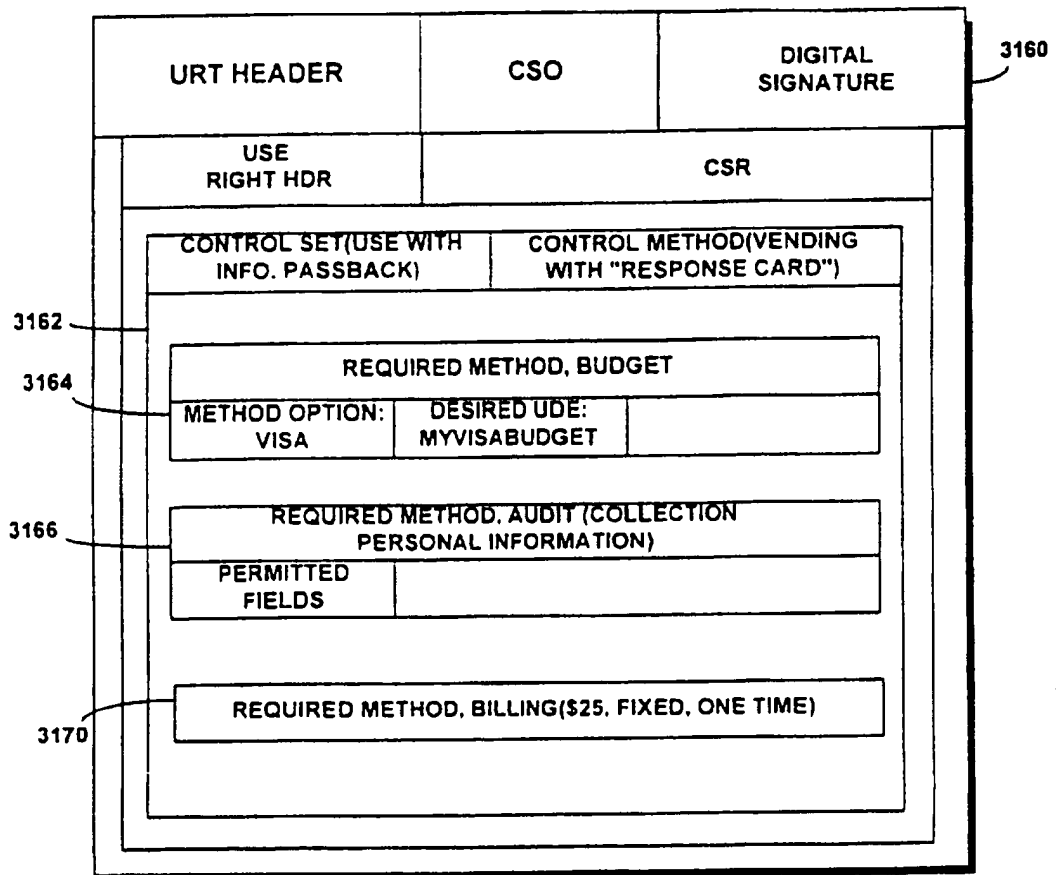
148/163

FIG. 75C



149/163

FIG. 75D



SUBSTITUTE SHEET (RULE 26)

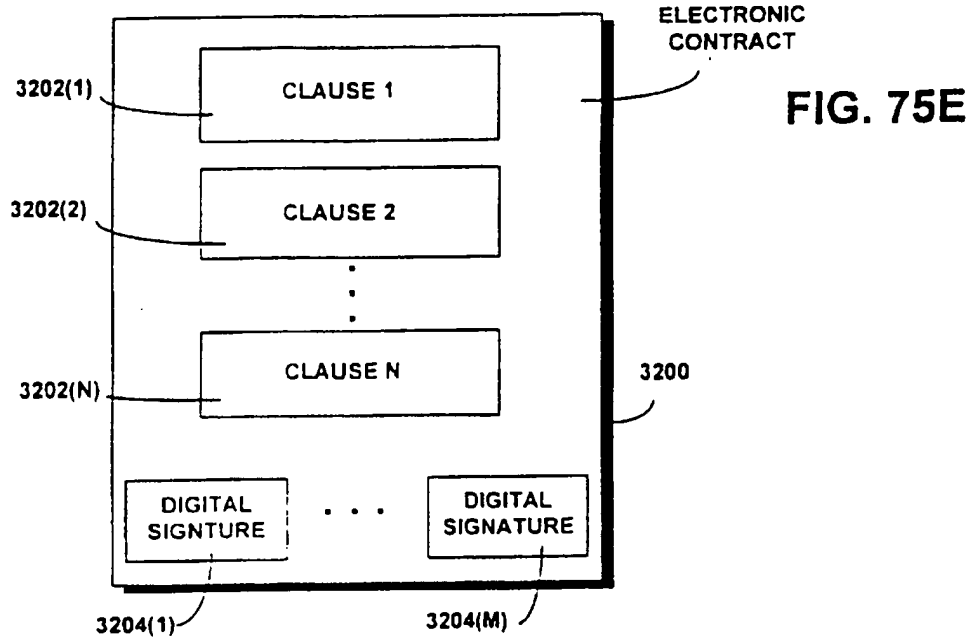


FIG. 75E

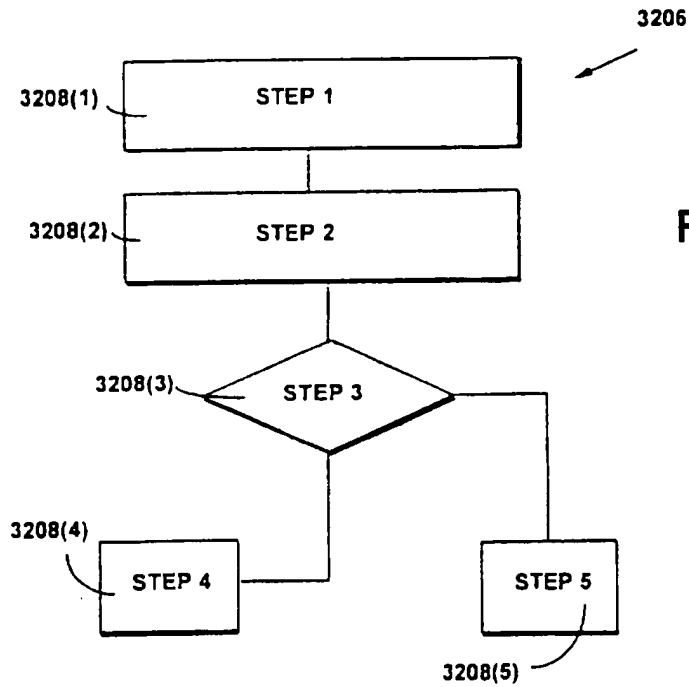
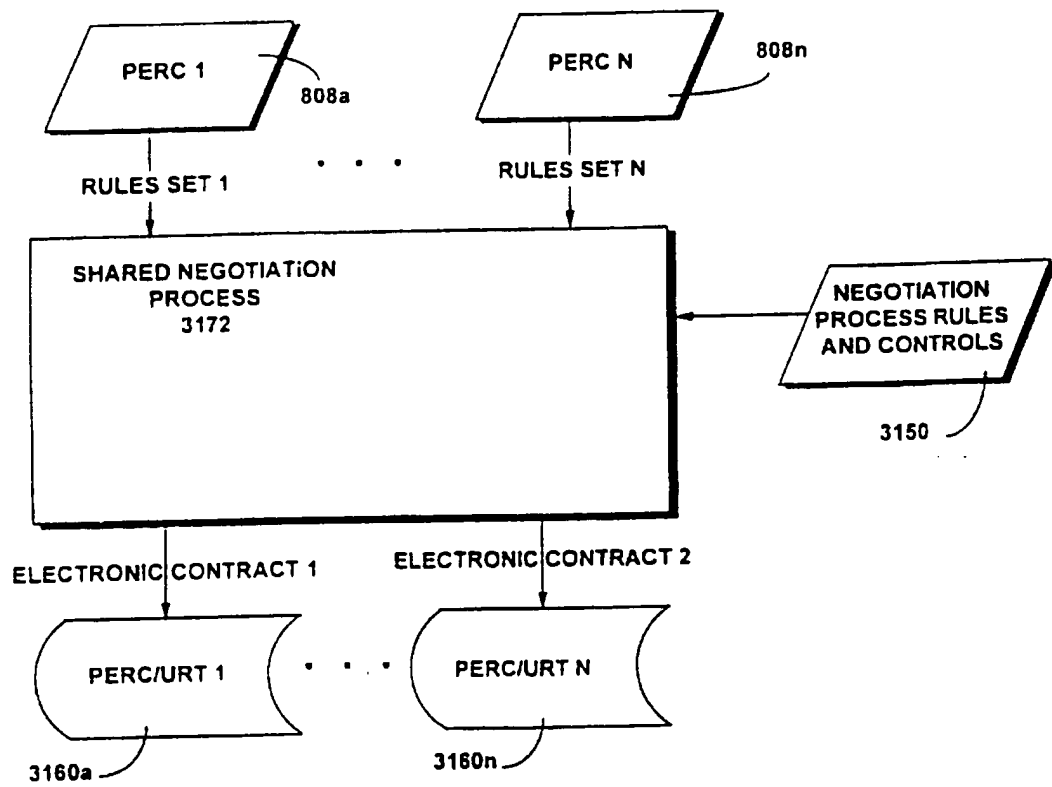


FIG. 75F

151/163

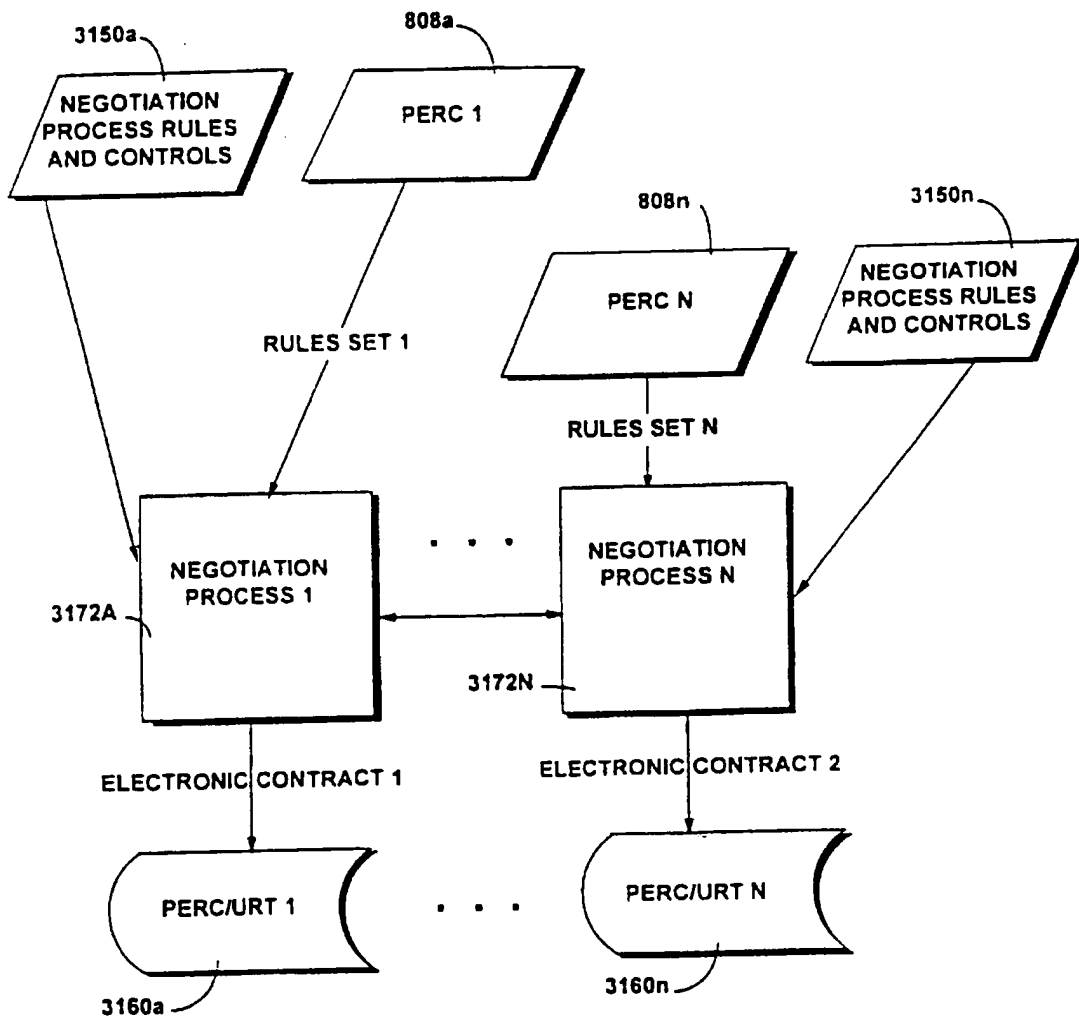
FIG. 76A



SUBSTITUTE SHEET (RULE 26)

152/163

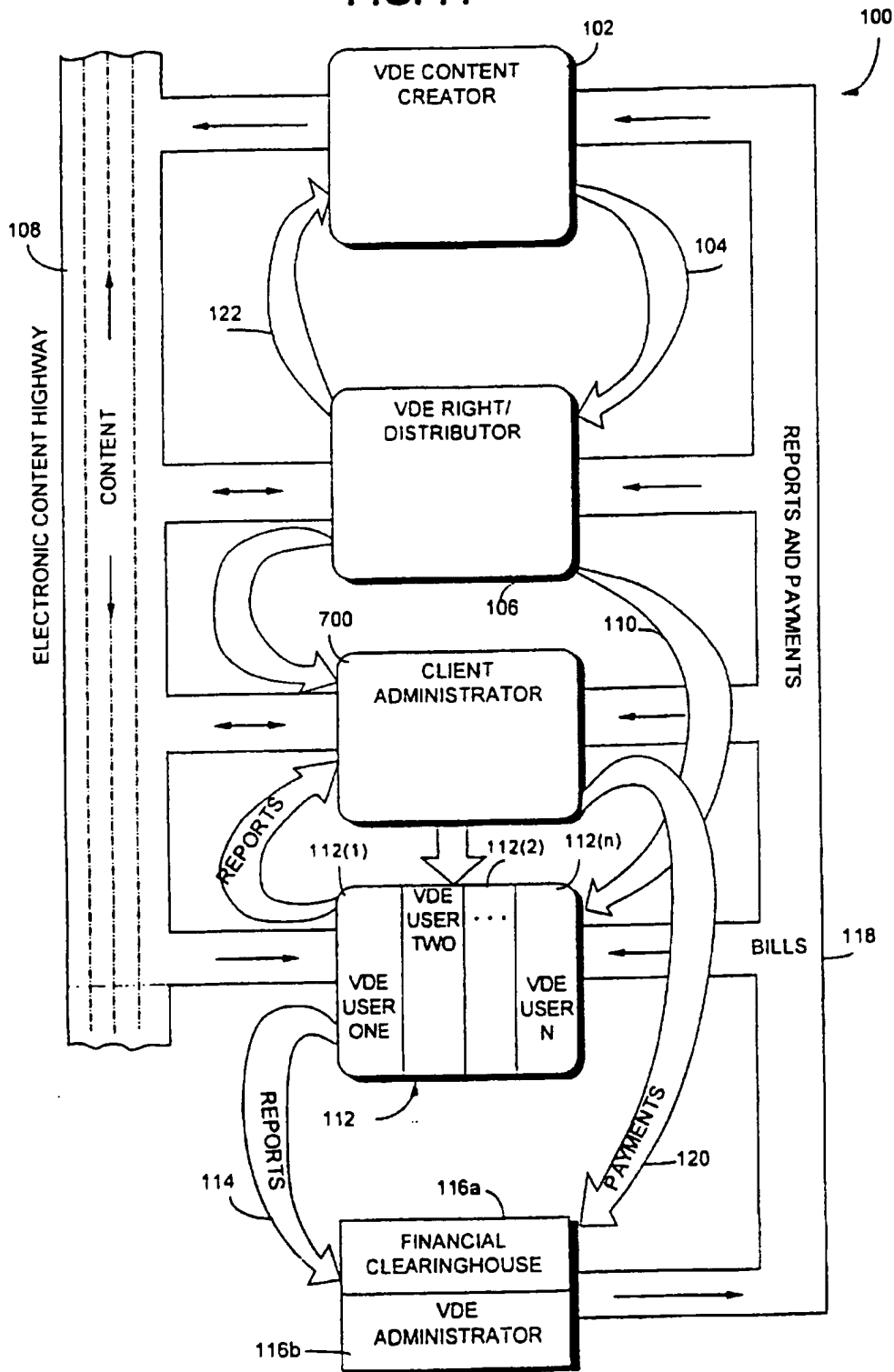
FIG. 76B



SUBSTITUTE SHEET (RULE 26)

153/163

FIG. 77



SUBSTITUTE SHEET (RULE 26)

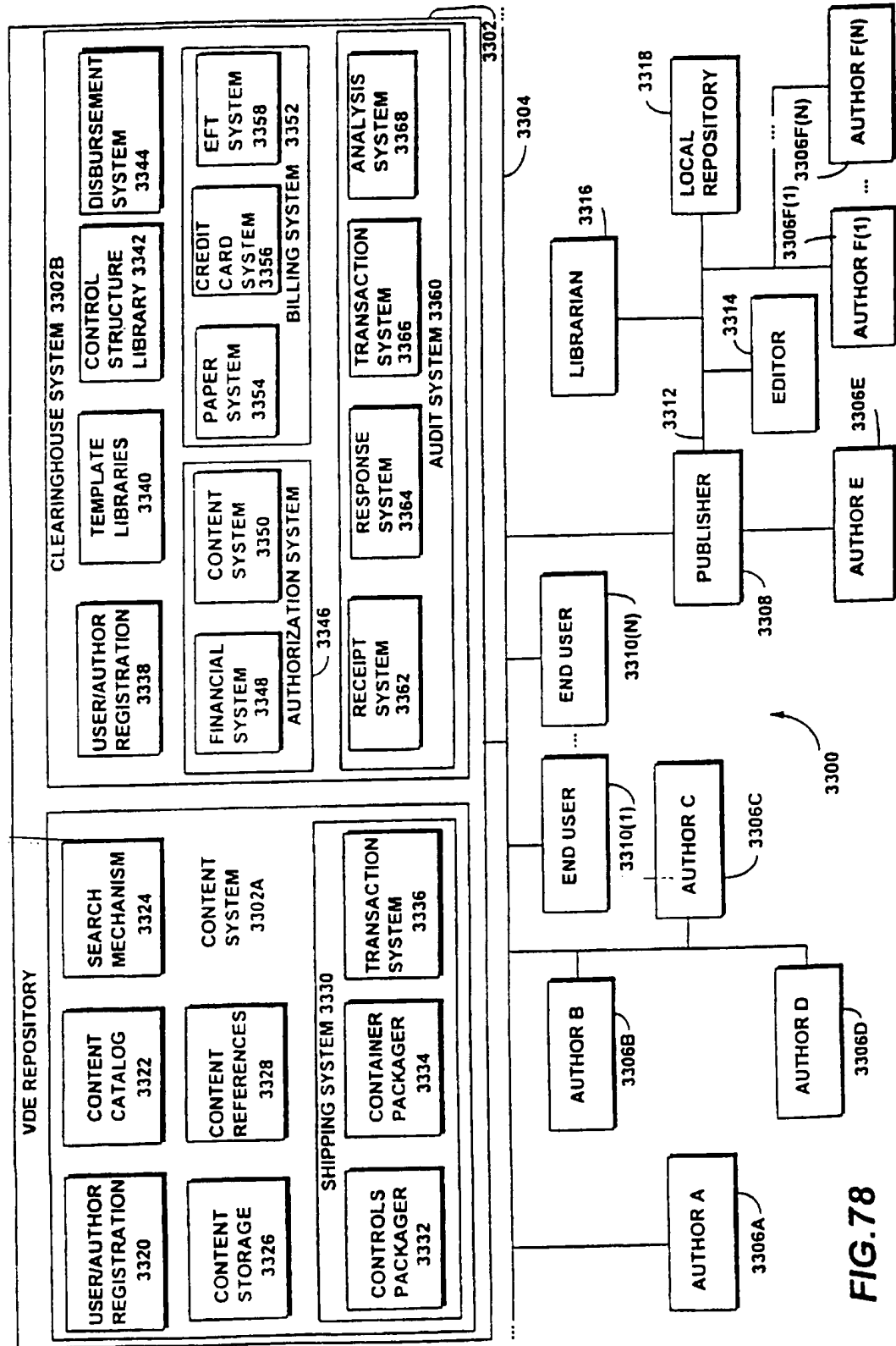
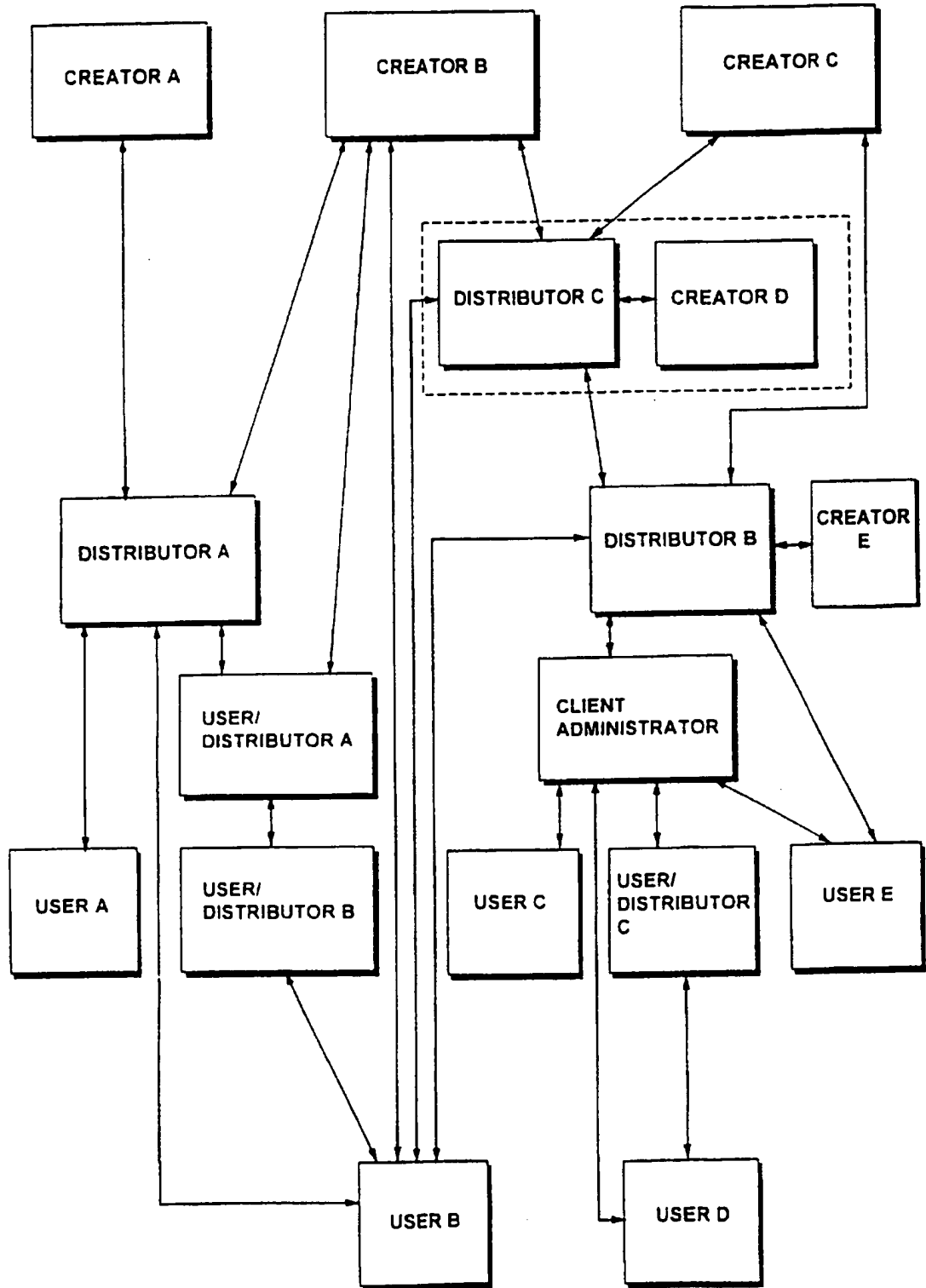


FIG. 78

155/163

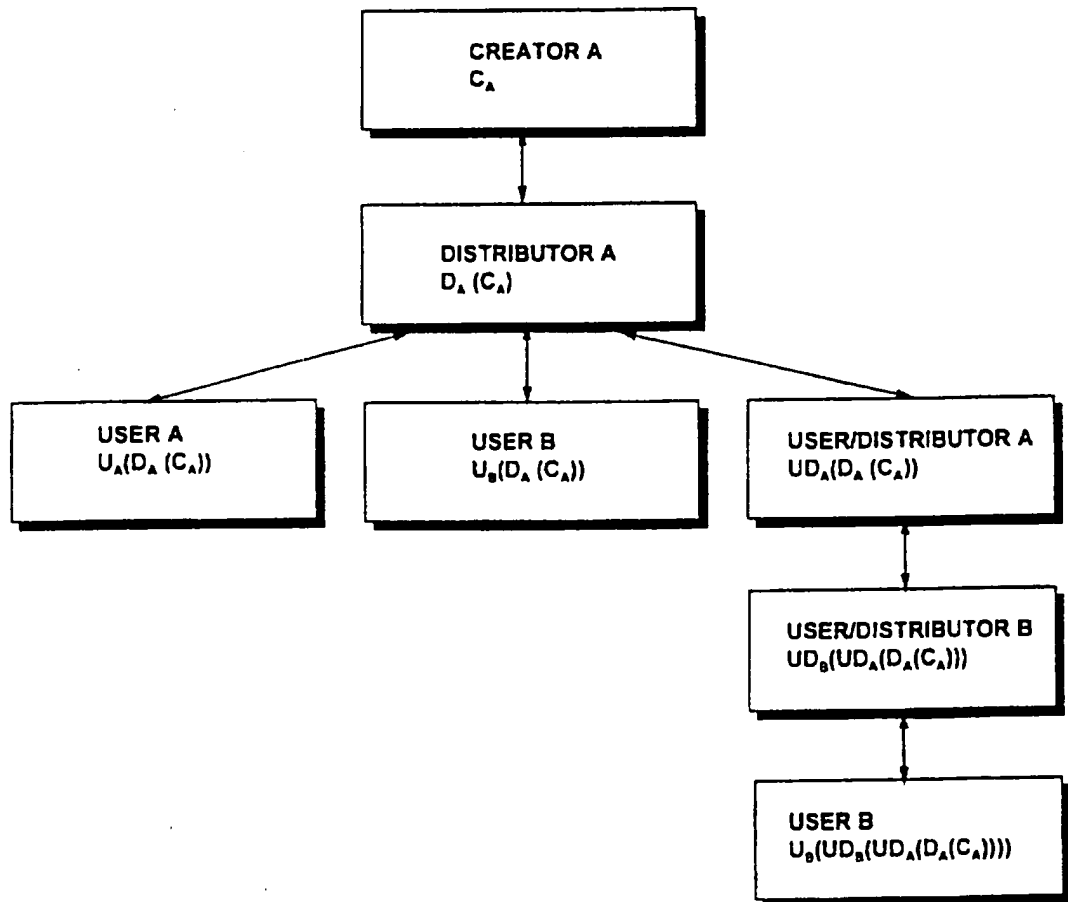
FIG. 79



SUBSTITUTE SHEET (RULE 26)

156/163

FIG. 80



SUBSTITUTE SHEET (RULE 26)

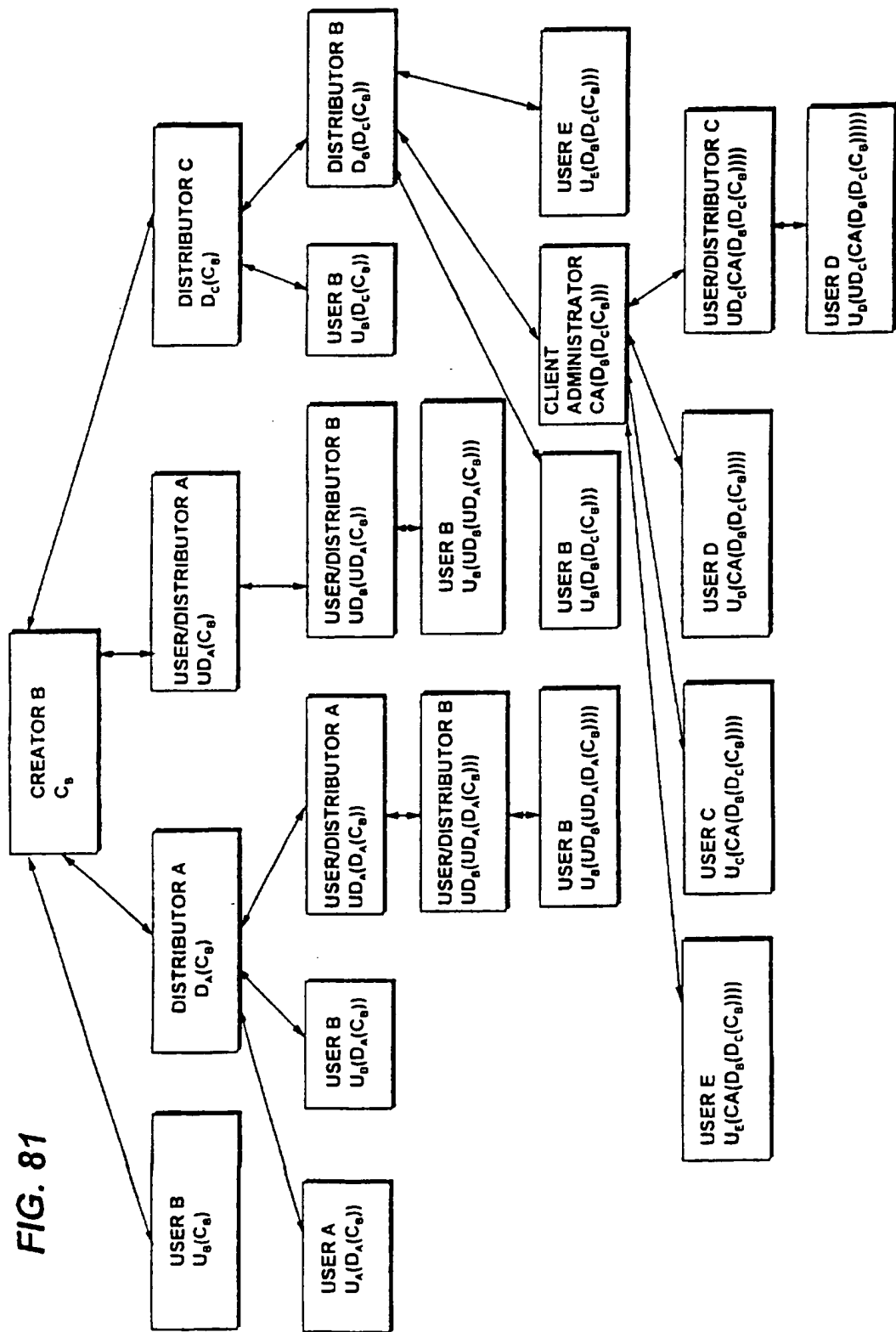
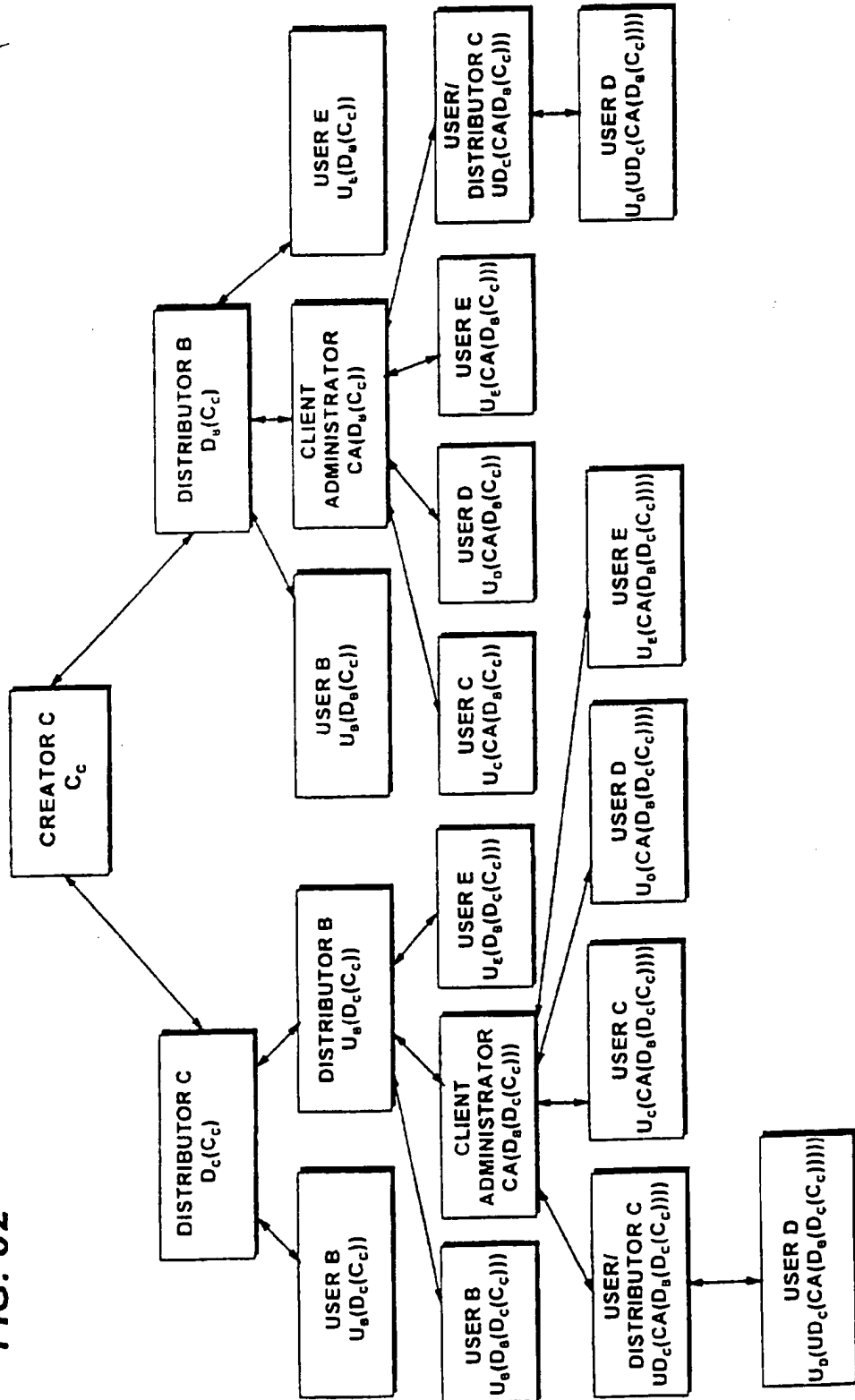


FIG. 81

FIG. 82



159/163

FIG. 83

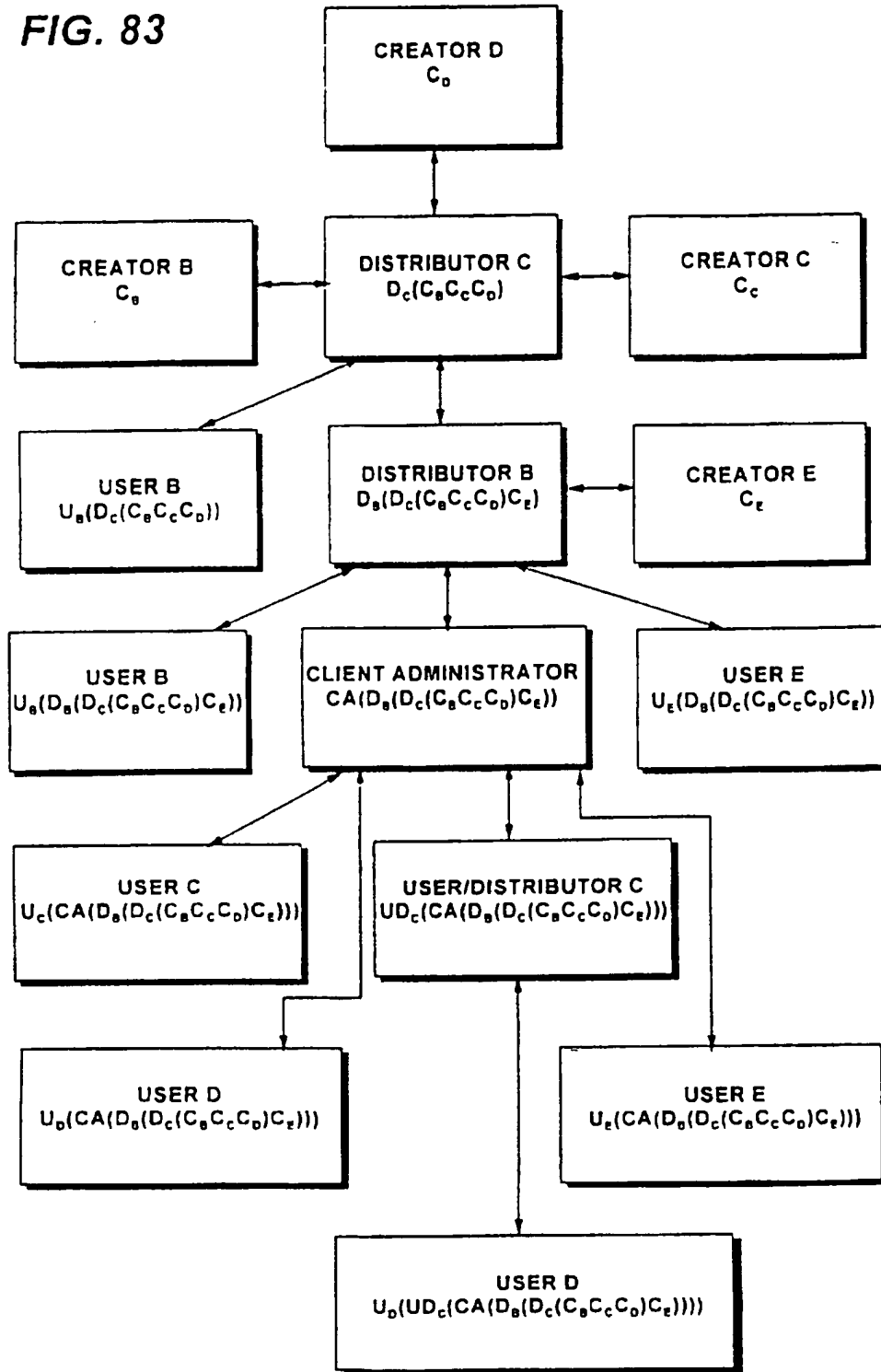
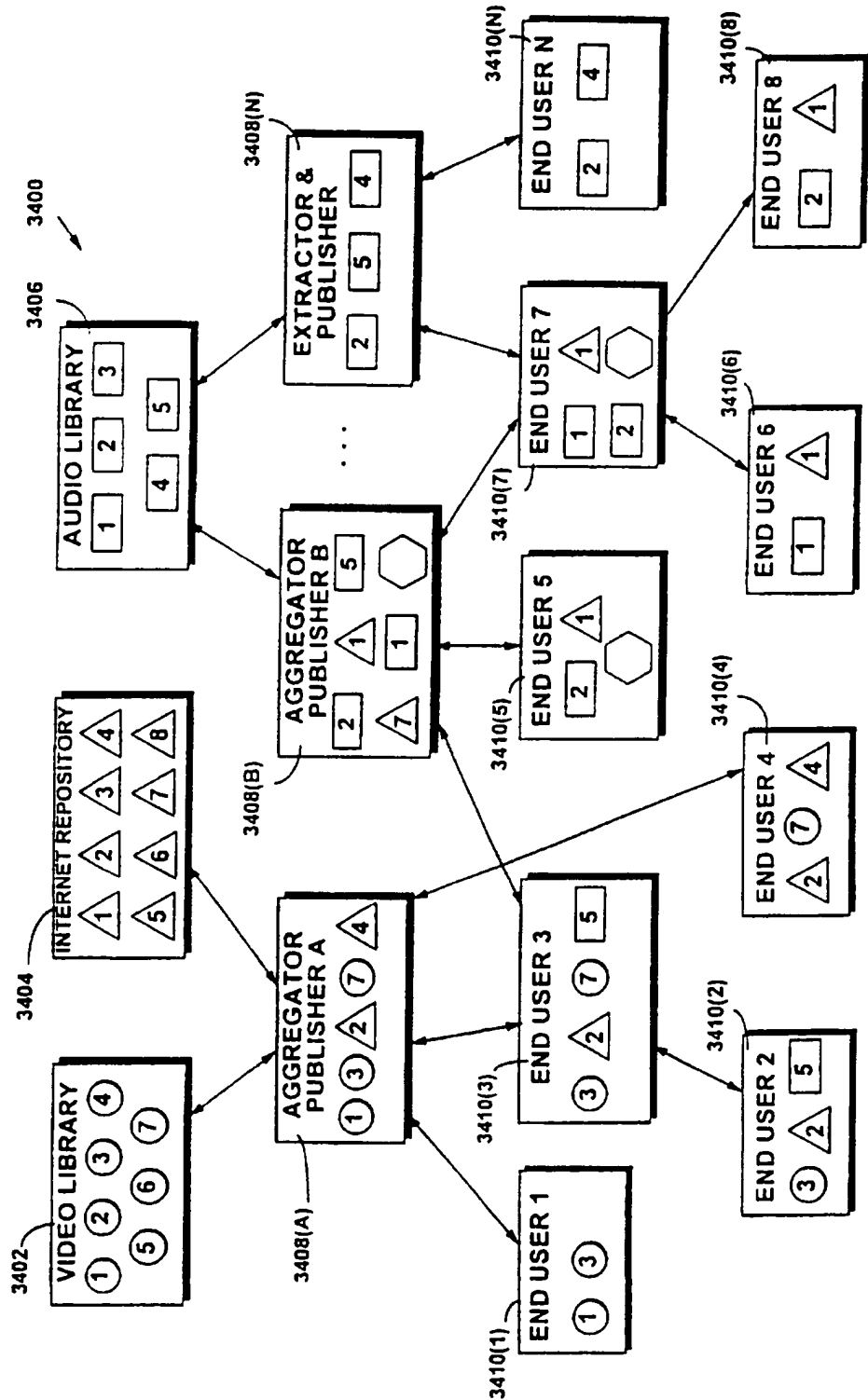
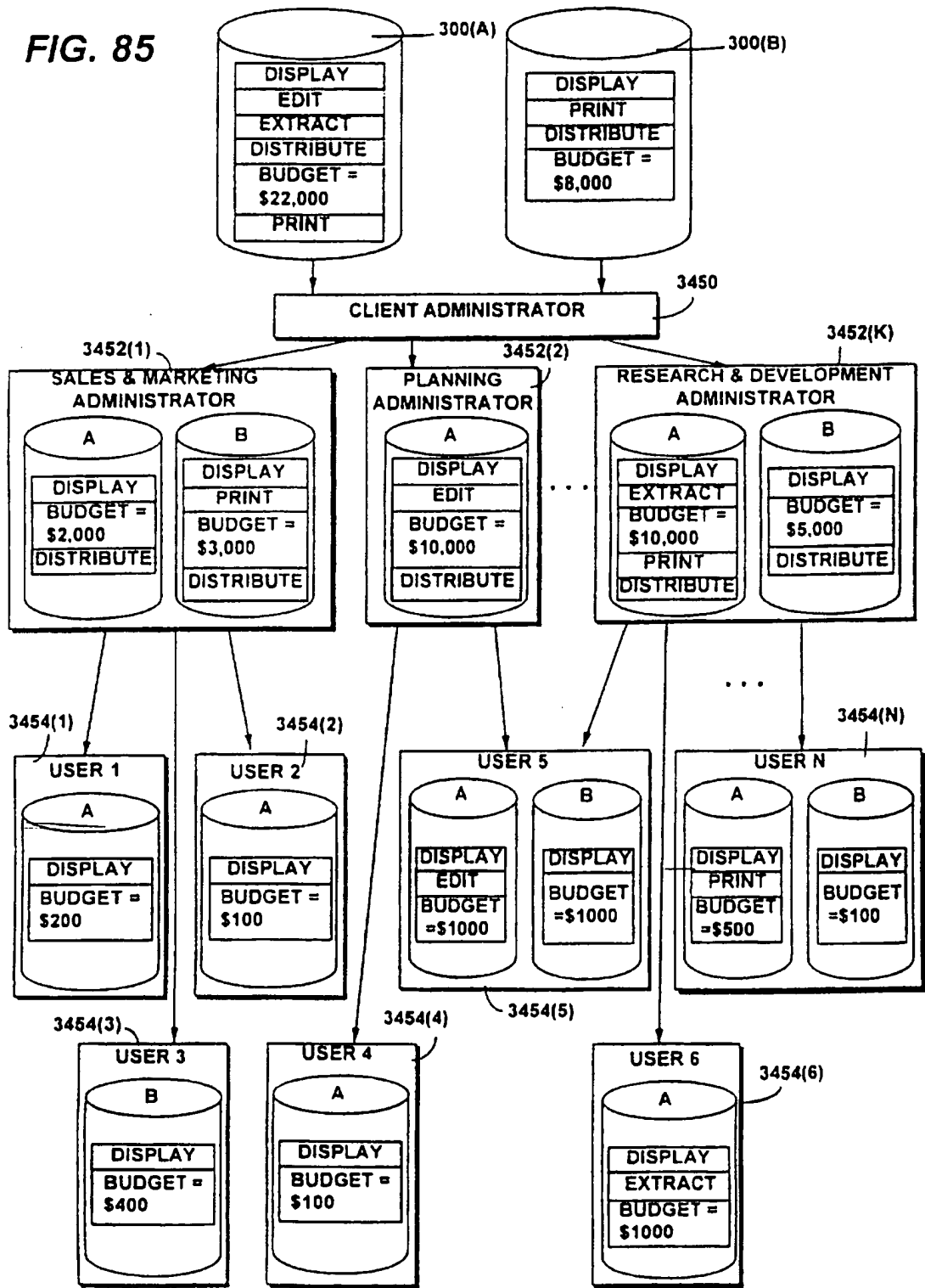


FIG. 84



161/163

FIG. 85



SUBSTITUTE SHEET (RULE 26)

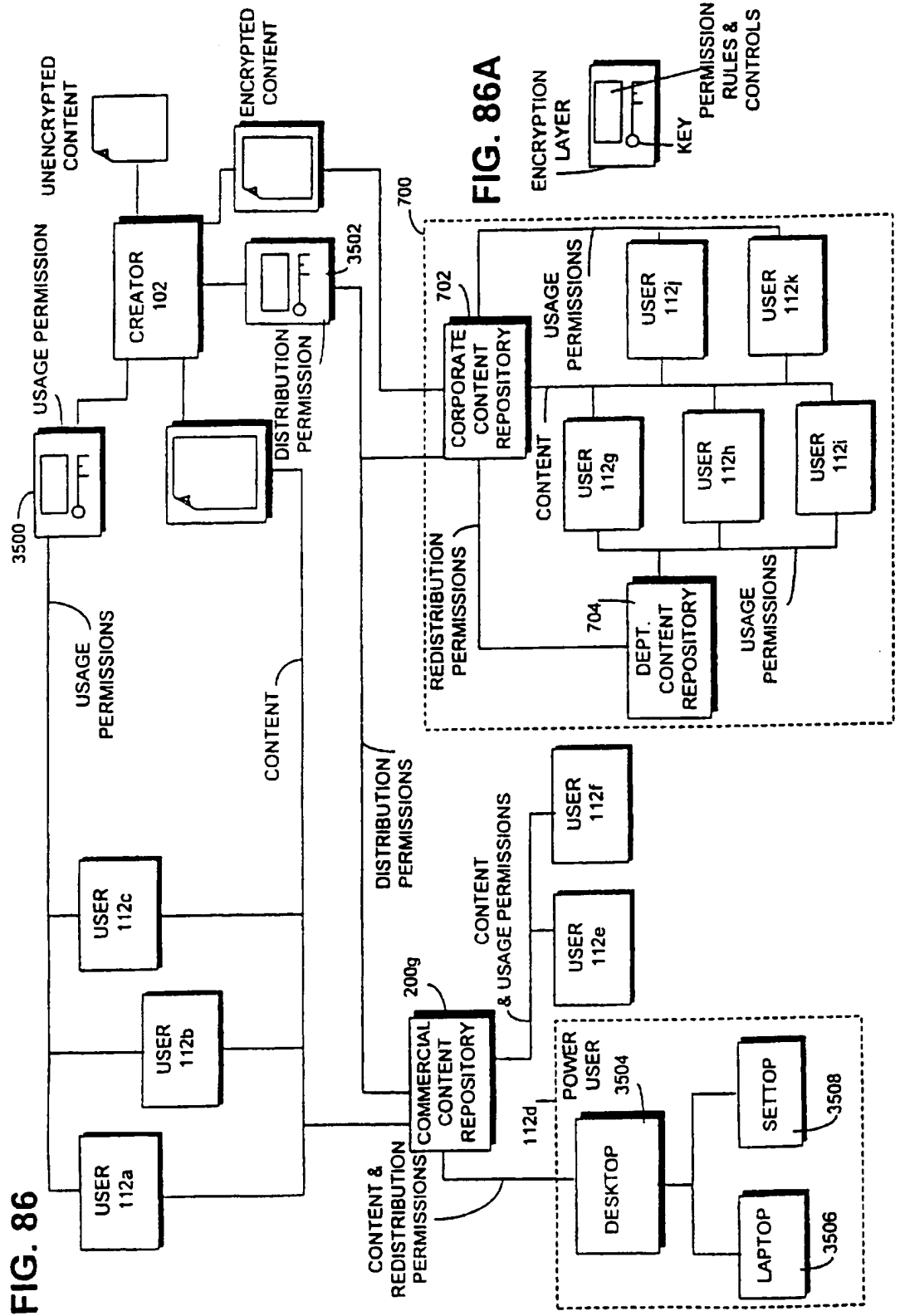


FIG. 86

FIG. 86A

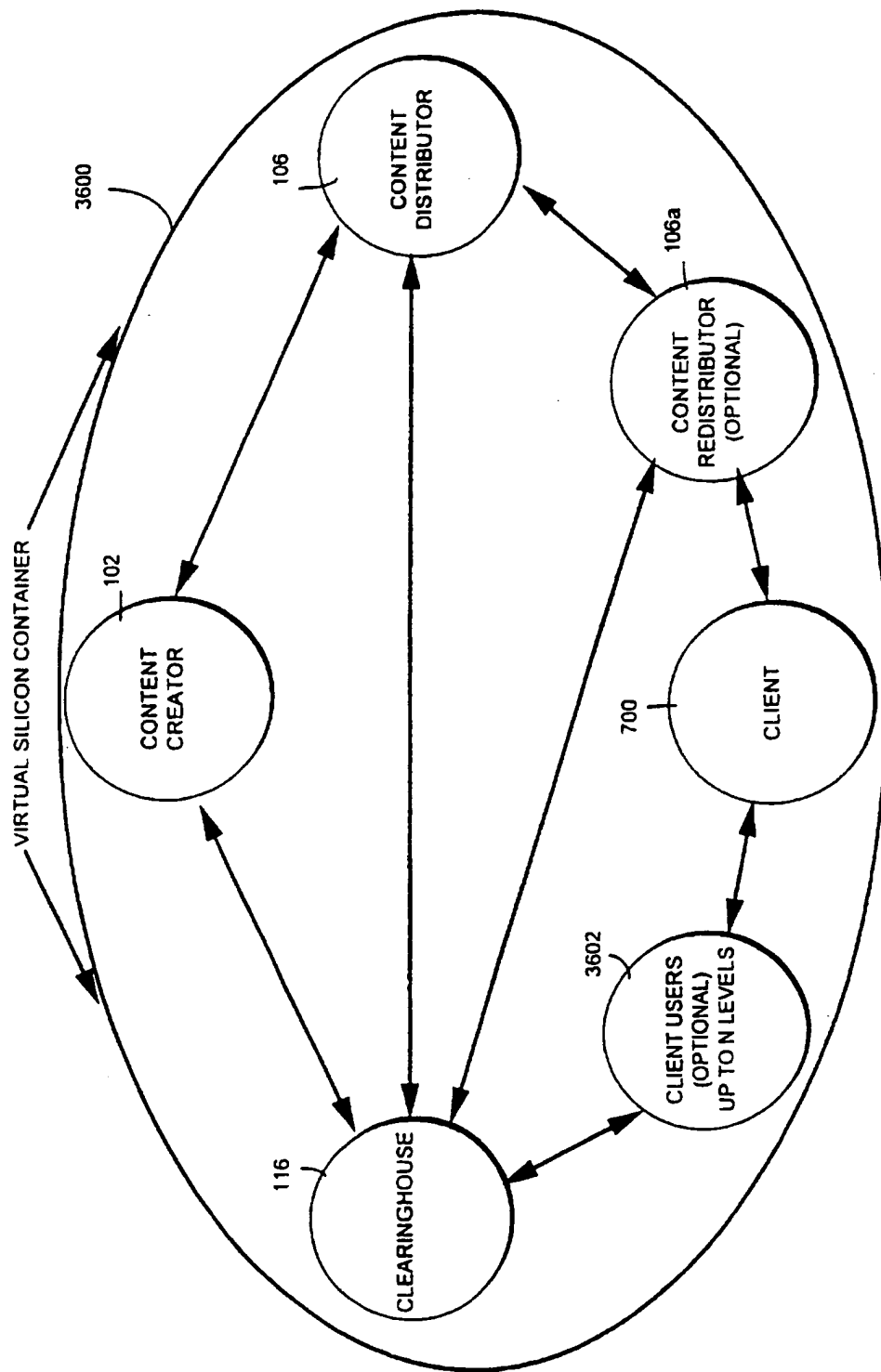


FIG. 87

INTERNATIONAL SEARCH REPORT

International Application No
PCT/US 97/15243

A. CLASSIFICATION OF SUBJECT MATTER IPC 6 G06F1/00		
According to International Patent Classification (IPC) or to both national classification and IPC		
B. FIELDS SEARCHED		
Minimum documentation searched (classification system followed by classification symbols) IPC 6 G06F		
Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched		
Electronic data base consulted during the international search (name of data base and, where practical, search terms used)		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CHOUDHURY A K ET AL: "COPYRIGHT PROTECTION FOR ELECTRONIC PUBLISHING OVER COMPUTER NETWORKS" IEEE NETWORK: THE MAGAZINE OF COMPUTER COMMUNICATIONS, vol. 9, no. 3 May 1995, pages 12-20, XP000505280 see the whole document	18
Y	WO 90 02382 A (INDATA CORP) 8 March 1990 see abstract; figures 2,12,13,15 see page 18, paragraph 3 - page 21, paragraph 2 see page 23, last paragraph - page 24, paragraph 1	1-10
A	--- -/--	11-17
<input checked="" type="checkbox"/> Further documents are listed in the continuation of box C.		
<input checked="" type="checkbox"/> Patent family members are listed in annex.		
* Special categories of cited documents :		
"A" document defining the general state of the art which is not considered to be of particular relevance "E" earlier document but published on or after the international filing date "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) "O" document referring to an oral disclosure, use, exhibition or other means "P" document published prior to the international filing date but later than the priority date claimed "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention "X" document of particular relevance: the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone "Y" document of particular relevance: the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art. "&" document member of the same patent family		
Date of the actual completion of the international search	Date of mailing of the international search report	
17 December 1997	29/12/1997	
Name and mailing address of the ISA European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx 31 651 epo nl, Fax: (+31-70) 340-3016	Authorized officer Powell, D	

INTERNATIONAL SEARCH REPORT

International Application No
PCT/US 97/15243

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	EP 0 715 246 A (XEROX CORP) 5 June 1996 see the whole document	1-10
A	-----	11-17
A	US 5 224 163 A (GASSER MORRIE ET AL) 29 June 1993 see the whole document	11-16
A	-----	
A	WO 94 01821 A (SECURE COMPUTING CORP) 20 January 1994 see the whole document	17
A	-----	
A	WO 94 03859 A (INT STANDARD ELECTRIC CORP) 17 February 1994 see the whole document	17

INTERNATIONAL SEARCH REPORT

Information on patent family members

(Int.) International Application No

PCT/US 97/15243

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9002382 A	08-03-90	AU 4188289 A EP 0472521 A US 5247575 A	23-03-90 04-03-92 21-09-93
EP 0715246 A	05-06-96	US 5638443 A JP 8263439 A	10-06-97 11-10-96
US 5224163 A	29-06-93	NONE	
WO 9401821 A	20-01-94	US 5596718 A AU 663406 B AU 4672693 A EP 0649546 A JP 7509086 T	21-01-97 05-10-95 31-01-94 26-04-95 05-10-95
WO 9403859 A	17-02-94	EP 0606401 A JP 7502847 T	20-07-94 23-03-95



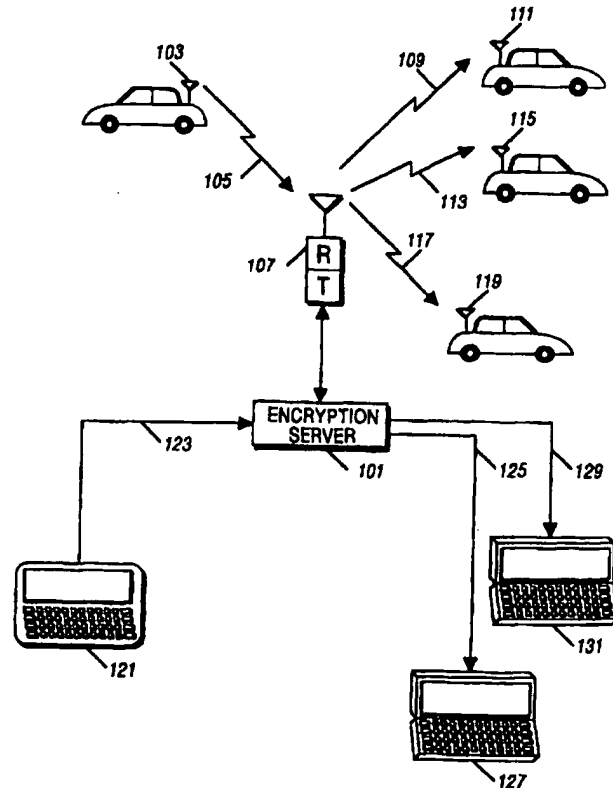
INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

<p>(51) International Patent Classification ⁶ : H04L</p>	<p>A2</p>	<p>(11) International Publication Number: WO 97/41661 (43) International Publication Date: 6 November 1997 (06.11.97)</p>
<p>(21) International Application Number: PCT/US97/06161 (22) International Filing Date: 16 April 1997 (16.04.97) (30) Priority Data: 08/639,457 29 April 1996 (29.04.96) US (71) Applicant: MOTOROLA INC. [US/US]; 1303 East Algonquin Road, Schaumburg, IL 60196 (US). (72) Inventor: DORENBOS, David; 241 N. Larch, Elmhurst, IL 60126 (US). (74) Agents: LUKASIK, Susan, L. et al.; Motorola Inc., Intellectual Property Dept., 1303 East Algonquin Road, Schaumburg, IL 60196 (US).</p>		<p>(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, HU, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, TJ, TM, TR, TT, UA, UG, UZ, VN, YU, ARIPO patent (GH, KE, LS, MW, SD, SZ, UG), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).</p> <p>Published <i>Without international search report and to be republished upon receipt of that report.</i></p>

(54) Title: USE OF AN ENCRYPTION SERVER FOR ENCRYPTING MESSAGES

(57) Abstract

An encryption server receives a first encrypted message (105) and decrypts (403) the encrypted message using a first key, yielding a decrypted message comprising a second encrypted message (105A), an identification of a sender of the first encrypted message, and an identification of a first recipient. The second encrypted message, the identification of the sender, and the identification of the first recipient are determined (405) from the decrypted message. The second encrypted message and the identification of the sender are encrypted (409) with a second key, yielding a third encrypted message (109). The third encrypted message (109) is transmitted to the first recipient.



FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakistan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

5 **USE OF AN ENCRYPTION SERVER FOR ENCRYPTING MESSAGES****Field of the Invention**

10 This invention relates to communication systems, including but not limited to encrypted communication systems.

Background of the Invention

15 Encrypted voice and data communication systems are well known. Many of these systems provide secure communications between two or more users by sharing one or more pieces of information between the users, which permits only those users knowing that information to properly
20 decrypt the message. This information is known as the encryption key, or key for short. Encryption keys may be private keys, where a single key is utilized for encryption and decryption, or public keys, where multiple keys are utilized for encryption and decryption.

25 Methods of encrypting using public-key encryption are well known in the art. Typically, a public-key encryption is a method of encryption by which a single message is encrypted using a sender's private key and then a recipient's public key. The recipient then decrypts the message using the recipient's private key and then the sender's public key. Typically, public keys are 512 bits long, although some public keys have as few as 256 bits.
30 Some encryption experts recommend using 1024-bit keys. Because the computational power required to break a key increases exponentially with the length of the key, longer keys provide more security. In addition, because two keys are needed to decrypt a message, two longer keys are more difficult to decrypt if neither key is known.

35

Today, secure communication systems are used to transmit data in an encrypted fashion. If a user wishes to send the same message to five different recipients, the user must encrypt the message five different times, each time using the public key of a different recipient for the message. The user then transmits the five messages to the five recipients. Such a process, however, is troublesome when the user wishes to transmit to, for example, 100 or more recipients. In this instance, the user must encrypt each message individually 100 or more times, one for each recipient. If the user has a portable communication device, such as a laptop computer, the user's battery may run out of power before encryption and transmission of each message has occurred. In addition, the encryption and transmission process can consume a lot of time and processing power for the portable device, rendering the portable device unavailable for other activities by the user during the encryption and transmission time period. Thus, such transmissions would be impractical for portable users.

Accordingly, there is a need for a method of transmitting encrypted data messages to multiple users without resulting in a time or power barrier to the user's communication device.

20

Brief Description of the Drawings

FIG. 1 is a block diagram of a communication system having an encryption server in accordance with the invention.

25

FIG. 2 is a block diagram of an encryption server in accordance with the invention.

FIG. 3 is a flowchart showing a method of transmission of a digital data message to an encryption server in accordance with the invention.

FIG. 4 is a flowchart showing a method of transmission of an encrypted message by an encryption server in accordance with the invention.

30

Description of a Preferred Embodiment

The following describes an apparatus for and method of using an encryption server for encrypting messages. Messages are encrypted twice, once with the sender's private key and then with an encryption server's public key before transmission of the messages to the encryption server. The encryption server decrypts received messages with the encryption server's private key, yielding an encrypted message, a user identification (ID), and one or more recipient IDs. The encryption server encrypts the encrypted message and the user ID individually with each of the recipient's public keys and transmits the resultant message(s) to the appropriate recipient. Each recipient decrypts the messages using the recipient's private key and the sender's public key. A secure communication system is thereby provided, wherein portable communication devices are neither tied up nor drained of power because the device's user wishes to send a single encrypted message to multiple recipients.

A method of using an encryption server for encrypting messages comprises the steps of, at a communication unit operated by a user generating a digital data message. The digital data message is encrypted using a first key, yielding a first encrypted message. An identification of the user and an identification of a first recipient are appended to the first encrypted message, yielding an appended first encrypted message. The appended first encrypted message is encrypted using a second key, yielding a second encrypted message. The second encrypted message is transmitted to an encryption server. At the encryption server, the second encrypted message is received. The second encrypted message is decrypted using a third key, yielding the appended first encrypted message. The first encrypted message, the identification of the user, and the identification of the first recipient are determined from the appended first encrypted message. The first encrypted message and the identification of the user are encrypted with a fourth key, yielding a third encrypted message. The third encrypted message is transmitted to the first recipient. In the preferred embodiment, the first key is a private key associated with the user, the second key is a public key associated with the encryption server, the third key is a private key associated with the encryption server, and the fourth key

is a public key associated with the first recipient. Alternatively, the second key and the third key may be identical. The transmitting steps may be performed over wireless communication resources, such as radio frequency communication resources, or wireline communication resources, such as
5 standard telephone lines or fiber optic cable.

In addition, the step of appending may further comprise the step of appending an identification of a second recipient to the first encrypted message, thereby yielding the appended first encrypted message. In this
10 case, the method further comprises the steps of encrypting, by the encryption server, the first encrypted message and the identification of the user with a fifth key, yielding a fourth encrypted message, and transmitting the fourth encrypted message to the second recipient. In the preferred
15 embodiment, the fifth key is a public key associated with the second recipient. Alternatively, the step of appending may comprise the step of appending three or more identifications of recipients to the first encrypted message, thereby yielding the appended first encrypted message.

A block diagram of a communication system having an encryption
20 server is shown in FIG. 1. An encryption server 101 is shown at the center of FIG. 1. Further details of the encryption server 101 are shown in FIG. 2 described below. A user of a first communication unit 103 utilizes the first communication unit 103 to generate a digital data message that is
25 encrypted in two stages in the preferred embodiment. In the first stage, the digital data message is encrypted using a first key, which is the user's private key in the preferred embodiment. The result of this encryption is a first-stage encrypted message. (In an alternate embodiment, the digital data message is not encrypted using the first key.) The user's identification (ID) and one or more recipient IDs are appended to the first-stage encrypted
30 message, yielding an appended message. The appended message is encrypted using a second key, yielding a second-stage encrypted message 105. In the preferred embodiment, the second key is the public key associated with the encryption server 101. The communication unit transmits the second-stage encrypted message 105 to the encryption server
35 via a wireless communication link to a wireless communication device 107, such as a radio frequency (RF) base station, repeater, or radio, or infrared

communication device. The second-stage encrypted message 105 is conveyed by the wireless communication device 107 to the encryption server 101.

5 The encryption server 101 decrypts the second-stage encrypted message 105 using an appropriate key. In the preferred embodiment, the appropriate key is the encryption server's private key. The encryption server 101 then determines the user's ID from the decrypted message and also determines the IDs of all recipients that the user indicated as intended
10 targets of the first-stage encrypted message. The encryption server 101 then encrypts the user's ID along with the first-stage encrypted message by encrypting with the public key of the first recipient. The resultant message 109 is transmitted to the first recipient, who utilizes communication unit 111. The encryption server then encrypts the first-stage encrypted message
15 along with the user's ID by encrypting with the public key of the second recipient and transmitting the resultant encrypted message 113 to the second recipient, who utilizes communication unit 115. This process continues until the encryption server reaches the last recipient ID on the user's list, and encrypts the first-stage encrypted message along with the
20 user's ID by encrypting with the public key of the last recipient and transmitting the resultant encrypted message 117 to the last recipient, who utilizes communication unit 119.

25 The encryption server 101 may also receive user requests for encryption from wireline communication devices 121 via wireline channels. As with the wireless transmission, the encryption server decrypts the received message 123 using the private key of the encryption server, then encrypts the resultant message individually for each different recipient using the appropriate recipient's individual public key. These recipients may be
30 wireline devices 127 and 131, which receive the messages 125 and 129 via wireline communication channels.

35 The above examples describe RF to RF transmission and wireline to wireline transmission of encrypted messages. Nevertheless, the method of the present invention is equally successful if a wireline device 121 requests transmission to wireless communication units 111, 115, and 119. Similarly,

a wireless communication unit 103 may request transmission from the encryption server 101 to wireline communication devices 127 and 131. In addition, the recipients may be a combination of both wireless and wireline communication units 111, 115, 119, 127, and 131, regardless of whether the sender uses a wireless communication unit 103 or a wireline communication device 121.

Upon receipt of the encrypted message from the encryption server, each recipient decrypts the message with the recipient's own private key, and after determining the user's ID, decrypts the resultant message with the user's public key, thereby yielding the original digital data message. The user is also referred to as the sender of the (second-stage) encrypted message 105.

A block diagram of an encryption server 101, including its input signals 105 and output signals 109, 113, 125, and 117, is shown in FIG. 2. In the preferred embodiment, the encryption server 101 is a Sun SparcServer2000 in a multiprocessor configuration, available from Sun Microsystems. The encryption server 101 comprises one or more processors 201, such as microprocessors or digital signal processors, as are well known in the art. The processors 201 have access to encryption and decryption algorithm(s) 203, a public key data base 205, and memory 211. The encryption/decryption algorithms 203 include public key algorithms, private algorithms, and other algorithms as may be used in the art. The public key data base 205 includes a list of IDs, as used by senders (users) and recipients, and the public keys associated with each of these IDs. The memory 211 includes programming and other data as is necessary to provide functionality as described herein for the encryption server 101. A receive block for wireline and wireless communications 207 and a transmit block for wireline and wireless communications 209 are also connected to the processors 201. The receive block for wireline and wireless communications 207 performs appropriate demodulation techniques on received messages 105 and 123. The transmit block for wireline and wireless communications 209 performs appropriate modulation techniques on messages 109, 113, 124, and 117 to be transmitted. In addition, the encryption server 101 may be equipped with hardware

and/or software to provide the encryption server 101 with over-the-air-rekeying capabilities.

As shown in FIG. 2, a user message 105 comprises a second-stage
5 encrypted (encrypted using the encryption server's public key) message
comprising the digital data message 105A, first-stage encrypted with the
user's (sender's) private key, in addition to the user ID and a number of
recipient IDs. Alternatively, the user message 105 may comprise an
unencrypted digital data message 105A, the user ID, and one or more
10 recipient IDs. The user message 105 is input to the receive
wireline/wireless block 207, the output of which is input to the processor(s)
201. The processor(s) 201 utilize(s) the encryption/decryption algorithm(s)
203 and the public key data base 205 to decrypt the message 105 using the
private key of the encryption server. The processor(s) 201 then determine(s)
15 the first-stage encrypted message 105A, the user ID, and the first recipient
ID from the decrypted message. The processor(s) 201 then determine(s) the
first recipient's public key from public key data base 205, and the encrypt the
first-stage encrypted message 105A and the user ID by using the
encryption/decryption algorithms 203 and the first recipient's public key.
20 The processor(s) 201 then append(s) the first recipient ID, thereby yielding a
message 109 that is sent to the transmit wireline/wireless block 209 for
transmitting to the first recipient's communication unit 111, as shown in
FIG. 1. A similar process is performed on the first-stage encrypted
message (or unencrypted digital data message) 105A and the user ID for
25 each of the recipients listed in the user's message 105.

In an alternate embodiment, the encryption server 101 may be physically
distributed as one or more encryption servers. In this embodiment, the
encryption server 101 encrypts the message using a second set of private
30 and public keys associated with a second server. The message so encrypted
is transmitted to the second encryption server. The second server decrypts
the message and then encrypts the message using the public key(s) of the
recipient(s). When traffic is heavy, the encryption server 101 may optimize
its efficiency by determining the computation required to transmit directly
35 to each recipient or transmit the request to one or more distributed servers.
This process is transparent to the user.

The flowchart of FIG. 3 shows a method for use by a communication unit in transmitting a digital data message to an encryption server 101. At step 301, a digital data message is generated. If at step 303 the digital data message is not to be encrypted, the process continues with step 307. If at 5 step 303 the digital data message is to be encrypted, the process continues with step 305, where the digital data message is encrypted using the private-key of the user who wishes to communicate the message. At step 307, it is determined if the IDs of the user and/or recipient(s) are to be encrypted. If 10 the IDs are to be encrypted, the process continues with step 309, where the user ID and recipient ID(s) are appended to the encrypted message from step 305 or the unencrypted message from step 301 if no encryption took place. At step 311, the message from step 309, including the appended IDs, is encrypted using the public key of the encryption server 101. The process 15 continues with step 317, where the encrypted message is transmitted to the encryption server 101. If at step 307 the IDs are not to be encrypted, the process continues with step 313, where the encrypted message of step 305 (or the unencrypted message from step 301 if no encryption took place) is encrypted with the public key of the encryption server 101. At step 315, the 20 user ID and recipient ID(s) are appended to the encrypted message of step 313, and the process continues with step 317.

In an alternative embodiment, i.e., when the digital data message is not to be encrypted at step 303 of FIG. 3, the sender or user may decrypt the 25 digital data message and, if desired, the recipient IDs only once, using the encryption server's public key. The encryption server then decrypts the message using the encryption server's private key, and encrypts the message individually for each of the recipients with the recipient's public key. The recipient then decrypts the message using only the recipient's 30 private key. This method requires the user to locally store only one public key, the key of the encryption server. With this method, a single symmetrical key may be used to encrypt and decrypt the messages between the user and the encryption server 101, and one or more keys may be used to encrypt the messages between the encryption server and the recipient. 35 Nevertheless, for better security, the encryption server 101 engaged in this embodiment should be a physically secured, e.g., locked away with limited

access, because unencrypted information is present inside the encryption server 101. An advantage of such a system includes enabling law enforcement officials the ability to read the decrypted message as available in the encryption server 101.

5

The flowchart of FIG. 4 shows the method performed by the encryption server 101 in accordance with the present invention. At step 401, the encryption server receives the encrypted message transmitted by the communication unit 103. At step 403, the encryption server decrypts the message received at step 401 with the private key of the encryption server 101. At step 405, the encryption server determines the user ID, the recipient ID(s), and the encrypted (generated at step 305 of FIG. 3) or unencrypted (generated at step 301 of FIG. 3) data message. In an alternate embodiment, the encryption server 101 may be equipped with the appropriate keys to decrypt the digital data message 105A (when the message 105A is encrypted) so that law enforcement agencies may have full access to all information transmitted in the system.

At step 407, it is determined if the IDs (i.e., the user ID and/or recipient ID(s)) are to be encrypted before transmission. If the IDs are to be encrypted, the process continues with step 409, where the encryption server encrypts the encrypted data message along with the user ID, and the recipient's ID if desired, with the recipient's public key. At step 411, the encryption server transmits the encrypted message to the recipient whose public key was used at step 409. If at step 413 there are more recipients identified by the user to which the encryption server has not yet encrypted and transmitted the message, the process continues with step 407. If there are no more recipients at step 413, the process ends. If at step 407, the IDs are not to be encrypted, the process continues with step 415, where the encrypted data message is encrypted with the recipient's public key, and the user ID and the recipient's ID are appended to that encrypted message without further encryption, and the process continues with step 411.

Optionally, all messages may be encrypted at one time, and then transmitted in succession at one time, rather than encrypting a first message with one public key, then transmitting the encrypted first message

right away, then encrypting a second message using another public key, and transmitting the encrypted second message immediately, and so forth.

5 The above text and associated drawings describe a method using public-key encryption. Private-key encryption, where the same key is used to encrypt and decrypt a message, may also be used. For example, the key used to encrypt the message send to the encryption server may be the same or identical key used to decrypt the encrypted message at the encryption server. In addition, the encryption method employed by the user to encrypt
10 the original digital data message 105A may also be private-key encryption, rather than public-key encryption. In addition, a different encryption algorithm may be utilized for the user's first stage of encryption than for the user's second stage of encryption, the result of which is transmitted to the encryption server.

15

In the above manner, the encryption server encrypts the user's data message individually for each different recipient using that particular recipient's public key. The encryption server has more computing resources available to it than an individual communication unit, and can
20 encrypt and transmit a message multiple times to many different users in a more efficient manner than can an individual communication unit. Individual communication units need not store all possible recipient's public keys, but instead need store only the encryption server's public key. Encryption of the recipient's ID(s) helps to secure the identity of the
25 recipient(s) and eliminates a source of information for traffic analysis by undesired readers/interceptors of such information.

What is claimed is:

Claims

1. A method comprising the steps of:
 - 5 at a communication unit operated by a user:
 - generating a digital data message;
 - 10 encrypting the digital data message using a first key, yielding a first encrypted message;
 - appending an identification of the user and an identification of a first recipient to the first encrypted message, yielding an appended first encrypted message;
 - 15 encrypting the appended first encrypted message using a second key, yielding a second encrypted message;
 - 20 transmitting the second encrypted message to the encryption server, wherein the encryption server is not the first recipient.
 - 25 2. The method of claim 1, wherein the first key is a private key associated with the user and wherein the second key is a public key associated with the encryption server.
 - 30 3. The method of claim 1, wherein the step of appending further comprises the step of appending an identification of a second recipient to the first encrypted message, thereby yielding the appended first encrypted message.

4. A method comprising the steps of:

at an encryption server:

5 receiving a first encrypted message;

decrypting the encrypted message using a first key, yielding a decrypted message comprising a second encrypted message, an identification of a sender of the first encrypted message, and an identification of a first

10 recipient;

determining the second encrypted message, the identification of the sender, and the identification of the first recipient from the decrypted message;

15 encrypting the second encrypted message and the identification of the sender with a second key, yielding a third encrypted message;

transmitting the third encrypted message to the first recipient.

20

5. The method of claim 4, wherein the first key is a private key associated with the encryption server and wherein the second key is a public key associated with the first recipient.

25

6. The method of claim 4, further comprising, when a second identification of a second recipient is part of the decrypted message, the steps of encrypting, by the encryption server, the second encrypted message and the identification of the sender with a third key, yielding a fourth encrypted message, and transmitting the fourth encrypted message to the second recipient.

30

7. A method comprising the steps of:
- at a communication unit operated by a user:
- 5 generating a digital data message;
- encrypting the digital data message using a first key, yielding a first encrypted message;
- 10 encrypting the first encrypted message using a second key, yielding a second encrypted message;
- appending an identification of the user and an identification of a first recipient to the second encrypted message, yielding an appended second encrypted message;
- 15 transmitting the appended second encrypted message to the encryption server;
- 20 at the encryption server:
- receiving the appended second encrypted message;
- determining the second encrypted message, the identification of the user, and the identification of the first recipient from the appended second encrypted message;
- 25 decrypting the second encrypted message using a third key, yielding the first encrypted message;
- 30 encrypting the first encrypted message with a fourth key, yielding a third encrypted message;
- transmitting the third encrypted message to the first recipient.
- 35

8. The method of claim 7, wherein the step of appending further comprises the step of appending an identification of a second recipient to the second encrypted message, thereby yielding the appended second encrypted message.

5

9. The method of claim 7, wherein the first key is a private key associated with the user, wherein the second key is a public key associated with the encryption server, wherein the third key is a private key associated with the encryption server, and wherein the fourth key is a public key associated with the first recipient.

10

10. The method of claim 7, wherein the identification of the user is encrypted using the second key before the step of appending.

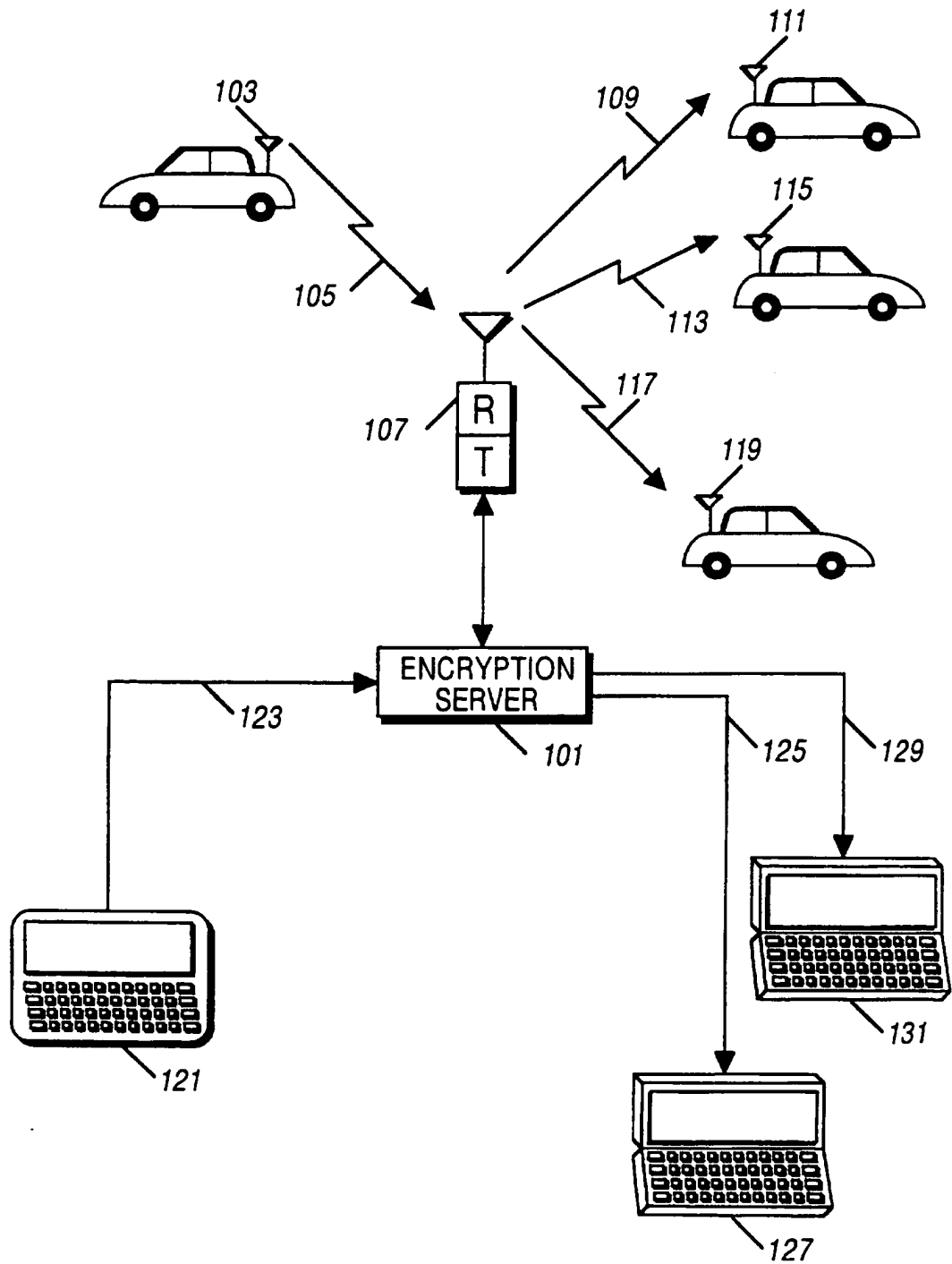


FIG. 1

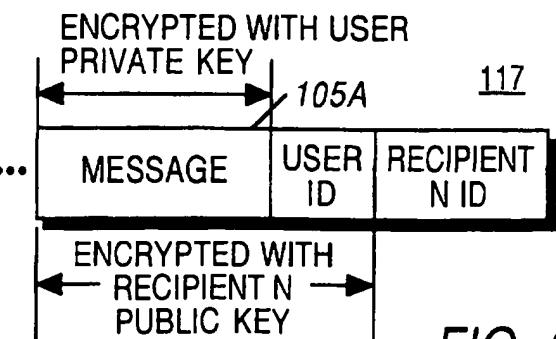
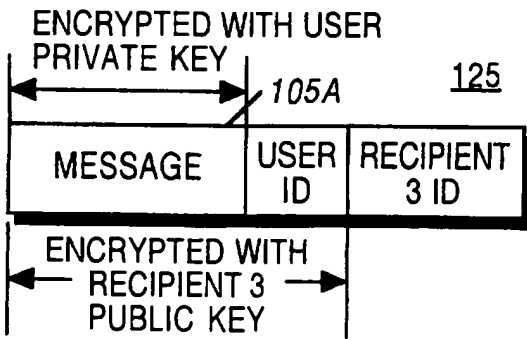
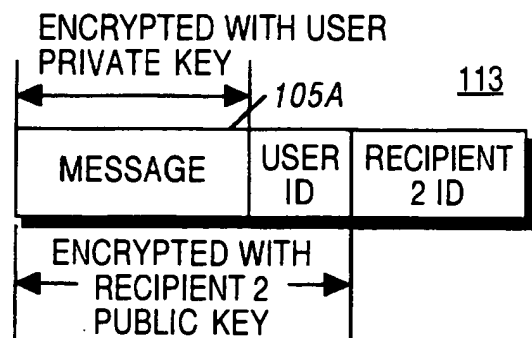
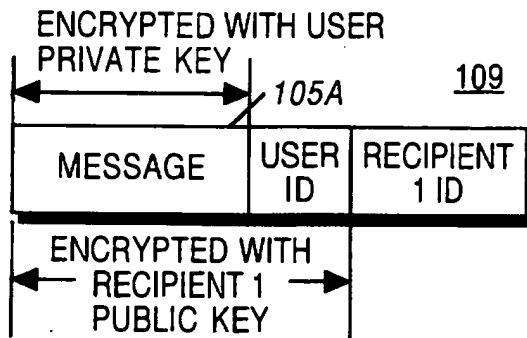
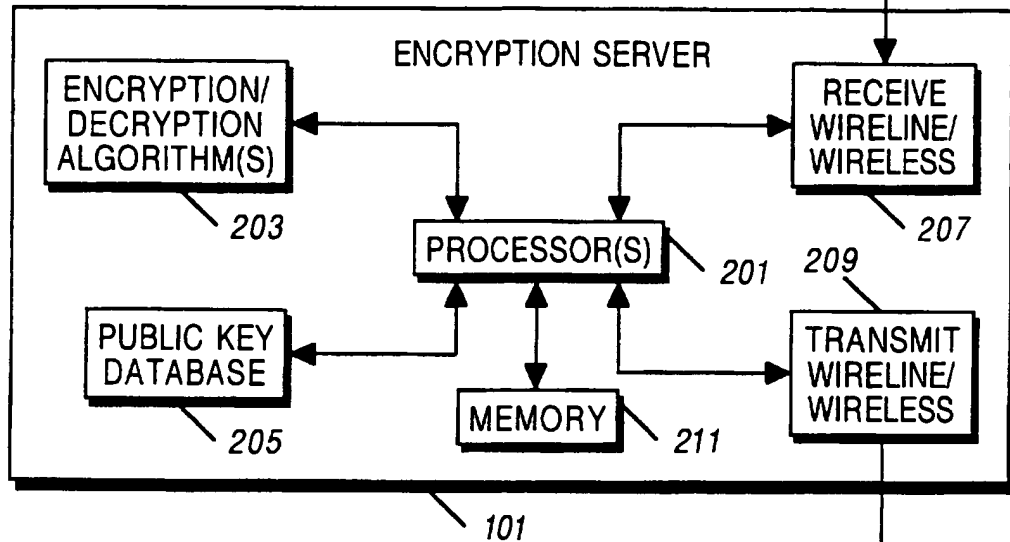
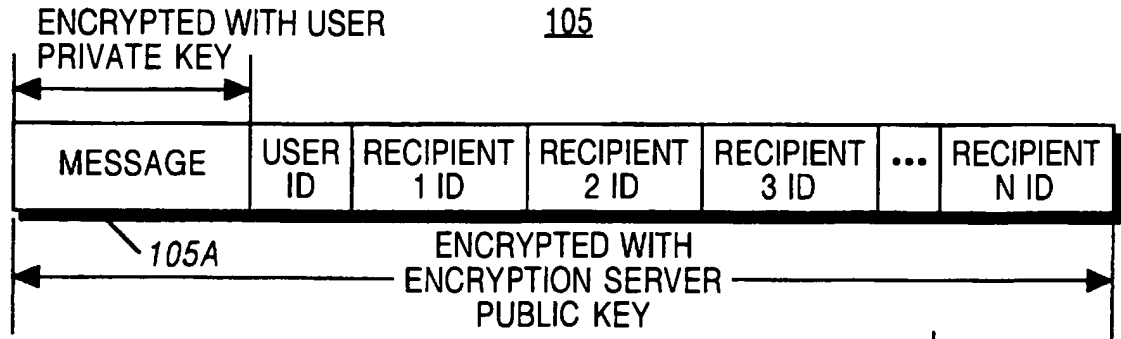


FIG. 2

FIG. 3

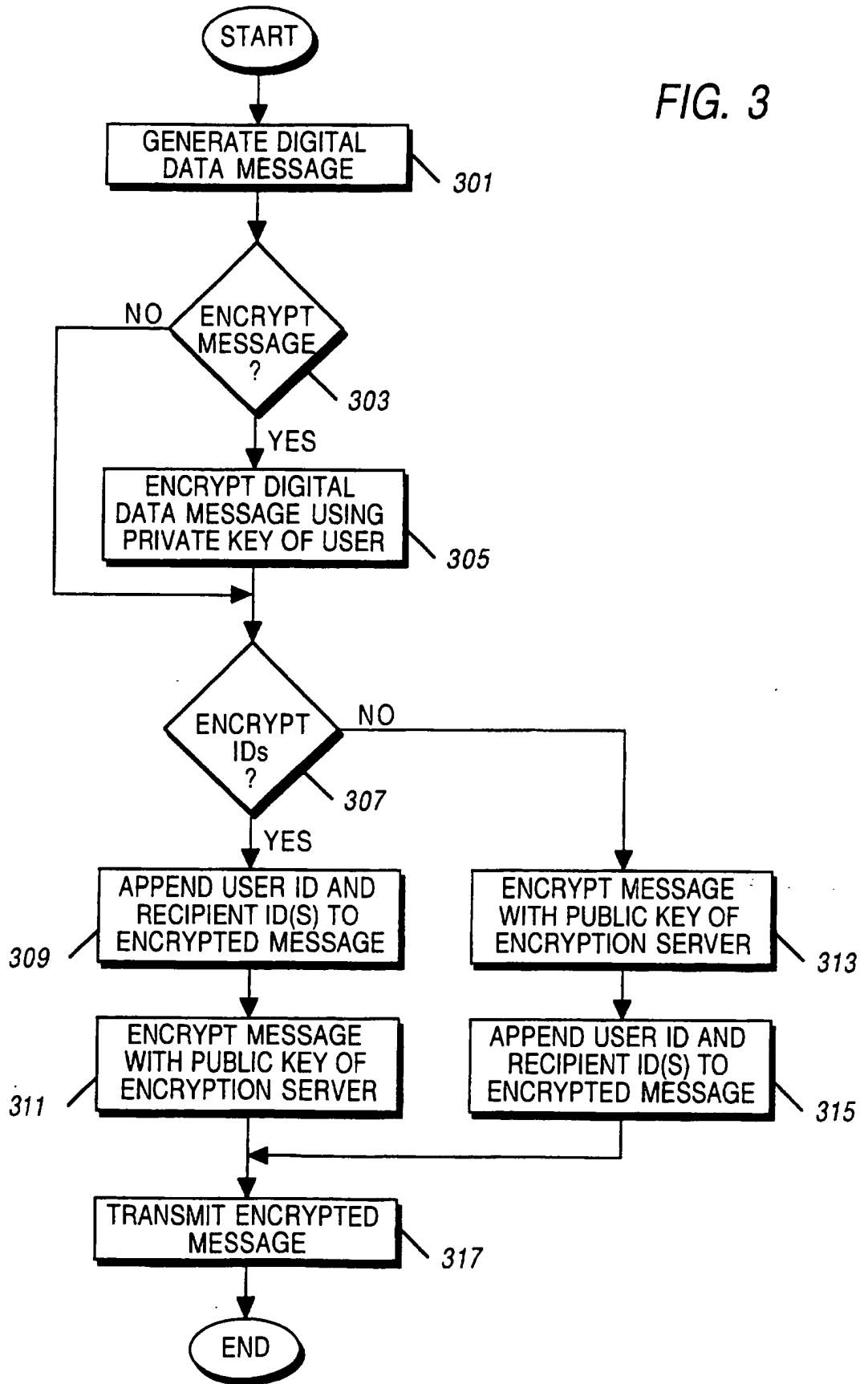
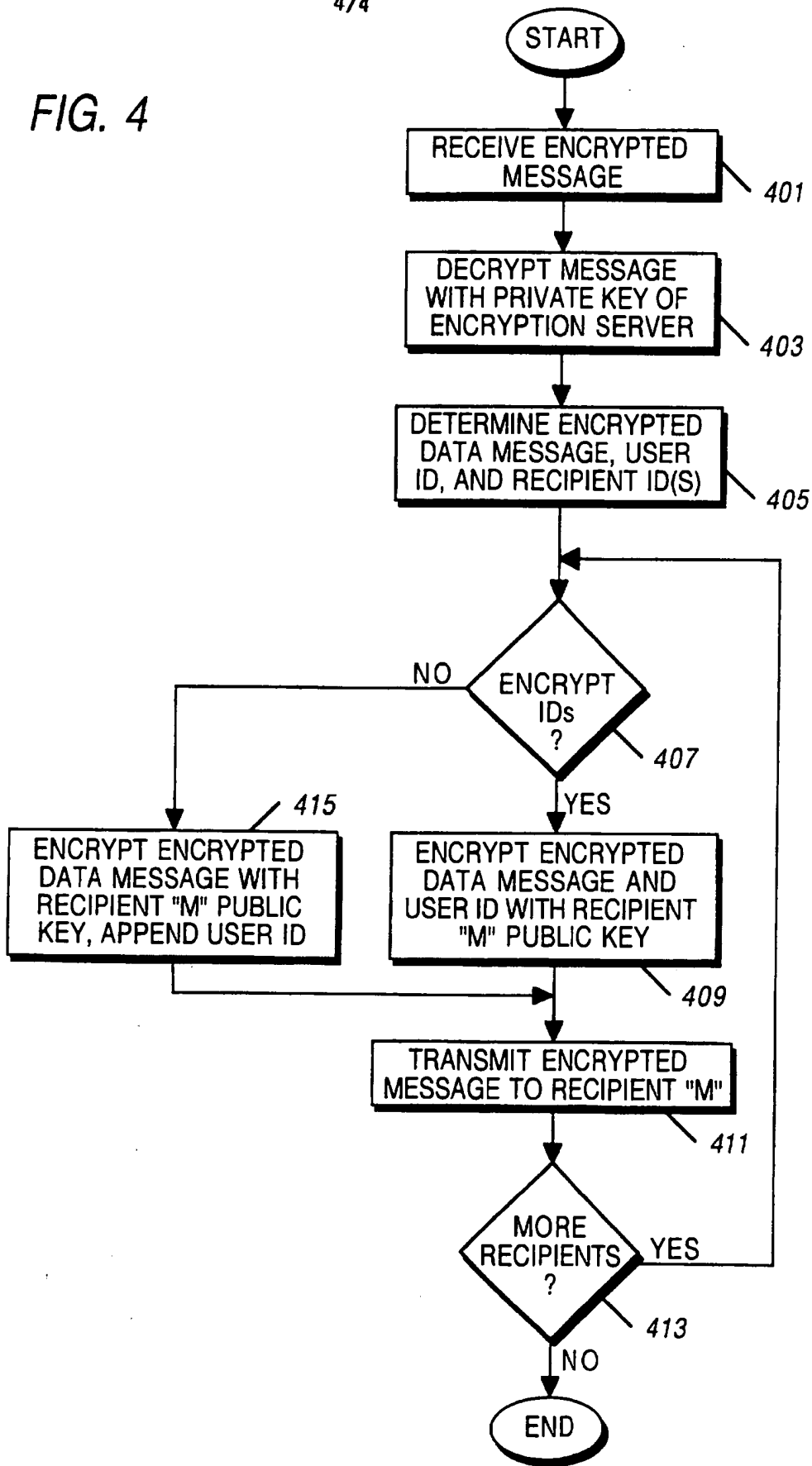


FIG. 4





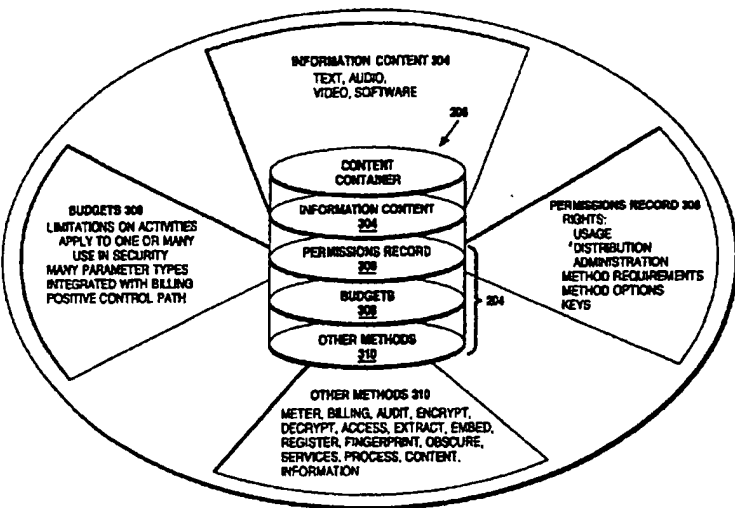
INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

<p>(51) International Patent Classification ⁶ : G11B 20/00</p>	<p>A2</p>	<p>(11) International Publication Number: WO 97/43761 (43) International Publication Date: 20 November 1997 (20.11.97)</p>
<p>(21) International Application Number: PCT/US97/08192 (22) International Filing Date: 15 May 1997 (15.05.97) (30) Priority Data: 60/017,722 15 May 1996 (15.05.96) US 60/018,132 22 May 1996 (22.05.96) US 08/689,606 12 August 1996 (12.08.96) US 08/689,754 12 August 1996 (12.08.96) US 08/699,712 12 August 1996 (12.08.96) US PCT/US96/14262 4 September 1996 (04.09.96) WO (34) Countries for which the regional or international application was filed: US et al. 60/037,931 14 February 1997 (14.02.97) US (71) Applicant (for all designated States except US): INTERTRUST TECHNOLOGIES CORP. [US/US]; 460 Oakmead Parkway, Sunnyvale, CA 94086 (US). (72) Inventors; and (75) Inventors/Applicants (for US only): SHEAR, Victor, H. [US/US]; 5203 Battery Lane, Bethesda, MD 20814 (US). SIBERT, Olin, W. [US/US]; 30 Ingleside Road, Lexington, MA 02173-2522 (US). VANWIE, David, M. [US/US]; Apartment 216, 965 E. El Camino Real, Sunnyvale, CA</p>		<p>94087 (US). WEBER, Robert, P. [US/US]; 215 Waverley Street #4, Menlo Park, CA 94025 (US). (74) Agent: FARIS, Robert, W.; Nixon & Vanderhye P.C., 8th floor, 1100 North Glebe Road, Arlington, VA 22201-4714 (US). (81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, HU, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, TJ, TM, TR, TT, UA, UG, US, UZ, VN, ARIPO patent (GH, KE, LS, MW, SD, SZ, UG), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG). Published Without international search report and to be republished upon receipt of that report.</p>

(54) Title: CRYPTOGRAPHIC METHODS, APPARATUS AND SYSTEMS FOR STORAGE MEDIAELECTRONIC RIGHTS MANAGEMENT IN CLOSED AND CONNECTED APPLIANCES

(57) Abstract

A rights management arrangement for storage media such as optical digital video disks (DVDs, also called digital versatile disks) provides adequate copy protection in a limited, inexpensive mass-producible, low-capability platform such as a dedicated home consumer disk player and also provides enhanced, more flexible security techniques and methods when the same media are used with platforms having higher security capabilities. A control object (or set) defines plural rights management rules for instance, price for performance or rules governing redistribution. Low capability platforms may enable only a subset of the control rules such as controls on copying or marking of played material. Higher capability platforms may enable all (or different subsets) of the rules. Cryptographically strong security is provided by encrypting at least some of the information carried by the media and enabling decryption based on the control set and/or other limitations. A secure "software container" can be used to protectively encapsulate (e.g., by cryptographic techniques) various digital property content (e.g., audio, video, game, etc.) and control object (i.e., set of rules) information. A standardized container format is provided for general use on/with various mediums and platforms. In addition, a special purpose container may be provided for DVD medium and appliances (e.g., recorders, players, etc.) that contains DVD program content (digital property) and DVD medium specific rules. The techniques, systems and methods disclosed herein are capable of achieving compatibility with other protection standards, such as for example, CGMA and Matsushita data protection standards adopted for DVDs. Cooperative rights management may also be provided, where plural networked rights management arrangements collectively control a rights management event on one or more of such arrangements.



FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakistan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

**CRYPTOGRAPHIC METHODS, APPARATUS
AND SYSTEMS FOR STORAGE MEDIA
ELECTRONIC RIGHTS MANAGEMENT IN
CLOSED AND CONNECTED APPLIANCES**

5 Cross-Reference to Related Applications and Patents

The specifications and drawings of the following prior,
commonly assigned published patent specifications are
incorporated by reference into this patent specification:

PCT Publication No. WO 96/27155 dated 6 September 1996

10 entitled "Systems And Methods For Secure Transaction
Management And Electronic Rights Protection", which is based
on PCT application no. PCT/US96/02303 filed 13 February 1996
and U.S. patent application serial no. 08/388,107 of Ginter et al.
entitled filed on February 13, 1995 (hereinafter "Ginter et al");

15 U.S. Patent No 4,827,508 entitled "Database Usage
Metering and Protection System and Method" dated May 2, 1989;

U.S. Patent No. 4,977,594 entitled "Database Usage
Metering and Protection System and Method" dated December 11,
1990;

U.S. Patent No. 5,050,213 entitled "Database Usage Metering and Protection System and Method" dated September 17, 1991; and

U.S. Patent No. 5,410,598 entitled "Database Usage Metering and Protection System and Method" dated April 25, 1995; and

European Patent No. EP 329681 entitled "Database Usage Metering and Protection System and Method" dated January 17, 1996.

10 In addition, the specifications and drawings of the following commonly-assigned prior-filed patent specifications are incorporated by reference into this patent application:

PCT Application No. PCT/US96/14262 filed 4 September 1996 entitled "Trusted Infrastructure Support Systems, Methods
15 And Techniques For Secure Electronic Commerce, Electronic Transactions, Commerce Process Control And Automation, Distributed Computing, And Rights Management," which corresponds to U.S. patent application serial no. 08/699,712 filed on August 12, 1996 (hereinafter "Shear et al.");

PCT Application No. _____ filed _____, 1997
entitled "Steganographic Techniques For Securely Delivering
Electronic Digital Rights Management Control Information Over
Insecure Communications Channels," which corresponds to U.S.
5 patent application serial no. 08/689,606 of Van Wie and Weber
filed on August 12, 1996 (hereinafter "Van Wie and Weber"); and

PCT Application No. _____ filed _____,
1997 based on U.S. Patent Application serial no.08/689,754
entitled "Systems and Methods Using Cryptography To Protect
10 Secure Computing Environments," of Sibert and Van Wie filed on
August 12, 1996 (hereinafter "Sibert and Van Wie").

FIELD OF THE INVENTION

This invention relates to information protection techniques
using cryptography, and more particularly to techniques using
15 cryptography for managing rights to information stored on
portable media -- one example being optical media such as Digital
Video Disks (also known as "Digital Versatile Disks" and/or
"DVDs"). This invention also relates to information protection
and rights management techniques having selectable applicability
20 depending upon, for example, the resources of the device being

used by the consumer (e.g., personal computer or standalone
player), other attributes of the device (such as whether the device
can be and/or typically is connected to an information network
("connected" versus "unconnected")), and available rights. This
5 invention further relates, in part, to cooperative rights management
-- where plural networked rights management arrangements
collectively control a rights management event on one or more of
such arrangements. Further, important aspects of this invention
can be employed in rights management for electronic information
10 made available through broadcast and/or network downloads
and/or use of non-portable storage media, either independent of, or
in combination with portable media.

BACKGROUND OF THE INVENTION

The entertainment industry has been transformed by the
15 pervasiveness of home consumer electronic devices that can play
video and/or audio from pre-recorded media. This transformation
began in the early 1900s with the invention of the
phonograph—which for the first time allowed a consumer to listen
to his or her favorite band, orchestra or singer in his or her home
20 whenever he or she wishes. The availability of inexpensive video

cassette recorders/players beginning in the early 1980s brought about a profound revolution in the movie and broadcast industries, creating an entirely new home consumer market for films, documentaries, music videos, exercise videos, etc.

5 The entertainment industry has long searched for optimal media for distributing content to home consumers. The original phonograph cylinders distributed by Thomas Edison and other phonograph pioneers had the advantage that they were difficult to copy, but suffered from various disadvantages such as high
10 manufacturing costs, low resistance to breakage, very limited playback time, relatively low playback quality, and susceptibility to damage from wear, scratching or melting. Later-developed wax and vinyl disks could hold more music material but suffered from many of the same disadvantages. Magnetic tapes, on the other
15 hand, could be manufactured very inexpensively and could hold a large amount of program material (e.g., 2, 4 or even 6 hours of video and/or audio). Such magnetic tapes could reproduce program material at relatively high quality, and were not as susceptible to damage or wearing out. However, despite the many
20 clear advantages that magnetic tape provides over other media, the

entertainment industry has never regarded it as an ideal or optimum medium because of its great susceptibility to copying.

Magnetic tape has the very flexible characteristic that it can be relatively easily recorded on. Indeed, the process for recording a magnetic tape is nearly as straightforward as that required for playing back pre-recorded content. Because of the relative ease by which magnetic tape can be recorded, home consumer magnetic tape equipment manufacturers have historically provided dual mode equipment that can both record and play back magnetic tapes. Thus, home audio and video tape players have traditionally had a "record" button that allows a consumer to record his or her own program material on a blank (un-recorded) magnetic tape. While this recording ability has given consumers additional flexibility (e.g., the ability to record a child's first words for posterity, and the ability to capture afternoon soap operas for evening viewing), it has unfortunately also been the foundation of an illegal multi-billion dollar content pirating industry that produces millions of illegal, counterfeit copies every year. This illegal pirating operation—which is international in scope—leeches huge amounts of revenue every year from the world's major

entertainment content producers. The entertainment industry must pass along these losses to honest consumers—resulting in higher box office prices, and higher video and audio tape sales and rental prices.

5 In the mid 1980s, the audio entertainment industry developed the optical compact disk as an answer to some of these problems. The optical compact disk—a thin, silvery plastic platter a few inches in diameter—can hold an hour or more of music or other audio programming in digital form. Such disks were later
10 also used for computer data. The disk can be manufactured very inexpensively, and provides extremely high quality playback that is resistant to noise because of the digital techniques used to record and recover the information. Because the optical disk can be made from plastic, it is light weight, virtually unbreakable, and
15 highly resistant to damage from normal consumer handling (unlike the prior vinyl records that were easily scratched or worn down even by properly functioning phonographs). And, because recording on an optical disk is, so far, significantly more difficult than playing back an optical disk, home consumer equipment
20 providing both recording and playback capabilities is unlikely, in

the near future, to be as cost-effective as play-only
equipment—greatly reducing the potential for illicit copying.
Because of these overwhelming advantages, the music industry
has rapidly embraced the new digital compact disk
5 technology—virtually replacing older audio vinyl disk media
within the space of a few short years.

Indeed, the threat of widespread and easy unauthorized
copying in the absence of rights management technologies
apparently has been an important contributing factor to the demise
10 of digital audio tape (DAT) as a media for music distribution and,
more importantly, home audio recording. Rightsholders in
recorded music vigorously opposed the widespread
commercialization of inexpensive DAT technology that lacked
rights management capabilities since the quality of the digital
15 recording was completely faithful to the digital source on, for
example, music CDs. Of course, the lack of rights management
was not the only factor at work, since compared with optical
media, tape format made random access difficult, for example,
playing songs out of sequence.

The video entertainment industry is on the verge of a revolution similar to that wrought by music CDs based on movies in digital format distributed on high capacity read-only optical media. For example, digital optical disk technology has advanced
5 to the point where it is now possible to digitally record, among other things, a full length motion picture (plus sound) on one side of a 5" plastic optical disk. This same optical disk can accommodate multiple high-quality digital audio channels (e.g., to record multi-channel "sensurround" sound for home theaters
10 and/or to record film dialog in multiple different languages on the same disk). This same technology makes it possible to access each individual frame or image of a movie for still image reproduction or—even more exciting—to provide an unprecedented "random access" playback capability that has never before existed
15 in home consumer equipment. This "random access" playback could be used, for example, to delete violence, foul language or nudity at time of playback so that parents could select a "PG" playback version of an "R" rated film at the press of a button. The "random access" capability also has exciting possibilities in terms
20 of allowing viewers to interact with the pre-recorded content (e.g.,

allowing a health enthusiast to select only those portions of an exercise video helpful to a particular day's workout). See, for example, "Applications Requirements for Innovative Video Programming," DVD Conference Proceedings (Interactive
5 Multimedia Association, 19-20 October 1995, Sheraton Universal Hotel, Universal City, California).

Non-limiting examples of the DVD family of optical media include:

- 10 • DVD (Digital Video Disk, Digital Versatile Disk), a non-limiting example of which includes consumer appliances that play movies recorded on DVD disks;
- 15 • DVD-ROM (DVD-Read Only Memory), a non-limiting example of which includes a DVD read-only drive and disk connected to a computer or other appliance;
- 20 • DVD-RAM (DVD Random Access Memory), a non-limiting example of which includes a read/write drive and optical media in, for example, consumer appliances for home recording and in a computer or other appliance

for the broadest range of specific applications;
and

- Any other high capacity optical media presently known or unknown.

5 "DVDs" are, of course, not limited to use with movies. Like
CDs, they may also be used for other kinds of information, for
example:

- sound recordings
- software
- 10 • databases
- games
- karaoke
- multimedia
- distance learning
- 15 • documentation
- policies and manuals

- any kind of digital data or other information
- any combination of kinds of digital data or other information
- any other uses presently known or unknown.

5 The broad range of DVD uses presents a technical challenge: how can the information content distributed on such disks, which might be any kind or combination of video, sound, or other data or information broadly speaking, be adequately protected while preserving or even maximizing consumer

10 flexibility? One widely proposed requirement for the new technology(mainly within the context of video), is, to the extent copying is permitted at all, to either: (a) allow a consumer to make a first generation copy of the program content for their own use, but prevent the consumer from making “copies of copies”, or

15 multi-generational copies of a given property (thus keeping honest people honest); or (b) to allow unlimited copying for those properties that rightsholders do not wish to protect against copying, or which consumers have made themselves.

However, providing only such simplistic and limited copy protection in a non-extensible manner may turn out to be extremely shortsighted—since more sophisticated protection and/or rights management objectives (e.g., more robust and selective application of copy protection and other protection techniques, enablement of pay-per-view models, the ability of the consumer to make use of enhanced functionality such as extracting material or interactivity upon paying extra charges, and receiving credit for redistribution, to name a few) could be very useful now or in the future. Moreover, in optimally approaching protection and rights management objectives, it is extremely useful to take differing business opportunities and threats into account that may relate to information delivered via DVD media, for example, depending upon available resources of the device and/or whether the device is connected or unconnected.

More sophisticated rights management capabilities will also allow studios and others who have rights in movies and/or sound recordings to better manage these important assets, in one example, to allow authorized parties to repurpose pieces of digital film, video and/or audio, whether specific and/or arbitrary pieces,

to create derivative works, multimedia games, in one non-limiting example. Solutions proposed to date for protecting DVD content have generally focused solely on limited copy protection objectives and have failed to adequately address or even recognize
5 more sophisticated rights management objectives and requirements. More specifically, one copy protection scheme for the initial generation of DVD appliances and media is based on an encryption method developed initially by Matsushita and the simple CGMA control codes that indicate permitted copying: a
10 one-generation copy, no copies, or unlimited copying.

SUMMARY OF THE INVENTIONS

Comprehensive solutions for protecting and managing information in systems that incorporate high capacity optical media such as DVD require, among other things, methods and
15 systems that address two broad sets of problems: (a) digital to analog conversion (and vice versa); and (b) the use of such optical media in both connected and unconnected environments. The inventions disclosed herein address these and other problems. For example, in the context of analog to digital conversion (and vice
20 versa), it is contemplated that, in accordance with the present

inventions, at least some of the information used to protect properties and/or describe rights management and/or control information in digital form could also be carried along with the analog signal. Devices that convert from one format and/or medium to another can, for example, incorporate some or all of the control and identifying information in the new context(s), or at least not actively delete such information during the conversion process. In addition, the present inventions provide control, rights management and/or identification solutions for the digital realm generally, and also critically important technologies that can be implemented in consumer appliances, computers, and other devices. One objective of the inventions is to provide powerful rights management techniques that are useful in both the consumer electronics and computer technology markets, and that also enable future evolution of technical capabilities and business models. Another non-limiting objective is to provide a comprehensive control, rights management and/or identification solution that remains compatible, where possible, with existing industry standards for limited function copy protection and for encryption.

The present inventions provide rights management and protection techniques that fully satisfy the limited copy protection objectives currently being voiced by the entertainment industry for movies while also flexibly and extensibly accommodating a wide
5 range of more sophisticated rights management options and capabilities.

Some important aspects of the present inventions (that are more fully discussed elsewhere in this application) include:

- 10 • Selection of control information associated with information recorded on DVD media (for example, rules and usage consequence control information, that comprise non-limiting example elements of a Virtual Distribution Environment (VDE)) that is based at least in
15 part on class of appliance, for example, type of appliance, available resources and/or rights;
- 20 • Enabling such selected control information to be, at least in part, a subset of control information used on other appliances and/or classes of appliance, or completely different control information;

- Protecting information output from a DVD device, such as applying rights management techniques disclosed in Ginter et al. and the present application to the signals transmitted using an IEEE 1394 port (or other serial interface) on a DVD player;
- Creation of protected digital content based on an analog source;
- Reflecting differing usage rights and/or content availability in different countries and/or regions of the world;
- Securely managing information on DVD media such that certain portions may be used on one or more classes of appliance (e.g., a standalone DVD player), while other portions may be used on the same or different classes of appliance (e.g., a standalone DVD player or a PC);
- Securely storing and/or transmitting information associated with payment, auditing, controlling and/or otherwise managing content recorded on DVD media, including techniques related to those disclosed in Ginter et al. and in Shear et al.;

- 5 • Updating and/or replacing encryption keys used in the course of appliance operation to modify the scope of information that may be used by appliances and/or classes of appliances;

- 10 • Protecting information throughout the creation, distribution, and usage process, for example, by initially protecting information collected by a digital camera, and continuing protection and rights management through the editing process, production, distribution, usage, and usage reporting.

- 15 • Allowing “virtual rights machines,” consisting of multiple devices and/or other systems that participate and work together in a permanently or in a temporarily connected network to share some or all of the rights management for a single and/or multiple nodes including, for example, allowing resources available in plural such devices and/or other systems, and/or rights associated with plural parties and/or groups using and/or controlling such devices and/or other systems, to be employed in concert (according to rights related rules and controls) so as to govern one or more electronic
- 20
- 25

events on any one or more of such devices
and/or other systems, such event governance
including, for example: viewing, editing,
subsetting, anthologizing, printing, copying,
5 titling, extracting, saving, and/or redistributing
rights protected digital content.

- Allowing for the exchange of rights among
peer-to-peer relating devices and/or other
systems, wherein such devices and/or other
10 systems participate in a temporary or
permanently connected network, and wherein
such rights are bartered, sold for currency,
and/or otherwise exchanged for value and/or
consideration where such value and/or
15 consideration is exchanged between such peer-
to-peer participating commercial and/or
consumer devices and/or other systems.

General Purpose DVD/Cost-effective Large Capacity Digital Media Rights Protection and Management

20 The inventions described herein can be used with any large
capacity storage arrangement where cost-effective distribution
media is used for commercial and/or consumer digital information
delivery and DVD, as used herein, should be read to include any
such system.

Copy protection and rights management are important in practical DVD systems and will continue to be important in other large capacity storage, playback, and recording systems, presently known or unknown, in the future. Protection is needed for some or all of the information delivered (or written) on most DVD media. Such protection against copying is only one aspect of rights management. Other aspects involve allowing rightsholders and others to manage their commercial interests (and to have them enforced, potentially at a distance in time and/or space) regardless of distribution media and/or channels, and the particular nature of the receiving appliance and/or device. Such rights management solutions that incorporate DVD will become even more significant as future generations of recordable DVD media and appliances come to market. Rightsholders will want to maintain and assert their rights as, for example, video, sound recordings, and other digital properties are transmitted from one device to another and as options for recording become available in the market.

The apparent convergence between consumer appliances and computers, increasing network and modem speeds, the declining cost of computer power and bandwidth, and the

increasing capacity of optical media will combine to create a world of hybrid business models in which digital content of all kinds may be distributed on optical media played on at least occasionally connected appliances and/or computers, in which the one-time purchase models common in music CDs and initial DVD movie offerings are augmented by other models, for example, lease, pay per view, and rent to own, to name just few. Consumers may be offered a choice among these and other models from the same or different distributors and/or other providers. Payment for use may happen over a network and/or other communications channel to some payment settlement service. Consumer usage and audit information may flow back to creators, distributors, and/or other participants. The elementary copy protection technologies for DVD now being introduced cannot support these and other sophisticated models.

As writable DVD appliances and media become available, additional hybrid models are possible, including, for example, the distribution of digital movies over satellite and cable systems. Having recorded a movie, a consumer may elect a lease, rental, pay-per-view, or other model if available. As digital television

comes to market, the ability of writable DVDs to make faithful
copies of on-air programming creates additional model
possibilities and/or rights management requirements. Here too,
simplistic copy protection mechanisms currently being deployed
5 for the initial read-only DVD technologies will not suffice.

Encryption Is A Means, Not An End

Encryption is useful in protecting intellectual properties in
digital format, whether on optical media such as DVD, on
magnetic media such as disk drives, in the active memory of a
10 digital device and/or while being transmitted across computer,
cable, satellite, and other kinds of networks or transmission
means. Historically, encryption was used to send secret messages.
With respect to DVD, a key purpose of encryption is to require the
use of a copy control and rights management system in order to
15 ensure that only those authorized to do so by rightsholders can
indeed use the content.

But encryption is more of a means, rather than an end. A
central issue is how to devise methods for ensuring, to the
maximal extent possible, that only authorized devices and parties
20 can decrypt the protected content and/or otherwise use information

only to the extent permitted by the rightsholder(s) and/or other relevant parties in the protected content.

The Present Inventions

The present inventions provide powerful right management capabilities. In accordance with one aspect provided by the present invention, encrypted digital properties can be put on a DVD in a tamper-resistant software "container" such as, for example, a "DigiBox" secure container, together with rules about "no copy" and/or "copy" and/or "numbers of permitted copies" that may apply and be enforced by consumer appliances. These same rules, and/or more flexible and/or different rules, can be enforced by computer devices or other systems that may provide more and/or different capabilities (e.g., editing, excerpting, one or more payment methods, increased storage capability for more detailed audit information, etc.). In addition, the "software container" such as for example, a "DigiBox" secure container, can store certain content in the "clear" (that is, in unencrypted form). For example, movie or music titles, copyright statements, audio samples, trailers, and/or advertising can be stored in the clear and/or could be displayed by any appropriate application or

device. Such information could be protected for authenticity (integrity) when available for viewing, copying, and/or other activities. At the same time, valuable digital properties of all kinds—film, video, image, text, software, and multimedia— may be
5 stored at least partially encrypted to be used only by authorized devices and/or applications and only under permitted, for example rightsholder-approved, circumstances.

Another aspect provided in accordance with the present invention (in combination with certain capabilities disclosed in
10 Ginter et al.) is that multiple sets of rules could be stored in the same "container" on a DVD disk. The software then applies rules depending on whether the movie, for example, was to be played by a consumer appliance or computer, whether the particular apparatus has a backchannel (e.g., an on-line connection), the
15 national and/or other legal or geographic region in which the player is located and/or the movie is being displayed, and/or whether the apparatus has components capable of identifying and applying such rules. For example, some usage rules may apply when information is played by a consumer device, while other
20 rules may apply when played by a computer. The choice of rules

may be left up to the rightsholder(s) and/or other participants-- or some rules may be predetermined (e.g., based on the particular environment or application). For example, film rightsholders may wish to limit copying and ensure that excerpts are not made

5 regardless of the context in which the property is played. This limitation might be applied only in certain legal or geographic areas. Alternatively, rightsholders of sound recordings may wish to enable excerpts of predetermined duration (e.g., no more than 20 seconds) and that these excerpts are not used to construct a new

10 commercial work. In some cases, governments may require that only "PG" versions of movies and/or the equivalent rating for TV programs may be played on equipment deployed in their jurisdiction, and/or that the applicable taxes, fees and the like are automatically calculated and/or collected if payments related to

15 content recorded on DVD is requested and/or performed (e.g., pay-per-use of a movie, game, database, software product, etc.; and/or orders from a catalog stored at least in part on DVD media, etc.).

In a microprocessor controlled (or augmented) digital

20 consumer appliance, such rules contemplated by the present

inventions can be enforced, for example, without requiring more than a relatively few additions to a central, controlling microprocessor (or other CPU, a IEEE 1394 port controller, or other content handling control circuitry), and/or making available

5 some ROM or flash memory to hold the necessary software. In addition, each ROM (or flash or other memory, which such memory may be securely connected to, or incorporated into, such control circuitry in a single, manufactured component) can, in one example, contain one or more digital documents or "certificate(s)"

10 that uniquely identifies a particular appliance, individual identity, jurisdiction, appliance class(es), and/or other chosen parameters. An appliance can, for example, be programmed to send a copy of a digital property to another digital device only in encrypted form and only inside a new, tamper-resistant "software container." The

15 container may also, for example, carry with it a code indicating that it is a copy rather than an original that is being sent. The device may also put a unique identifier of a receiving device and/or class of devices in the same secure container.

Consequently, for example, in one particular arrangement, the

20 copy may be playable only on the intended receiving device,

class(es) of devices, and/or devices in a particular region in one non-limiting example and rights related to use of such copy may differ according to these and/or other variables.

The receiving device, upon detecting that the digital property is indeed a copy, can, for example, be programmed not to make any additional copies that can be played on a consumer device and/or other class(es) of devices. If a device detects that a digital property is about to be played on a device and/or other class(es) of devices other than the one it was intended for, it can be programmed to refuse to play that copy (if desired).

The same restrictions applied in a consumer appliance can, for example, be enforced on a computer equipped to provide rights management protection in accordance with the present inventions. In this example, rules may specify not to play a certain film and/or other content on any device other than a consumer appliance and/or classes of appliances, for example. Alternatively, these same powerful capabilities could be used to specify different usage rules and payment schemes that would apply when played on a computer (and/or in other appliances and/or classes of appliances), as the rightsholder(s) may desire, for example,

different pricing based upon different geographic or legal locales where content is played.

In addition, if "backchannels" are present—for example, set-top boxes with bi-directional communications or computers
5 attached to networks—the present inventions contemplate electronic, independent delivery of new rules if desired or required for a given property. These new rules may, for example, specify discounts, time-limited sales, advertising subsidies, and/or other information if desired. As noted earlier, determination of these
10 independently delivered rules is entirely up to the rightsholder(s) and/or others in a given model.

The following are two specific examples of a few aspects of the present invention discussed above:

1. An Analog To Digital Copying Example

- 15 a) Bob has a VHS tape he bought (or rented) and wants to make a copy for his own use. The analog film has copy control codes embedded so that they do not interfere with the quality of the signal. Bob has a writable DVD appliance

that is equipped to provide rights management protection in accordance with the present invention. Bob's DVD recorder detects the control codes embedded in the analog signal (for example, such recorder may detect watermarks and/or fingerprints carrying rights related control and/or usage information), creates a new secure container to hold the content rules and describe the encoded film, and creates new control rules (and/or delivers to a secure VDE system for storage and reporting certain usage history related information such as user name, time, etc.) based on the analog control codes and/or other information it detected and that are then placed in the DigiBox and/or into a secure VDE installation data store such as a secure data base. Bob can play that copy back on his DVD appliance whenever he chooses.

5

10

15

- b) Bob gives the DVD disk he recorded to Jennifer who wishes to play it on computer that has a DVD drive. Her computer is equipped to provide rights management protection in accordance with the present invention. Her computer opens the "DigiBox," detects that this copy is being used on a device different from the one that recorded it (an unauthorized device) and refuses to play the copy.
- 5
- c) Bob gives the DVD disk to Jennifer as before, but now Jennifer contacts electronically a source of new rules and usage consequences, which might be the studio, a distributor, and/or a rights and permissions clearinghouse, (or she may have sufficient rights already on her player to play the copy). The source sends a DigiBox container to Jennifer with rules and consequences that permit playing the movie on her
- 10
- 15
- 20

computer while at the same time
charging her for use, even though the
movie was recorded on DVD by Bob
rather than by the studio or other value
5 chain participant.

2. A Digital To Analog Copying Example

- a) Jennifer comes home from work, inserts a
rented or owned DVD into a player connected
to, or an integral part of her TV, and plays the
10 disk. In a completely transparent way, the film
is decrypted, the format is converted from
digital to analog, and displayed on her analog
TV.
- b) Jennifer wishes to make a copy for her own
15 use. She plays the film on an DVD device
incorporating rights management protection in
accordance with the present invention, that
opens the DigiBox secure container, accesses
the control information, and decrypts the film.

She records the analog version on her VCR
which records a high-quality copy.

- 5 c) Jennifer gives the VCR copy to Doug who
wishes to make a copy of the analog tape for
his own use, but the analog control information
forces the recording VCR to make a lower-
quality copy, or may prevent copying. In
another non-limiting example, more
comprehensive rights management information
10 may be encoded in the analog output using the
methods and/or systems described in more
detail in the above referenced Van Wie and
Weber patent application.

In accordance with one aspect provided by this invention,
15 the same portable storage medium, such as a DVD, can be used
with a range of different, scaled protection environments
providing different protection capabilities. Each of the different
environments may be enabled to use the information carried by the
portable storage medium based on rights management techniques
20 and/or capabilities supported by the particular environment. For

example, a simple, inexpensive home consumer disk player may support copy protection and ignore more sophisticated and complex content rights the player is not equipped to enable. A more technically capable and/or secure platform (e.g., a personal computer incorporating a secure processing component possibly supported by a network connection, or a "smarter" appliance or device) may, for example, use the same portable storage medium and provide enhanced usage rights related to use of the content carried by the medium based on more complicated rights management techniques (e.g., requiring payment of additional compensation, providing secure extraction of selected content portions for excerpting or anthologizing, etc.). For example, a control set associated with the portable storage medium may accommodate a wide variety of different usage capabilities—with the more advanced or sophisticated uses requiring correspondingly more advanced protection and rights management enablement found on some platforms and not others. Lower-capability environments can, as another example, ignore (or not enable or attempt to use) rights in the control set that they don't understand, while higher-capability environments (having awareness of the

overall capabilities they provide), may, for example, enable the rights and corresponding protection techniques ignored by the lower-capability environments.

In accordance with another aspect provided by the invention, a media- and platform-independent security component can be scaled in terms of functionality and performance such that the elementary rights management requirements of consumer electronics devices are subsets of a richer collection of functionality that may be employed by more advanced platforms.

5 The security component can be either a physical, hardware component, or a "software emulation" of the component. In accordance with this feature, an instance of medium (or more correctly, one version of the content irrespective of media) can be delivered to customers independently of their appliance or

10 platform type with the assurance that the content will be protected. Platforms less advanced in terms of security and/or technical capabilities may provide only limited rights to use the content, whereas more advanced platforms may provide more expansive rights based on correspondingly appropriate security conditions

15 and safeguards.

20

In accordance with a further aspect provided by the present invention, mass-produced, inexpensive home consumer DVD players (such as those constructed, for example, with minimum complexity and parts count) can be made to be compatible with
5 the same DVDs or other portable storage media used by more powerful and/or secure platforms (such as, for example, personal computers) without degrading advanced rights management functions the storage media may provide in combination with the more powerful and/or secure platforms. The rights management
10 and protection arrangement provided and supported in accordance with this aspect of the invention thus supports inexpensive basic copy protection and can further serve as a commercial convergence technology supporting a bridging that allows usage in accordance with rights of the same content by a limited resource
15 consumer device while adequately protecting the content and further supporting more sophisticated security levels and capabilities by (a) devices having greater resources for secure rights management, and/or (b) devices having connectivity with other devices or systems that can supply further secure rights
20 management resources. This aspect of the invention allows

multiple devices and/or other systems that participate and work together in a permanently or temporarily connected network to share the rights management for at least one or more electronic events (e.g., managed through the use of protected processing environments such as described in Ginter et al.) occurring at a single, or across multiple nodes and further allows the rights associated with parties and/or groups using and/or controlling such multiple devices and/or other systems to be employed according to underlying rights related rules and controls, this allowing, for example, rights available through a corporate executive's device to be combined with or substitute for, in some manner, the rights of one or more subordinate corporate employees when their computing or other devices of these parties are coupled in a temporary networking relationship and operating in the appropriate context. In general, this aspect of the invention allows distributed rights management for DVD or otherwise packaged and delivered content that is protected by a distributed, peer-to-peer rights management. Such distributed rights management can operate whether the DVD appliance or other electronic information usage device is participating,

permanently or temporarily connected network and whether or not the relationships among the devices and/or other systems participating in the distributed rights management arrangement are relating temporarily or have a more permanent operating relationship. In this way, the same device may have different rights available depending on the context in which that device is operating (e.g., in a corporate environment such as in collaboration with other individuals and/or with groups, in a home environment internally and/or in collaboration with external one or more specified individuals and/or other parties, in a retail environment, in a classroom setting as a student where a student's notebook might cooperate in rights management with a classroom server and/or instructor PC, in a library environment where multiple parties are collaboratively employing differing rights to use research materials, on a factory floor where a hand held device works in collaboration with control equipment to securely and appropriately perform proprietary functions, and so on).

For example, coupling a limited resource device arrangement, such as a DVD appliance, with an inexpensive network computer (NC), or a personal computer (PC), may allow

an augmenting (or replacing) of rights management capabilities and/or specific rights of parties and/or devices by permitting rights management to be a result of a combination of some or all of the rights and/or rights management capabilities of the DVD

5 appliance and those of an Network or Personal Computer (NC or PC). Such rights may be further augmented, or otherwise modified or replaced by the availability of rights management capabilities provided by a trusted (secure) remote network rights authority.

10 These aspects of the present invention can allow the same device, in this example a DVD appliance, to support different arrays, e.g., degrees, of rights management capabilities, in disconnected and connected arrangements and may further allow

15 available rights to result from the availability of rights and/or rights management capabilities resulting from the combination of rights management devices and/or other systems. This may include one or more combinations of some or all of the rights available through the use of a "less" secure and/or resource poor device or system which are augmented, replaced, or otherwise

20 modified through connection with a device or system that is

“more” or “differently” secure and/or resource rich and/or possesses differing or different rights, wherein such connection employs rights and/or management capabilities of either and/or both devices as defined by rights related rules and controls that
5 describe a shared rights management arrangement.

In the latter case, connectivity to a logically and/or physically remote rights management capability can expand (by, for example, increasing the available secure rights management resources) and/or change the character of the rights available to
10 the user of the DVD appliance or a DVD appliance when such device is coupled with an NC, personal computer, local server, and/or remote rights authority. In this rights augmentation scenario, additional content portions may be available, pricing may change, redistribution rights may change (e.g., be expanded),
15 content extraction rights may be increased, etc.

Such “networking rights management” can allow for a combination of rights management resources of plural devices and/or other systems in diverse logical and/or physical relationships, resulting in either greater or differing rights through
20 the enhanced resources provided by connectivity with one or more

“remote” rights authorities. Further, while providing for increased and/or differing rights management capability and/or rights, such a connectivity based rights management arrangement can support multi-locational content availability, by providing for seamless
5 integration of remotely available content, for example, content stored in remote, Internet world wide web-based, database supported content repositories, with locally available content on one or more DVD discs.

In this instance, a user may experience not only increased or
10 differing rights but may use both local DVD content and supplementing content (i.e., content that is more current from a time standpoint, more costly, more diverse, or complementary in some other fashion, etc.). In such an instance, a DVD appliance and/or a user of a DVD appliance (or other device or system
15 connected to such appliance) may have the same rights, differing, and/or different rights applied to locally and remotely available content, and portions of local and remotely available content may themselves be subject to differing or different rights when used by a user and/or appliance. This arrangement can support an overall,
20 profound increase in user content opportunities that are seamlessly

integrated and efficiently available to users in a single content searching and/or usage activity by exploiting the rights management and content resources of plural, connected arrangements.

5 Such a rights augmenting remote authority may be directly coupled to a DVD appliance and/or other device by modem, or directly or indirectly coupled through the use of an I/O interface, such as a serial 1394 compatible controller (e.g., by communicating between a 1394 enabled DVD appliance and a
10 local personal computer that functions as a smart synchronous or asynchronous information communications interface to such one or more remote authorities, including a local PC or NC or server that serves as a local rights management authority augmenting and/or supplying the rights management in a DVD appliance).

15 In accordance with yet another aspect provided by this invention, rights provided to, purchased, or otherwise acquired by a participant and/or participant DVD appliance or other system can be exchanged among such peer-to-peer relating devices and/or other systems through the use of one or more permanently or
20 temporarily networked arrangements. In such a case, rights may be

bartered, sold, for currency, otherwise exchanged for value, and/or
loaned so long as such devices and/or other systems participate in
a rights management system, for example, such as the Virtual
Distribution Environment described in Ginter, et al., and employ
5 rights transfer and other rights management capabilities described
therein. For example, this aspect of the present invention allows
parties to exchange games or movies in which they have
purchased rights. Continuing the example, an individual might
buy some of a neighbor's usage rights to watch a movie, or
10 transfer to another party credit received from a game publisher for
the successful superdistribution of the game to several
acquaintances, where such credit is transferred (exchanged) to a
friend to buy some of the friend's rights to play a different game a
certain number of times, etc. In accordance with yet another aspect
15 provided by this invention, content carried by a portable storage
medium such as a DVD is associated with one or more encryption
keys and a secure content identifier. The content itself (or
information required to use the content) is at least partially
cryptographically encrypted—with associated decryption keys
20 being required to decrypt the content before the content can be

used. The decryption keys may themselves be encrypted in the form of an encrypted key block. Different key management and access techniques may be used, depending on the platform.

In accordance with still yet another aspect provided by this invention, electronic appliances that "create" digital content (or even analog content) —e.g., a digital camera/video recorder or audio recorder—can be readily equipped with appropriate hardware and/or software so as to produce content that is provided within a secure container at the outset. For example, content recorded by a digital camera could be immediately packaged in a secure container by the camera as it is recording. The camera could then output content already packaged in a secure container(s). This could preclude the need to encapsulate the content at a later point in time or at a later production stage, thus, saving at least one production-process step in the overall implementation of electronic rights management in accordance with the present invention. Moreover, it is contemplated that the very process of "reading" content for use in the rights management environment might occur at many steps along a conventional production and distribution process (such as during editing and/or

the so called "pressing" of a master DVD or audio disk, for
example). Accordingly, another significant advantage of the
present invention is that rights management of content essentially
can be extended throughout and across each appropriate content
5 creation, editing, distribution, and usage stages to provide a
seamless content protection architecture that protects rights
throughout an entire content life cycle.

In one example embodiment, the storage medium itself
carries key block decryption key(s) in a hidden portion of the
10 storage medium not normally accessible through typical access
and/or copying techniques. This hidden key may be used by a
drive to decrypt the encrypted key block—such decrypted key
block then being used to selectively decrypt content and related
information carried by the medium. The drive may be designed in
15 a secure and tamper-resistant manner so that the hidden keys are
never exposed outside of the drive to provide an additional
security layer.

In accordance with another example embodiment, a video
disk drive may store and maintain keys used to decrypt an
20 encrypted key block. The key block decryption keys may be

stored in a drive key store, and may be updatable if the video disk drive may at least occasionally use a communications path provided, for example, by a set top box, network port or other communications route.

5 In accordance with a further example embodiment, a virtual distribution environment secure node including a protected processing environment such as a hardware-based secure processing unit may control the use of content carried by a portable storage medium such as a digital video disk in accordance
10 with control rules and methods specified by one or more secure containers delivered to the secure node on the medium itself and/or over an independent communications path such as a network.

Certain conventional copy protection for DVD currently
15 envisions CGMA copy protection control codes combined with certain encryption techniques first proposed apparently by Matsushita Corporation. Notwithstanding the limited benefits of this approach to digital property protection, the present invention is capable of providing a supplementary, compatible, and far more
20 comprehensive rights management system while also providing

additional and/or different options and solutions. The following are some additional examples of advantageous features provided in accordance with the inventions:

- 5 • Strong security to fully answer content supplier needs.

- 10 • Value chain management automation and efficiencies including distributed rights protection, "piece of the tick" payment disaggregation to value chain participants, cost-effective micro-transaction management, and superdistribution, including offline micropayment and microtransaction support for at least occasionally connected devices.

- 15 • Simplified, more efficient channel management including support for the use of the same content deliverable on limited resource, greater resource, standalone, and/or connected devices.

- 20 • Can be used with any medium and application type and/or all forms of content and content models -- not just compressed video and sound as in some prior techniques and supports the use of copies of the same or materially the

5 same content containers across a wide variety
of media delivery systems (e.g., broadcast,
Internet repository, optical disc, etc) for
operation on a wide variety of different
electronic appliances (e.g., digital cameras,
digital editing equipment, sound recorders,
sound editing equipment, movie theater
projectors, DVD appliances, broadcast tape
players, personal computers, smart televisions,
10 etc).

- 15 • Asset management and revenue and/or other
consideration maximizing through important
new content revenue and/or other consideration
opportunities and the enhancement of value
chain operating efficiencies.
- 20 • Is capable of providing 100% compatibility
with the other protection techniques such as,
for example, CGMA protection codes and/or
Matsushita data scrambling approaches to
DVD copy protection.
- Can be employed with a variety of existing
data scrambling or protection systems to
provide very high degrees of compatibility
and/or level of functionality.

- Allows DVD technology to become a reusable, programmable, resource for an unlimited variety of entertainment, information commerce, and cyberspace business models.

- 5 • Enables DVD drive and/or semiconductor component manufacturers and/or distributors and/or other value adding participants to become providers of, and rights holders in, the physical infrastructure of the emerging,
10 connected world of the Internet and Intranets where they may charge for the use of a portion (e.g., a portion they provided) of the distributed, physical infrastructure as that portion participates in commercial networks. Such manufacturers and/or distributors and/or
15 other value adding participants can enjoy the revenue benefits resulting from participation in a “piece of the tick” by receiving a small portion of the revenue received as a result of a
20 participating transaction.

- Provides automated internationalization, regionalization, and rights management in that:
 - DVD content can be supplied with arrays of different rule sets for

automatic use depending on rights and
identity of the user; and

-- Societal rights, including taxes, can be
handled transparently.

5 In addition, the DVD rights management method and
apparatus of the present invention provides added benefits to
media recorders/publishers in that it:

- Works with a current "keep honest people
honest" philosophy.
- 10 • Can provide 100% compatibility with other
protection schemes such as for example,
Matsushita data scrambling and/or CGMA
encoded discs.
- 15 • Can work with and/or supplement other
protection schemes to provide desired degree
and/or functionality, or can be used in addition
to or instead of other approaches to provide
additional and/or different functionality and
features.

- Provides powerful, extensible rights management that reaches beyond limited copy protection models to rights management for the digitally convergent world.
- 5 • Empowers recording/publishing studios to create sophisticated asset management tools.
- Creates important business opportunities through controlled use of studio properties in additional multimedia contexts.
- 10 • Uniquely ties internationalization, regionalization, superdistribution, repurposing, to content creation processes and/or usage control.

Other aspects of the present invention provide benefits to
15 other types of rightsholders, such as for example:

- Persistent, transparent protection of digital content—globally, through value chain and process layers.
- Significant reduction in revenue loss from
20 copying and pass-along.

- Converts "pass-along," copying, and many forms of copyright infringement from a strategic business threat to a fundamental business opportunity.
- 5 • A single standard for all digital content regardless of media and/or usage locality and other rights variables.
- Major economies of scale and/or scope across industries, distribution channels, media, and
10 content type.
- Can support local usage governance and auditing within DVD players allowing for highly efficient micro-transaction support, including multiparty microtransactions and
15 transparent multiparty microtransactions.
- Empowers rightsholders to employ the broadest range of pricing, business models, and market strategies—as they see fit.

Further aspects of the present invention which may prove
20 beneficial to DVD and other digital medium appliance
manufacturers are:

- Capable of providing bit for bit compatibility with existing discs.
- Content type independent.
- Media independent and programmable/reusable.
- Highly portable transition to next generation of appliances having higher density devices and/or a writable DVD and/or other optical media format(s).
- Participation in revenue flow generated using the appliance.
- Single extensible standard for all digital content appliances.
- Ready for the future "convergent" world in which many appliances are connected in the home using, as one example, IEEE 1394 interfaces or other means (e.g., some appliances will be very much like computers and some computers will be very much like appliances).

Aspects of the present inventions provide many benefits to computer and OS manufacturers such as for example:

- 5 • Implementation in computers as an extension to the operating system, via for example, at least one transparent plug-in, and does not require modifications to computer hardware and/or operating systems.
- Easy, seamless integration into operating systems and into applications.
- 10 • Extremely strong security, especially when augmented with "secure silicon" (i.e., hardware/firmware protection apparatus fabricated on chip).
- Transforms user devices into true electronic commerce appliances.
- 15 • Provides a platform for trusted, secure rights management and event processing.
- Programmable for customization to specialized requirements.

Additional features and advantages provided in accordance with the inventions include, for example:

- 5 • Information on the medium (for example, both properties and metadata) may be encrypted or not.

- 10 • Different information (for example, properties, metadata) may be encrypted using different keys. This provides greater protection against compromise, as well as supporting selective usage rights in the context of a sophisticated rights management system.

- 15 • There may be encrypted keys stored on the medium, although this is not required. These keys may be used to decrypt the protected properties and metadata. Encrypted keys are likely to be used because that allows more keying material for the information itself, while still keeping access under control of a single key.

- 20 • Multiple sets of encrypted keys may be stored on the medium, either to have different sets of keys associated with different information, or to allow multiple control regimes to use the

same information, where each control regime may use one or more different keys to decrypt the set of encrypted keys that it uses.

- 5 • To support the ability of the player to access rights managed containers and/or content, a decryption key for the encrypted keys may be hidden on the medium in one or more locations that are not normally accessible. The "not normally accessible" location(s) may be
10 physically enabled for drives installed in players, and disabled for drives installed in computers. The enablement may be different firmware, a jumper on the drive, etc.

- 15 • The ability of the player to access rights managed containers and/or content may also be supported by one or more stored keys inside the player that decrypts certain encrypted keys on the medium.

- 20 • Keys in a player may allow some players to play different properties than others. Keys could be added to, and/or deleted from the player by a network connection (e.g., to a PC, a cable system, and/or a modem connection to a source of new and/or additional keys and/or

key revocation information) or automatically loaded by "playing" a key distribution DVD.

- 5 • Controlling computer use may be supported by some or all of the same techniques that control player use of content and/or rights management information.

- 10 • Controlling computer use of content and/or rights management information may be supported by having a computer receive, through means of a trusted rights management system, one or more appropriate keys.

- 15 • A computer may receive additional keys that permit decryption of certain encrypted keys on the medium.

- 20 • A computer may receive additional keys that permit decryption of one or more portions of encrypted data directly. This may permit selective use of information on the medium without disclosing keys (e.g., a player key that decrypts any encrypted keys).

In accordance with further aspects provided by the present invention, a secure "software container" is provided that allows:

- Cryptographically protected encapsulation of content, rights rules, and usage controls.
- Persistent protection for transport, storage, and value chain management.
- 5 • Sophisticated rules interface architecture.

Elements can be delivered independently, such as new controls, for example, regarding discount pricing (e.g. sale pricing, specific customer or group discounts, pricing based on usage patterns, etc.) and/or other business model changes, can be

10 delivered after the property has been distributed (this is especially beneficial for large properties or physical distribution media (e.g., DVD, CD-ROM) since redistribution costs may be avoided and consumers may continue to use their libraries of discs). In addition, encrypted data can be located "outside" the container.

15 This can allow, for example, use of data stored independently from the controls and supports "streaming" content as well as "legacy" systems (e.g., CGMS).

BRIEF DESCRIPTION OF THE DRAWINGS

These and other features and advantages provided in accordance with these inventions may be better and more completely understood by referring to the following detailed description of presently preferred examples in conjunction with the drawings, of which:

Figure 1A shows example home consumer electronics equipment for using portable storage media such as digital video disks;

10 Figure 1B shows example secure node equipment for using the same portable storage media but providing more advanced rights management capabilities;

Figure 1C shows an example process for manufacturing protected optical disks;

15 Figure 2A shows an example architecture of the Figure 1A consumer electronics equipment;

Figure 2B shows an example architecture for the Figure 1B secure node equipment;

Figure 3 shows example data structures used by the Figure 1A equipment;

Figure 3A and 3B show example control set definitions;

Figures 4A and 4B show example usage techniques provided by the Figure 1A appliance;

Figure 5 shows example data structures used by the Figure 1B secure node for accessing information on the storage medium;

Figure 6 shows an example usage technique performed by the Figure 1B secure node;

Figure 7 is a block diagram illustrating an example of a special secure software container contained on a DVD;

Figure 8 is a block diagram illustrating an example of a secure container along with the video property content stored on a DVD medium;

Figure 9 is a block diagram illustrating another example of a standard container stored on a DVD medium including an additional container having a more complex rule arrangement for use, for example, with a secure node;

Figure 10 shows an example use of a DVD having a container (i.e., stored on the medium) with a DVD player provided with a secure rights management node, and also shows use of the same DVD with a DVD player that does not have a secure rights management node;

Figure 11 is a block diagram illustrating use of a DVD that does not have a container on a DVD player that is provided with rights management secure node in accordance with the present invention as compared with use of the same DVD with a DVD player that does not have a secure node;

Figures 12-14 show example network configurations; and

Figures 15A-15C show an example virtual rights process.

15 **DETAILED DESCRIPTION OF
PRESENTLY PREFERRED EXAMPLE
EMBODIMENTS**

Overall Example Digital Video Disk Usage System

Figure 1A shows example inexpensive mass-produced home consumer electronics equipment 50 for using information stored on a storage medium 100 such as a portable digitally-encoded optical disk (e.g., a digital video disk or "DVD").

Consumer equipment 50 includes a dedicated disk player 52, that
in some embodiments, may also have the capability to write
optical media (writeable DVD disks, or "DVD-RAM") for
example) as well, connected to a home color television set 54. A
5 remote control unit 56 may be used to control the disk player 52
and/or television set 54.

In one example, disk 100 may store a feature length motion
picture or other video content. Someone wishing to watch the
content stored on disk 100 may purchase or rent the disk, insert
10 the disk into player 52 and use remote control 56 (and/or controls
58 that may be provided on player 52) to control the player to play
back the content via home television set 54.

In some embodiments, remote control 56 (and/or controls
58 that may be provided on device 52) may be used to control the
15 recording of a movie, for example. Player 52 reads the digitized
video and audio information carried by disk 100, converts it into
signals compatible with home color television set 54, and provides
those signals to the home color television set.

In some embodiments, television set 54 (and/or a set top box) provide the video signals to be recorded by device 52 on writable optical media, DVD-RAM in one non-limiting example. Television set 54 produces images on screen 54a and produces
5 sounds through loudspeakers 54b based on the signals player 52 provides to the television set.

The same disk 100 may be used by a more advanced platform 60 shown in Figure 1B. Platform 60 may include, for example, a personal computer 62 connected to a display monitor
10 64, a keyboard 66, a mouse pointing device 68, and a loudspeaker 70. In this example, platform 60 may be able to play back the content stored on disk 100 in the same way as dedicated disk player 52, but may also be capable of more sophisticated and/or advanced uses of the content as enabled by the presence of secure
15 node 72 within the platform. (In some embodiments, platform 60 may also be able to record content on writable optical media, DVD-RAM, in one non-limiting example.) For example, it may be possible, using platform 60 and its secure node 72, to interactively present the motion picture or other content such that the user may
20 input choices via keyboard 66 and/or mouse pointing device 68

that, in real time, change the presentation provided via display 64 and loudspeaker 60.

As one example, the platform 60 user selects from options displayed on display 64 that cause the content presentation sequence to change (e.g., to provide one of a number of different endings, to allow the user to interactively control the flow of the images presented, etc.). Computer 62 may also be capable of using and manipulating digital data including for example computer programs and/or other information stored on disk 100 that player 52 cannot handle.

Secure node 72 provides a secure rights management facility that may, for example, permit more invasive or extensive use of the content stored on disk. For example, dedicated player 52 may prevent any copying of content stored by disk 100, or it may allow the content to be copied only once and never again. Platform 60 including secure node 72, on the other hand, may allow multiple copies of some or all of the same content—but only if certain conditions are met (e.g., the user of equipment 60 falls within a certain class of people, compensation at an agreed on rate is securely provided for each copy made, only certain excerpts of

the content are copied, a secure audit trail is maintained and reported for each copy so made, etc.). (In some embodiments, dedicated player 52 may send protected content only to devices authenticated as able to enforce securely rights management rules and usage consequences. In some embodiments, devices may authenticate using digital certificates, one non-limiting example being certificates conforming to the X.509 standard.) Hence, platform 60 including secure node 72 can, in this example, use the content provided by disk 100 in a variety of flexible, secure ways that are not possible using dedicated player 52—or any other appliance that does not include a secure node.

Example Secure Disk Creation and Distribution Process

Figure 1C shows an example secure process for creating a master multimedia DVD disk 100 for use with players 50, 60. In this example, a digital camera 350 converts light images (i.e., pictures) into digital information 351 representing one or a sequence of images. Digital camera 350 in this example includes a secure node 72A that protects the digital information 351 before it leaves camera 350. Such protection can be accomplished, for

example, by packaging the digital information within one or more containers and/or associating controls with the digital information.

In this example, digital camera 350 provides the protected digital image information 351 to a storage device such as, for example, a digital tape recorder 352. Tape recorder 352 stores the digital image information 351 (along with any associated controls) onto a storage medium such as magnetic tape cartridge 354 for example. Tape recorder 352 may also include a secure node 72B. Secure node 72B in this example can understand and enforce the controls that the digital camera secure node 72A applies to and/or associated with the digital information 351, and/or it may apply its own controls to the stored information.

The same or different tape recorder 352 may play back protected digital information 351 to a digital mixing board 356. Digital mixing board 356 may mix, edit, enhance or otherwise process the digital information 351 to generate processed digital information 358 representing one or a sequence of images. Digital mixing board 356 may receive additional inputs from other devices such as for example other tape recorders, other digital cameras, character generators, graphics generators, animators, or

any other image-based devices. Any or all of such devices may also include secure nodes 72 to protect the information they generate. In some embodiments, some of the digital information can be derived from equipment including a secure node, and other
5 digital information can be derived from equipment that has no secure node. In still other embodiments, some of the digital information provided to digital mixer 356 is protected and some is not protected.

Digital mixing board 356 may also include a secure node
10 72C in this example. The digital mixing board secure node 72C may enforce controls applied by digital camera secure node 72A and/or tape recorder secure node 72B, and/or it may add its own protections to the digital information 358 it generates.

In this example, an audio microphone 361 receives sound
15 and converts the sound into analog audio signals. The audio signals in this example are inputted to a digital audio tape recorder 362. In the example shown, tape recorder 362 and audio mixer 364 are digital devices. However, in other embodiments, one, the other or both of these devices may operate in the analog domain.
20 In the example shown, digital audio tape recorder 362 converts the

analog audio signals into digital information representing the sounds, and stores the digital information (and any associated controls) onto a tape 362.

In this example, audio tape recorder 362 includes a secure
5 node 72E that may associate controls with the information stored on tape 363. Such controls may be stored with the information on the tape 363. In another embodiment, microphone 361 may include its own internal secure node 72 that associates control information with the audio information (e.g., by
10 steganographically encoding the audio information with control information). The tape recorder 362 may enforce such controls applied by microphone 361.

Alternatively, microphone 361 may operate in the digital domain and provide digital representations of audio, perhaps
15 including control information supplied by secure node 72 optionally incorporated in microphone 361, directly to connected devices such as audio tape recorder 362. Digital representations may optionally be substituted for analog representations of any signals between the devices in the example Figure 1C.

The same or different tape recorder 362 may play back the information recorded on tape 363, and provide the information 366 to an audio mixer 364. Audio mixer 364 may edit, mix, or otherwise process the information 366 to produce information 368
5 representing one or a sequence of sounds. Audio mixer 364 may also receive inputs from other devices such as for example other tape recorders, other microphones, sound generators, musical synthesizers, or any other audio-based devices. Any or all of such devices may also include secure nodes 72 to protect the
10 information they generate. In some embodiments, some of the digital information is derived from equipment including a secure node, and other digital information is derived from equipment that has no secure node. In still other embodiments, some of the digital information provided to audio mixer 364 is protected and
15 some is not protected.

Audio mixer 364 in this example includes a secure node 72F that enforces the controls, if any, applied by audio tape recorder secure node 72E; and/or applies its own controls.

Digital image mixer 356 may provide digital information
20 358 to "DVD-RAM" equipment 360 that is capable of writing to

master disks 100 and/or to disks from which master disks may be created. Similarly, audio mixer 364 may provide digital information 368 to equipment 360. Equipment 360 records the image information 358 and audio information 368 onto master disk 100. In this example, equipment 360 may include a secure node 72D that enforces controls applied by digital camera secure node 72A, tape recorder secure node 72B, digital mixer secure node 72C audio tape recorder secure node 72E and/or audio mixer secure node 72F; and/or it may add its own protections to the digital information 358 it writes onto master disks 100. A disk manufacturer can then mass-produce disks 100(1)-100(N) based on the master disk 100 using conventional disk mass-production equipment for distribution through any channels (e.g., video and music stores, websites, movie theaters, etc.). Consumer appliances 50 shown in Figures 1A and 1B may play back the disks 100 – enforcing the controls applied to the information stored on the disks 100. Secure nodes 72 thus maintain end-to-end, persistent secure control over the images generated by digital camera 350 and the sounds generated by microphone 361 during the entire process of making, distributing and using disks 100.

In the Figure 1C example shown, the various devices may communicate with one another over so-called "IEEE 1394" high-speed digital serial busses. In this context, "IEEE 1394" refers to hardware and software standards set forth in the following

5 standards specification incorporated by reference herein: 1394-1995 IEEE Standard for a High Performance Serial Bus, No. 1-55937-583-3 (Institute of Electrical and Electronics Engineers 1995). This specification describes a high-speed memory mapped digital serial bus that is self-configuring, hot pluggable, low cost

10 and scalable. The bus supports isochronous and asynchronous transport at 100, 200 or 400 Mbps, and flexibly supports a number of different topologies. The specification describes a physical level including two power conductors and two twisted pairs for signalling. The specification further describes physical, link and

15 transaction layer protocols including serial bus management.

Alternatively, any other suitable electronic communication means may be substituted for the "IEEE 1394" medium shown in Figure 1C, including other wired media (e.g., Ethernet, universal serial bus), and/or wireless media based on radio-frequency (RF)

transmission, infra-red signals, and/or any other means and/or types of electronic communication.

Example Dedicated Player Architecture

Figure 2A shows an example architecture for dedicated player 52. In this example, player 52 includes a video disk drive 80, a controller 82 (e.g., including a microprocessor 84, a memory device such as a read only memory 86, and a user interface 88), and a video/audio processing block 90. Video disk drive 80 optically and physically cooperates with disk 100, and reads digital information from the disk. Controller 82 controls disk drive 80 based on program instructions executed by microprocessor 84 and stored in memory 86 (and further based on user inputs provided by user interface 88 which may be coupled to controls 58 and/or remote control unit 56). Video/audio processing block 90 converts digital video and audio information read by disk drive 80 into signals compatible with home color television set 54 using standard techniques such as video and audio decompression and the like. Video/audio processing block 90 may also insert a visual marking indicating the ownership and/or protection of the video program. Block 90 may also

introduce a digital marking indicating to a standard recording device that the content should not be recorded.

Example Secure Node Architecture

Figure 2B shows an example architecture for platform 60 shown in Figure 1B—which in this example is built around a personal computer 62 but could comprise any number of different types of appliances. In this example, personal computer 62 may be connected to an electronic network 150 such as the Internet via a communications block 152. Computer equipment 62 may include a video disk drive 80' (which may be similar or identical to the disk drive 80 included within example player 52). Computer equipment 62 may further include a microprocessor 154, a memory 156 (including for example random access memory and read only memory), a magnetic disk drive 158, and a video/audio processing block 160. Additionally, computer equipment 62 may include a tamper-resistant secure processing unit 164 or other protected processing environment. Secure node 72 shown in Figure 1B may thus be provided by a secure processing unit 164, software executing on microprocessor 154, or a combination of

the two. Different embodiments may provide secure node 72 using software-only, hardware-only, or hybrid arrangements.

Secure node 72 in this example may provide and support a a general purpose Rights Operating System employing reusable
5 kernel and rights language components. Such a commerce-enabling Rights Operating System provides capabilities and integration for advanced commerce operating systems of the future. In the evolving electronic domain, general purpose, reusable electronic commerce capabilities that all participants can
10 rely on will become as important as any other capability of operating systems. Moreover, a rights operating system that provides, among other things, rights and auditing operating system functions can securely handle a broad range of tasks that relate to a virtual distribution environment. A secure processing unit can,
15 for example, provide or support many of the security functions of the rights and auditing operating system functions. The other operating system functions can, for example, handle general appliance functions. The overall operating system may, for example, be designed from the beginning to include the rights and
20 auditing operating system functions plus the other operating

system functions, or the rights and auditing operating system functions may, in another example, be an add-on to a preexisting operating system providing the other operating system functions. Any or all of these features may be used in combination with the
5 invention disclosed herein.

Example Disk Data Structures and Associated Protections

Figure 3 shows some example data structures stored on disk 100. In this example, disk 100 may store one or more properties or other content 200 in protected or unprotected form. Generally,
10 in this example, a property 200 is protected if it is at least in part encrypted and/or associated information needed to use the property is at least in part encrypted and/or otherwise unusable without certain conditions having being met. For example,
15 property 200(1) may be completely or partially encrypted using conventional secure cryptographic techniques. Another property 200(2) may be completely unprotected so that it can be used freely without any restriction. Thus, in accordance with this example, disk 100 could store both a movie as a protected property 200(1)
20 and an unprotected interview with the actors and producers or a

"trailer" as unprotected property 200(2). As shown in this example, disk 100 may store any number of different properties 200 in protected or unprotected form as limited only by the storage capacity of the disk.

5 In one example, the protection mechanisms provided by disk 100 may use any or all of the protection (and/or other) structures and/or techniques described in the above-referenced Shear patents. The Shear patents describe, by way of non-exhaustive example, means for solving the problem of how to
10 protect digital content from unauthorized use. For example, the Shear patent specifications describe, among other things, means for electronically "overseeing" -- through distributed control nodes present in client computers -- the use of digital content. This includes means and methods for fulfilling the consequences
15 of any such use.

Non-limiting examples of certain elements described in the Shear patent specifications include:

(a) decryption of encrypted information,

- (b) metering,
- (c) usage control in response to a combination of derived metering information and rules set by content providers,
- 5 (d) securely reporting content usage information,
- (e) use of database technology for protected information storage and delivery,
- (f) local secure maintenance of budgets, including, for example, credit budgets,
- 10 (g) local, secure storage of encryption key and content usage information,
- (h) local secure execution of control processes, and
- (i) in many non-limiting instances, the use of optical media.

15 Any or all of these features may be used in combination in or with the inventions disclosed herein.

Certain of the issued Shear patents' specifications also involve database content being local and remote to users.

Database information that is stored locally at the end-user's system and complemented by remote, "on-line" database information, can, for example, be used to augment the local information, which in one example, may be stored on optical media (for example, DVD and/or CD-ROM). Special purpose semiconductor hardware can, for example, be used to provide a secure execution environment to ensure a safe and reliable setting for digital commerce activities.

The Shear patents also describe, among other things, database usage control enabled through the use of security, metering, and usage administration capabilities. The specifications describe, *inter alia*, a metering and control system in which a database, at least partially encrypted, is delivered to a user (e.g., on optical media). Non-limiting examples of such optical media may, for example, include DVD and CD-ROM. Subsequent usage can, for example, be metered and controlled in any of a variety of ways, and resulting usage information can be transmitted to a responsible party (as one example).

The Shear patent specifications also describe the generation of a bill in response to the transmitted information. Other

embodiments of the Shear patents provide, for example, unique information security inventions which involve, for example, digital content usage being limited based on patterns of usage such as the quantity of particular kinds of usage. These capabilities

5 include monitoring the “contiguousness,” and/or “logical relatedness” of used information to ensure that the electronic “conduct” of an individual does not exceed his or her licensed rights. Still other aspects of the Shear patents describe, among other things, capabilities for enabling organizations to securely

10 and locally manage electronic information usage rights. When a database or a portion of a database is delivered to a client site, some embodiments of the Shear patents provide, for example, optical storage means (non-exhaustive examples of which include DVD and CD-ROM) as the mechanism of delivery. Such storage

15 means can store, for example, a collection of video, audio, images, software programs, games, etc., in one example, on optical media, such as DVD and/or CD-ROM, in addition to other content such as a collection of textual documents, bibliographic records, parts catalogs, and copyrighted or uncopied materials of all kinds.

Any or all of these features may be used in the embodiments herein.

One specific non-limiting embodiment could, for example, involve a provider who prepares a collection of games. The provider prepares a database "index" that stores information pertaining to the games, such as for example, the name, a description, a creator identifier, the billing rates, and the maximum number of times or total elapsed time each game may be used prior to a registration or re-registration requirement. Some or all of this information could be stored in encrypted form, in one example, on optical media, non-limiting examples of which include DVD and CD-ROM. The provider may then encrypt some or all portions of the games such that a game could not be used unless one or more encrypted portions were decrypted. Typically, decryption would not occur unless provider specified conditions were satisfied, in one example, unless credit was available to compensate for use and audit information reflecting game usage was being stored. The provider could determine, for example: which user activities he or she would allow, whether to meter such activities for audit and/or control purposes, and what, if any, limits

would be set for allowed activities. This might include, for example, the number of times that a game is played, and the duration of each play. Billing rates might be discounted, for example, based on total time of game usage, total number of
5 games currently registered for use, or whether the customer was also registered for other services available from the same provider, etc.

In the non-limiting example discussed above, a provider might, for example, assemble all of the prepared games along with
10 other, related information, and publish the collection on optical media, non-limiting examples of which include CD-ROM and/or DVD. The provider might then distribute this DVD disk to prospective customers. The customers could then select the games they wish to play, and contact the provider. The provider, based
15 on its business model, could then send enabling information to each authorized customer, such as for example, including, or enabling for use, decryption keys for the encrypted portion of the selected games (alternatively, authorization to use the games may have arrived with the DVD and/or CD-ROM disk, or might be
20 automatically determined, based on provider set criteria, by the

user's secure client system, for example, based on a user's participation in a certified user class). Using the user's client decryption and metering mechanism the customer could then make use of the games. The mechanism might then record usage information, such as for example, the number of times the game was used, and, for example, the duration of each play. It could periodically transmit this information the game provider, thus substantially reducing the administration overhead requirements of the provider's central servers. The game provider could receive compensation for use of the games based upon the received audit information. This information could be used to either bill their customers or, alternatively, receive compensation from a provider of credit.

Although games provide one convenient, non-limiting example, many of these same ideas can be easily applied to all kinds of content, all kinds of properties, including, by way of non-limiting examples:

- video,
- digitized movies,

- audio,
- images,
- multimedia,
- software,
- 5 • games,
- any other kind of property
- any combination of properties.

Other non-limiting embodiments of the Shear patent

10 specifications support, for example, securely controlling different kinds of user activities, such as displaying, printing, saving electronically, communicating, etc. Certain aspects further apply different control criteria to these different usage activities. For example, information that is being browsed may be distinguished
15 from information that is read into a host computer for the purpose of copying, modifying, or telecommunicating, with different cost rates being applied to the different activities (so that, for example,

the cost of browsing can be much less than the cost of copying or printing).

The Shear patent specifications also, for example, describe management of information inside of organizations by both publishers and the customer. For example, an optional security system can be used to allow an organization to prevent usage of all or a portion of an information base unless the user enters his security code. Multiple levels of security codes can be supported to allow restriction of an individual's use according to his security authorization level. One embodiment can, for example, use hardware in combination with software to improve tamper resistance, and another embodiment could employ an entirely software based system. Although a dedicated hardware/software system may under certain circumstances provide assurance against tampering, techniques which may be implemented in software executing on a non-dedicated system may provide sufficient tamper resistance for some applications. Any or all of these features may be used in combination with the technology disclosed in this patent specification.

Figures 3 Disks May Also Store Metadata, Controls and Other Information

In this example, disk 100 may also store "metadata" in protected and/or unprotected form. Player 52 uses metadata 202 to assist in using one or more of the properties 200 stored by disk 100. For example, disk 100 may store one metadata block 202(1) in unprotected form and another metadata block 202(2) in protected form. Any number of metadata blocks 202 in protected and/or unprotected form may be stored by disk 100 as limited only by the disk's storage capacity. In this example, metadata 202 comprises information used to access properties 200. Such metadata 202 may comprise, for example, frame sequence or other "navigational" information that controls the playback sequence of one or more of the properties 200 stored on disk 100. As one example, an unprotected metadata block 202 may access only selected portions of a protected property 200 to generate an abbreviated "trailer" presentation, while protected metadata block 202 may contain the frame playback sequence for the entire video presentation of the property 200. As another example, different metadata blocks 202 may be provided for different "cuts" of the

same motion picture property 200 (e.g., an R-rated version, a PG-rated version, a director's cut version, etc.).

In this example, disk 100 may store additional information for security purposes. For example, disk 100 may store control
5 rules in the form of a control set 204—which may be packaged in the form of one or more secure containers 206. Commerce model participants can securely contribute electronic rules and controls that represent their respective “electronic” interests. These rules and controls extend a “Virtual Presence™” through which the
10 commerce participants may govern remote value chain activities according to their respective, mutually agreed to rights. This Virtual Presence may take the form of participant specified electronic conditions (e.g., rules and controls) that must be satisfied before an electronic event may occur. These rules and
15 controls can be used to enforce the party’s rights during “downstream” electronic commerce activities. Control information delivered by, and/or otherwise available for use with, VDE content containers may, for example, constitute one or more “proposed” electronic agreements which manage the use and/or
20 consequences of the use of such content and which can enact the

terms and conditions of agreements involving multiple parties and their various rights and obligations.

The rules and controls from multiple parties can be used, in one example, to form aggregate control sets (“Cooperative Virtual Presence™”) that ensure that electronic commerce activities will be consistent with the agreements amongst value chain participants. These control sets may, for example, define the conditions which govern interaction with protected digital content (disseminated digital content, appliance control information, etc.).

10 These conditions can, for example, be used to control not only digital information use itself, but also the consequences of such use. Consequently, the individual interests of commerce participants are protected and cooperative, efficient, and flexible electronic commerce business models can be formed. These

15 models can be used in combination with the present invention.

Disks May Store Encrypted Information

Disk 100 may also store an encrypted key block 208. In this example, disk 100 may further store one or more hidden keys 210. In this example, encrypted key block 208 provides one or more

20 cryptographic keys for use in decrypting one or more properties

200 and/or one or more metadata blocks 202. Key block 208 may provide different cryptographic keys for decrypting different properties 200 and/or metadata blocks 202, or different portions of the same property and/or metadata block. Thus, key block 208
5 may comprise a large number of cryptographic keys, all of which are or may be required if all of the content stored by disk 100 is to be used. Although key block 208 is shown in Figure 3 as being separate from container 206, it may be included within or as part of the container if desired.

10 Cryptographic key block 208 is itself encrypted using one or more additional cryptographic keys. In order for player 52 to use any of the protected information stored on disk 100, it must first decrypt corresponding keys within the encrypted key block 208—and then use the decrypted keys from the key block to
15 decrypt the corresponding content.

In this example, the keys required to decrypt encrypted key block 208 may come from several different (possibly alternative) sources. In the example shown in Figure 3, disk 100 stores one or more decryption keys for decrypting key block 208 on the medium
20 itself in the form of a hidden key(s) 210. Hidden key(s) 210 may

be stored, for example, in a location on disk 100 not normally accessible. This "not normally accessible" location could, for example, be physically enabled for drives 80 installed in players 52 and disabled for drives 80' installed in personal computers 62.

5 Enablement could be provided by different firmware, a jumper on drive 80, etc. Hidden key(s) 210 could be arranged on disk 100 so that any attempt to physically copy the disk would result in a failure to copy the hidden key(s). In one example a hidden key(s) could be hidden in the bit stream coding sequences for one or

10 more blocks as described by J. Hogan (Josh Hogan, "DVD Copy Protection," presentation to DVD copy protect technical meeting #4, 5/30/96, Burbank, CA.)

Alternatively, and/or in addition, keys required to decrypt encrypted key block 208 could be provided by disk drive 80. In

15 this example, disk drive 80 might include a small decryption component such as, for example, an integrated circuit decryption engine including a small secure internal key store memory 212 having keys stored therein. Disk drive 80 could use this key store 212 in order to decrypt encrypted key block 208 without exposing

20 either keys 212 or decrypted key block 208—and then use the

decrypted key from key block 208 to decrypt protected content
200, 202.

Disks May Store and/or Use Secure Containers

In yet another example, the key(s) required to decrypt
5 protected content 200, 202 is provided within secure container
206. Figure 3A shows a possible example of a secure container
206 including information content 304 (properties 200 and
metadata 202 may be external to the container—or alternatively,
most or all of the data structures stored by video disk 100 may be
10 included as part of a logical and/or actual protected container).
The control set 204 shown in Figure 3 may comprise one or more
permissions record 306, one or more budgets 308 and/or one or
more methods 310 as shown in Figure 3A. Figure 3B shows an
example control set 204 providing one or more encryption keys
15 208, one or more content identifiers 220, and one or more controls
222. In this example, different controls 222 may apply to different
equipment and/or classes of equipment such as player 52 and/or
computer equipment 62 depending upon the capabilities of the
particular platform and/or class of platform. Additionally,
20 controls 220 may apply to different ones of properties 200 and/or

different ones of metadata blocks 202. For example, a control 222(1) may allow property 200(1) to be copied only once for archival purposes by either player 52 or computer equipment 62. A control 222(2) (which may be completely ignored by player 52 because it has insufficient technical and/or security capabilities but which may be useable by computer equipment 62 with its secure node 72) may allow the user to request and permit a public performance of the same property 200(1) (e.g., for showing in a bar or other public place) and cause the user's credit or other account to be automatically debited by a certain amount of compensation for each showing. A third control 222(3) may, for example, allow secure node 72 (but not player 52) to permit certain classes of users (e.g., certified television advertisers and journalists) to extract or excerpt certain parts of protected property 200(1) for promotional uses. A further control 222(4) may, as another example, allow both video player 52 and secure node 72 to view certain still frames within property 200(1)—but might allow only secure node 72 to make copies of the still frames based on a certain compensation level.

**Example Disks and/or System May Make Use of
Trusted Infrastructure**

Controls 222 may contain pointers to sources of additional control sets for one or more properties, controls, metadata, and/or other content on the optical disk. In one example, these additional controls may be obtained from a trusted third party, such as a rights and permissions clearinghouse and/or from any other value chain participant authorized by at least one rightsholder to provide at least one additional control set. This kind of rights and permissions clearinghouse is one of several distributed electronic administrative and support services that may be referred to as the "Distributed Commerce Utility," which, among other things, is an integrated, modular array of administrative and support services for electronic commerce and electronic rights and transaction management. These administrative and support services can be used to supply a secure foundation for conducting financial management, rights management, certificate authority, rules clearing, usage clearing, secure directory services, and other transaction related capabilities functioning over a vast electronic network such as the Internet and/or over organization internal Intranets, or even in-home networks of electronic appliances. Non-

limiting examples of these electronic appliances include at least occasionally connected optical media appliances, examples of which include read-only and/or writable DVD players and DVD drives in computers and convergent devices, including, for
5 example, digital televisions and settop boxes incorporating DVD drives.

These administrative and support services can, for example, be adapted to the specific needs of electronic commerce value chains in any number of vertical markets, including a wide variety
10 of entertainment applications. Electronic commerce participants can, for example, use these administrative and support services to support their interests, and/or they can shape and reuse these services in response to competitive business realities. Non-
exhaustive examples of electronic commerce participants include
15 individual creators, film and music studios, distributors, program aggregators, broadcasters, and cable and satellite operators.

The Distributed Commerce Utility can, for example, make optimally efficient use of commerce administration resources, and can, in at least some embodiments, scale in a practical fashion to

optimally accommodate the demands of electronic commerce growth.

The Distributed Commerce Utility may, for example, comprise a number of Commerce Utility Systems. These
5 Commerce Utility Systems can provide a web of infrastructure support available to, and reusable by, the entire electronic community and/or many or all of its participants. Different support functions can, for example, be collected together in hierarchical and/or in networked relationships to suit various
10 business models and/or other objectives. Modular support functions can, for example, be combined in different arrays to form different Commerce Utility Systems for different design implementations and purposes. These Commerce Utility Systems can, for example, be distributed across a large number of
15 electronic appliances with varying degrees of distribution.

The "Distributed Commerce Utility" provides numerous additional capabilities and benefits that can be used in conjunction with the particular embodiments shown in the drawings of this application, non-exhaustive examples of which include:

- Enables practical and efficient electronic commerce and rights management.
- Provides services that securely administer and support electronic interactions and consequences.
- 5 • Provides infrastructure for electronic commerce and other forms of human electronic interaction and relationships.
- Optimally applies the efficiencies of modern distributed computing and networking.
- 10 • Provides electronic automation and distributed processing.
- Supports electronic commerce and communications infrastructure that is modular, programmable, distributed and optimally computerized.
- 15 • Provides a comprehensive array of capabilities that can be combined to support services that perform various administrative and support roles.

- Maximizes benefits from electronic automation and distributed processing to produce optimal allocation and use of resources across a system or network.
- 5 • Is efficient, flexible, cost effective, configurable, reusable, modifiable, and generalizable.
- Can economically reflect users' business and privacy requirements.
- Can optimally distribute processes -- allowing commerce models to be flexible, scaled to demand and to match
10 user requirements.
- Can efficiently handle a full range of activities and service volumes.
- Can be fashioned and operated for each business model, as a mixture of distributed and centralized processes.
- 15 • Provides a blend of local, centralized and networked capabilities that can be uniquely shaped and reshaped to meet changing conditions.

- Supports general purpose resources and is reusable for many different models; in place infrastructure can be reused by different value chains having different requirements.
- 5 • Can support any number of commerce and communications models.
- Efficiently applies local, centralized and networked resources to match each value chain's requirements.
- Sharing of common resources spreads out costs and maximizes efficiency.
- 10 • Supports mixed, distributed, peer-to-peer and centralized networked capabilities.
- Can operate locally, remotely and/or centrally.
- Can operate synchronously, asynchronously, or support both modes of operation.
- 15 • Adapts easily and flexibly to the rapidly changing sea of commercial opportunities, relationships and constraints of "Cyberspace."

Any or all of these features may be used in combination with the inventions disclosed herein.

The Distributed Commerce Utility provides, among other advantages, comprehensive, integrated administrative and support services for secure electronic commerce and other forms of electronic interaction. These electronic interactions supported by the Distributed Commerce Utility may, in at least some embodiments, entail the broadest range of appliances and distribution media, non-limiting examples of which include networks and other communications channels, consumer appliances, computers, convergent devices such as WebTV, and optical media such as CD-ROM and DVD in all their current and future forms.

Example Access Techniques

Figures 3, 4A and 4B show example access techniques provided by player 52. In this example, upon disk 100 being loaded into player disk drive 80 (Figure 4A, block 400), the player controller 82 may direct drive 80 to fetch hidden keys 210 from disk 100 and use them to decrypt some or all of the encrypted key block 208 (Figure 4A, block 402). In this example, drive 80 may

97

store the keys so decrypted without exposing them to player controller 82 (e.g., by storing them within key store 212 within a secure decryption component such as an integrated circuit based decryption engine) (Figure 4A, block 404). The player 52 may
5 control drive 80 to read the control set 204 (which may or may not be encrypted) from disk 100 (Figure 4A, block 406). The player microprocessor 82 may parse control set 204, ignore or discard those controls 222 that are beyond its capability, and maintain permissions and/or rights management information corresponding
10 to the subset of controls that it can enforce (e.g., the "copy once" control 222(1)).

Player 52 may then wait for the user to provide a request via control inputs 58 and/or remote control unit 56. If the control input is a copy request ("yes" exit to Figure 4A, decision block
15 408), then player microprocessor 84 may query control 222(1) to determine whether copying is allowed, and if so, under what conditions (Figure 4A, decision block 410). Player 52 may refuse to copy the disk 100 if the corresponding control 222(1) forbids copying ("no" exit to Figure 4A, decision block 410), and may
20 allow copying (e.g., by controlling drive 80 to sequentially access

all of the information on disk 100 and provide it to an output port not shown) if corresponding control 222(1) permits copying ("yes" exit to Figure 4A, decision block 410; block 412). In this example, player 52 may, upon making a copy, store an identifier associated with disk 100 within an internal, non-volatile memory (e.g., controller memory 86) or elsewhere if control 222(1) so requires. This stored disk identifier can be used by player 52 to enforce a "copy once" restriction (i.e., if the user tries to use the same player to copy the same disk more than once or otherwise as forbidden by control 222(1), the player can deny the request).

If the user requests one of properties 200 to be played or read ("yes" exit to Figure 4A, decision block 414), player controller 82 may control drive 80 to read the corresponding information from the selected property 200 (e.g., in a sequence as specified by metadata 202) and decrypt the read information as needed using the keys initially obtained from key block 208 and now stored within drive key storage 212 (Figure 4A, block 416).

Figure 4B is a variation on the Figure 4A process to accommodate a situation in which player 52 itself provides decryption keys for decrypting encrypted key block 208. In this

example, controller 82 may supply one or more decryption keys to drive 80 using a secure protocol such a Diffie-Hellman key agreement, or through use of a shared key known to both the drive and some other system or component to which the player 52 is or
5 once was coupled (Figure 4B, block 403). The drive 80 may use these supplied keys to decrypt encrypted key block 208 as shown in Figure 4A, block 404, or it may use the supplied keys to directly decrypt content such as protected property 200 and/or protected metadata 202(2).

10 As a further example, the player 52 can be programmed to place a copy it makes of a digital property such as a film in encrypted form inside a tamper-resistant software container. The software container may carry with it a code indicating that the digital property is a copy rather than an original. The sending
15 player 52 may also put its own unique identifier (or the unique identifier of an intended receiving device such as another player 52, a video cassette player or equipment 50) in the same secure container to enforce a requirement that the copy can be played only on the intended receiving device. Player 52 (or other
20 receiving device) can be programmed to make no copies (or no

additional copies) upon detecting that the digital property is a copy rather than an original. If desired, a player 52 can be programmed to refuse to play a digital property that is not packaged with the player's unique ID.

5 **Example Use of Analog Encoding Techniques**

In another example, more comprehensive rights management information may be encoded by player 52 in the analog output using methods for watermarking and/or fingerprinting. Today, a substantial portion of the “real world” is analog rather than digital. Despite the pervasiveness of analog signals, existing methods for managing rights and protecting copyright in the analog realm are primitive or non-existent. For example:

- 15 • Quality degradation inherent in multigenerational analog copying has not prevented a multi-billion dollar pirating industry from flourishing.

- Some methods for video tape copy and pay per view protection attempt to prevent any copying at all of commercially released content, or allow only one

generation of copying. These methods can generally be easily circumvented.

- Not all existing devices respond appropriately to copy protection signals.
- 5 • Existing schemes are limited for example to “copy/no copy” controls.
- Copy protection for sound recordings has not been commercially implemented.

A related problem relates to the conversion of information
10 between the analog and digital domains. Even if information is effectively protected and controlled initially using strong digital rights management techniques, an analog copy of the same information may no longer be securely protected.

For example, it is generally possible for someone to make
15 an analog recording of program material initially delivered in digital form. Some analog recordings based on digital originals are of quite good quality. For example, a Digital Versatile Disk

(“DVD”) player may convert a movie from digital to analog format and provide the analog signal to a high quality analog home VCR. The home VCR records the analog signal. A consumer now has a high quality analog copy of the original
5 digital property. A person could re-record the analog signal on a DVD-RAM. This recording will in many circumstances have substantial quality – and would no longer be subject to “pay per view” or other digital rights management controls associated with the digital form of the same content.

10 Since analog formats will be with us for a long time to come, rightsholders such as film studios, video rental and distribution companies, music studios and distributors, and other value chain participants would very much like to have significantly better rights management capabilities for analog film,
15 video, sound recordings and other content. Solving this problem generally requires a way to securely associate rights management information with the content being protected.

In combination with other rights management capabilities, watermarking and/or fingerprinting, may provide “end to end”

secure rights management protection that allows content providers
and rights holders to be sure their content will be adequately
protected -- irrespective of the types of devices, signaling formats
and nature of signal processing within the content distribution
5 chain. This "end to end" protection also allows authorized analog
appliances to be easily, seamlessly and cost-effectively integrated
into a modern digital rights management architecture.

Watermarking and/or fingerprinting may carry, for example,
control information that can be a basis for a Virtual Distribution
10 Environment ("VDE") in which electronic rights management
control information may be delivered over insecure (e.g., analog)
communications channels. This Virtual Distribution Environment
is highly flexible and convenient, accommodating existing and
new business models while also providing an unprecedented
15 degree of flexibility in facilitating ad hoc creation of new
arrangements and relationships between electronic commerce and
value chain participants -- regardless of whether content is
distributed in digital and/or analog formats.

Watermarking together with distributed, peer-to-peer rights management technologies provides numerous advantages, including, but not limited to:

- 5 • An indelible and invisible, secure technique for providing rights management information.

- An indelible method of associating electronic commerce and/or rights management controls with analog content such as film, video, and sound recordings.

- 10 • Persistent association of the commerce and/or rights management controls with content from one end of a distribution system to the other -- regardless of the number and types of transformations between signaling formats (for example, analog to digital, and digital to
15 analog).

- The ability to specify “no copy/ one copy/ many copies” rights management rules, and also more

complex rights and transaction pricing models (such as, for example, “pay per view” and others).

- 5 • The ability to fully and seamlessly integrate with comprehensive, general electronic rights management solutions.

- Secure control information delivery in conjunction with authorized analog and other non-digital and/or non-secure information signal delivery mechanisms.

- 10 • The ability to provide more complex and/or more flexible commerce and/or rights management rules as content moves from the analog to the digital realm and back.

- 15 • The flexible ability to communicate commerce and/or rights management rules implementing new, updated, or additional business models to authorized analog and/or digital devices.

Any or all of these features may be used in combination in and/or with the inventions disclosed in the present specification.

Briefly, watermarking and/or fingerprinting methods may, using “steganographical” techniques, substantially indelibly and substantially invisibly encode rights management and/or electronic commerce rules and controls within an information signal such as, for example, an analog signal or a digitized (for example, sampled) version of an analog signal, non-limiting examples of which may include video and/or audio data, that is then decoded and utilized by the local appliance. The analog information and stenographically encoded rights management information may be transmitted via many means, non-limiting examples of which may include broadcast, cable TV, and/or physical media, VCR tapes, to mention one non-limiting example.

Any or all of these techniques may be used in combination in accordance with the inventions disclosed herein.

Watermarking and/or fingerprinting methods enable at least some rights management information to survive transformation of the video and/or other information from analog to digital and from

Figure 5 shows an example arrangement permitting platform 60 including secure node 72 to have enhanced and/or different capabilities to use information and/or rights management information on disk 100, and Figure 6 shows an example access technique provided by the secure node. Referring to Figure 5, secure node 72 may be coupled to a network 150 whereas player 52 may not be—giving the secure node great additional flexibility in terms of communicating security related information such as audit trails, compensation related information such as payment requests or orders, etc. This connection of secure node 72 to network 150 (which may be replaced in any given application by some other communications technique such as insertion of a replaceable memory cartridge) allows secure node 72 to receive and securely maintain rights management control information such as an additional container 206' containing an additional control set 204'. Secure node 72 may use control set 204' in addition or in lieu of a control set 204 stored on disk 100. Secure node 72 may also maintain a secure cryptographic key store 212 that may provide cryptographic keys to be used in lieu of or in addition to any keys 208, 210 that may be stored on disk 100.

Because of its increased security and/or technical capabilities,
secure node 72 may be able to use controls 222 within control set
204 that player 52 ignores or cannot use—and may be provided
with further and/or enhanced rights and/or rights management
5 capabilities based on control set 204' (which the user may, for
example, order specially and which may apply to particular
properties 200 stored on disk 100 and/or particular sets of disks).

Example Secure Node Access Techniques

The Figure 6 example access technique (which may be
10 performed by platform 60 employing secure node 72, for example)
involves, in this particular example, the secure node 72 fetching
property identification information 220 from disk 100 (Figure 6,
block 502), and then locating applicable control sets and/or rules
204 (which may be stored on disk 100, within secure node 72,
15 within one or more repositories the secure node 72 accesses via
network 150, and/or a combination of any or all of these
techniques) (Figure 6, block 504). Secure node 72 then loads the
necessary decryption keys and uses them to decrypt information as
required (Figure 6, block 506). In one example, secure node 72
20 obtains the necessary keys from secure containers 206 and/or 206'

and maintains them within a protected processing environment such as SPU 164 or a software-emulated protected processing environment without exposing them externally of that environment. In another example, the secure node 72 may load
5 the necessary keys (or a subset of them) into disk drive 82' using a secure key exchange protocol for use by the disk drive in decrypting information much in the same manner as would occur within player 52 in order to maintain complete compatibility in drive hardware.

10 Secure node 72 may monitor user inputs and perform requested actions based on the particular control set 204, 204'. For example, upon receiving a user request, secure node 72 may query the control set 204, 204' to determine whether it (they) permits the action the user has requested (Figure 6, block 508) and, if
15 permitted, whether conditions for performing the requested operation have been satisfied (Figure 6, block 510). In this example, secure node 72 may effect the operations necessary to satisfy any such required conditions such as by, for example, debiting a user's locally-stored electronic cash wallet, securely
20 requesting an account debit via network 150, obtaining and/or

checking user certificates to ensure that the user is within an appropriate class or is who he or she says he is, etc.—using network 150 as required (Figure 6, block 510). Upon all necessary conditions being satisfied, secure node 72 may perform the

5 requested operation (and/or enable microprocessor 154 to perform the operation) (e.g., to release content) and may then generate secure audit records which can be maintained by the secure node and/or reported at the time or later via network 150 (Figure 6, block 512).

10 If the requested operation is to release content (e.g., make a copy of the content), platform 60 (or player 52 in the example above) may perform the requested operation based at least in part on the particular controls that enforce rights over the content. For example, the controls may prevent platform 60 from releasing

15 content except to certain types of output devices that cannot be used to copy the content, or they may release the content in a way that discourages copying (e.g., by "fingerprinting" the copy with an embedded designation of who created the copy, by intentionally degrading the released content so that any copies

20 made from it will be inferior, etc.). As one specific example, a

video cassette recorder (not shown) connected to platform 60 may be the output device used to make the copy. Because present generations of analog devices such as video cassette recorders are incapable of making multigenerational copies without significant loss in quality, the content provider may provide controls that permit content to be copied by such analog devices but not by digital devices (which can make an unlimited number of copies without quality loss). For example, platform 60 may, under control of digital controls maintained by secure node 72, release content to the video cassette recorder only after the video cassette recorder supplies the platform a digital ID that designates the output device as a video cassette recorder -- and may refuse to provide any output at all unless such a digital ID identifying the output device as a lower quality analog device is provided.

15 Additionally or in the alternative, platform 60 may intentionally degrade the content it supplies to the video cassette recorder to ensure that no acceptable second-generation copies will be made. In another example, more comprehensive rights management information may be encoded by platform 60 in the analog output

20 using watermarking and/or fingerprinting.

Additional Examples of Secure Container Usage

Figure 7 shows a basic example of a DVD medium 700 containing a kind of secure container 701 for use in DVDs in accordance with the present invention. As shown in this example, container 701 ("DigiBox for DVDs") could be a specialized version of a "standard" container tailored especially for use with DVD and/or other media, or it could, alternatively (in an arrangement shown later in Figure 8), be a fully "standard" container. As shown in this example, the specialized container 701 incorporates features that permit it to be used in conjunction with content information, metadata, and cryptographic and/or protection information that is stored on the DVD medium 700 in the same manner as would have been used had container 701 not been present. Thus, specialized container 701 provides compatibility with existing data formats and organizations used on DVDs and/or other media. In addition, a specialized container 701 can be tailored to support only those features necessary for use in support of DVD and/or other media, so that it can be processed and/or manipulated using less powerful or less expensive computing resources than would be required for complete support of a "standard" container object.

In this example, specialized "DVD only" container 701 includes a content object (a property) 703 which includes an "external reference" 705 to video title content 707, which may be stored on the DVD and/or other medium in the same manner as would have been used for a medium not including container 701. The video title content 707 may include MPEG-2 and/or AC-3 content 708, as well as scrambling (protection) information 710 and header, structure and/or meta data 711. External reference 705 contains information that "designates" (points to, identifies, and/or describes) specific external processes to be applied/executed in order to use content and other information not stored in container 701. In this example, external reference 705 designates video title content 707 and its components 708, 710, and 711. Alternatively, container 701 could store some or all of the video title content in the container itself, using a format and organization that is specific to container 701, rather than the standard format for the DVD and/or other medium 700.

In this example, container 701 also includes a control object (control set) 705 that specifies the rules that apply to use of video title content 707. As indicated by solid arrow 702, control object

705 "applies to" content object (property) 703. As shown in this example, rule 704 can specify that protection processes, for example CGMA or the Matsushita data scrambling process, be applied, and can designate, by external reference 709 contained in
5 rule 704, data scrambling information 710 to be used in carrying out the protection scheme. The shorthand "do CGMA" description in rule 704 indicates that the rule requires that the standard CGMA protection scheme used for content on DVD media is to be used in conjunction with video title content 707, but a different example
10 could specify arbitrary other rules in control object 705 in addition to or instead of the "do CGMA" rule, including other standard DVD protection mechanisms such as the Matsushita data scrambling scheme and/or other rights management mechanisms. External reference 709 permits rule 704 to be based on protection
15 information 710 that is stored and manipulated in the same format and manner as for a DVD medium that does not incorporate container 701 and/or protection information that is meaningful only in the context of processing container 701.

Figure 8 shows a example of a DVD medium 800
20 containing a "standard" secure container 801. In this example, the

"standard" container provides all of the functionality (if desired) of the Figure 7 container, but may offer additional and/or more extensive rights management and/or content use capabilities than available on the "DVD only" container (e.g., the capacity to
5 operate with various different platforms that use secure nodes).

Figure 9 shows a more complex example of DVD medium 800 having a standard container 901 that provides all of the functionality (if desired) of the Figure 7 container, and that can function in concert with other standard containers 902 located
10 either on the same DVD medium or imported from another remote secure node or network. In this example, standard container 902 may include a supplementary control object 904 which applies to content object 903 of standard container 901. Also in this
example, container 902 may provide an additional rule(s) such as,
15 for example, a rule permitting/extending rights to allow up to a certain number (e.g., five) copies of the content available on DVD 900. This arrangement, for example, provides added flexibility in controlling rights management of DVD content between multiple platforms via access through "backchannels" such as via a set-top

box or other hardware having bi-directional communications capabilities with other networks or computers.

Additional Use of A DVD Disk With A Secure Container

5 Figure 10 illustrates the use of a "new" DVD disk—i.e., one that includes a special DVD secure container in the medium. This container may, in one example, be used in two possible use scenarios: a first situation in which the disk is used on an "old" player (DVD appliance, i.e., a DVD appliance that is not equipped
10 with a secure node to provide rights management in accordance with the present invention; and a second situation in which the disk is used on a "new" player—i.e., a DVD appliance which is equipped with a secure node to provide rights management in accordance with the present invention. In this example, a secure
15 node within the "new" player is configured with the necessary capabilities to process other copy protection information such as, for example, CGMA control codes and data scrambling formats developed and proposed principally by Matsushita.

For example, in the situation shown in Figure 10, the "new"
20 player (which incorporates a secure node in accordance with the

present invention) can recognize the presence of a secure container on the disk. The player may then load the special DVD secure container from the disk into the resident secure node. The secure node opens the container, and implements and/or enforces

5 appropriate rules and usage consequences associated with the content by applying rules from the control object. These rules are extremely flexible. In one example, the rules may, for example, call for use of other protection mechanisms (such as, for example, CGMA protection codes and Matsushita data scrambling) which

10 can be found in the content (or property) portion of the container.

In another example shown in Figure 10, the special DVD container on the disk still allows the "old" player to use to a predetermined limited amount content material which may be used in accordance with conventional practices.

15 **Example Use of A DVD Disk With No Secure Container**

Referring now to Figure 11, a further scenario is discussed. Figure 11 illustrates use of an "old" DVD disk with two possible use examples: a first example in which the disk is used on an "old"

20 player—i.e., a DVD appliance that is not equipped with a secure

node for providing rights management in accordance with the present invention—and a second example in which the disk is used on a "new" player (i.e., equipped with a secure node).

In the first case, the "old" player will play the DVD content in a conventional manner. In the second scenario, the "new" player will recognize that the disk does not have a container stored in the medium. It therefore constructs a "virtual" container in resident memory of the appliance. To do this, it constructs a container content object, and also constructs a control object containing the appropriate rules. In one particular example, the only applicable rule it need apply is to "do CGMA" -- but in other examples, additional and/or different rules could be employed. The virtual container is then provided to the secure node within the "new" player for implementing management of use rights in accordance with the present invention. Although not shown in Figures 10 and 11, use of "external references" may also be provided in both virtual and non-virtual containers used in the DVD context.

**Example Illustrative Arrangements for Sharing,
Brokering and Combining Rights When Operating in At Least
Occasionally Connected Scenarios**

5 As described above, the rights management resources of
several different devices and/or other systems can be flexibly
combined in diverse logical and/or physical relationships,
resulting for example in greater and/or differing rights. Such
rights management resource combinations can be effected through
10 connection to one or more remote rights authorities. Figures 12-
14 show some non-limiting examples of how rights authorities can
be used in various contexts.

For example, Figure 12 shows a rights authority broker
1000 connected to a local area network (LAN) 1002. LAN 1002
15 may connect to wide area network if desired. LAN 1002 provides
connectivity between rights authority broker 1000 and any number
of appliances such as for example a player 50, a personal
computer 60, a CD "tower" type server 1004. In the example
shown, LAN 1002 includes a modem pool (and/or network

protocol server, not shown)1006 that allows a laptop computer
1008 to connect to the rights authority broker 1000 via dial-up
lines 1010. Alternatively, laptop 1008 could communicate with
rights authority broker 1000 using other network and/or
5 communication means, such as the Internet and/or other Wide
Area Networks (WANs). A disk player 50A may be coupled to
laptop 1008 at the laptop location. In accordance with the
teachings above, any or all of devices shown in Figure 12 may
include one or more secure nodes 72.

10 Rights authority broker 1000 may act as an arbiter and/or
negotiator of rights. For example, laptop 1008 and associated
player 50A may have only limited usage rights when operating in
a stand-alone configuration. However, when laptop 1008 connects
to rights authority broker 1000 via modem pool 1006 and LAN
15 1002 and/or by other communication means, the laptop may
acquire different and/or expanded rights to use disks 100 (e.g.,
availability of different content portions, different pricing,
different extraction and/or redistribution rights, etc.) Similarly,
player 50, equipment 60 and equipment 1004 may be provided
20 with an enhanced and/or different set of disk usage rights through

communication with rights authority broker 1000 over LAN 1002.
Communication to and from rights authority broker 1000 is preferably secured through use of containers of the type disclosed in the above-referenced Ginter et al. patent specification.

5 Figure 13 shows another example use of a rights authority broker 1000 within a home environment. In this example, the laptop computer 1008 may be connected to a home-based rights authority broker 1000 via a high speed serial IEEE 1394 bus and/or by other electronic communication means. In addition,
10 rights authority broker 1000 can connect with any or all of:

- a high definition television 1100,
- one or more loudspeakers 1102 or other audio transducers,
- one or more personal computers 60,
- 15 • one or more set-top boxes 1030,
- one or more disk players 50,
- one or more other rights authority brokers 1000A-1000N
and

- any other home or consumer equipment or appliances.

Any or all of the equipment listed above may include a secure node 72.

Figure 14 shows another example use of a rights authority

5 broker 1000. In this example, rights authority broker 1000 is connected to a network 1020 such as a LAN, a WAN, the Internet, etc. Network 1020 may provide connectivity between rights authority broker 1000 and any or all of the following equipment:

- 10 • one or more connected or occasionally connected disk players 50A, 50B;
- one more networked computers 1022;
- one or more disk reader towers/servers 1004;
- one or more laptop computers 1008;
- 15 • one or more Commerce Utility Systems such as a rights and permissions clearinghouse 1024 (see Shear et al., “Trusted Infrastructure...” specification referenced above);

- one or more satellite or other communications uplinks
1026;
- one or more cable television head-ends 1028;
- one or more set-top boxes 1030 (which may be
5 connected to satellite downlinks 1032 and/or disk
players 50C);
- one or more personal computer equipment 60;
- one or more portable disk players 1034 (which may be
connected through other equipment, directly, and/or
10 occasionally unconnected;
- one or more other rights authority brokers 1000A-
1000N; and
- any other desired equipment.

Any or all of the above-mentioned equipment may
15 include one or more secure nodes 72. Rights authority
broker 1000 can distribute and/or combine rights for use by
any or all of the other components shown in Figure 14. For
example, rights authority broker 100 can supply further

secure rights management resources to equipment
connected to the broker via network 1020. Multiple
equipment shown in Figure 14 can participate and work
together in a permanently or temporarily connected network
5 1020 to share the rights management for a single node.
Rights associated with parties and/or groups using and/or
controlling such multiple devices and/or other systems can
be employed according to underlying rights related rules
and controls. As one example, rights available through a
10 corporate executive's laptop computer 1008 might be
combined with or substituted for, in some manner, the rights
of one or more subordinate corporate employees when their
computing or other devices 60 are coupled to network 1020
in a temporary networking relationship. In general, this
15 aspect of the invention allows distributed rights
management for DVD or otherwise packaged and delivered
content that is protected by a distributed, peer-to-peer rights
management. Such a distributed rights management can
operate whether the DVD appliance or other content usage
20 device is participating in a permanently or temporarily

connected network 1020, and whether or not the relationships among the devices and/or other systems participating in the distributed rights management arrangement are relating temporarily or have a more
5 permanent operating relationship.

For example, laptop computer 1008 may have different rights available depending on the context in which that device is operating. For example, in a general corporate environment such as shown in Figure 12, the laptop 1008 may have one set of rights.
10 However, the same laptop 1008 may be given a different set of rights when connected to a more general network 1020 in collaboration with specified individuals and/or groups in a corporation. The same laptop 1008 may be given a still different set of rights when connected in a general home environment such
15 as shown by example in Figure 13. The same laptop 1008 could be given still different rights when connected in still other environments such as, by way of non-limiting example:

- a home environment in collaboration with specified individuals and/or groups,

- a retail environment,
 - a classroom setting as a student,
 - a classroom setting in collaboration with an instructor, in a library environment,
- 5
- on a factory floor,
 - on a factory floor in collaboration with equipment enabled to perform proprietary functions, and so on.

As one more particular example, coupling a limited resource device arrangement such as a DVD appliance 50 shown in Figure 10 14 with an inexpensive network computer (NC) 1022 may allow an augmenting (or replacing) of rights management capabilities and/or specific rights of parties and/or devices by permitting rights management to be a result of a combination of some or all of the rights and/or rights management capabilities of the DVD 15 appliance and those of an Network or Personal Computer (NC or PC). Such rights may be further augmented, or otherwise modified or replaced by the availability of rights management capabilities provided by a trusted (secure) remote network rights authority 1000.

The same device, in this example a DVD appliance 50, can thus support different arrays, e.g., degrees, of rights management capabilities, in disconnected and connected arrangements and may further allow available rights to result from the availability of

5 rights and/or rights management capabilities resulting from the combination of rights management devices and/or other systems. This may include one or more combinations of some or all of the rights available through the use of a “less” secure and/or resource poor device or system which are augmented, replaced, or

10 otherwise modified through connection with a device or system that is “more” or “differently” secure and/or resource rich and/or possesses differing or different rights, wherein such connection employs rights and/or management capabilities of either and/or both devices as defined by rights related rules and controls that

15 describe a shared rights management arrangement.

In the latter case, connectivity to a logically and/or physically remote rights management capability can expand (by, for example, increasing the available secure rights management resources) and/or change the character of the rights available to

20 the user of the DVD appliance 50 or a DVD appliance when such

device is coupled with an NC 1022, personal computer 60, and/or
remote rights authority 1000. In this rights augmentation scenario,
additional content portions may be available, pricing may change,
redistribution rights may change (e.g., be expanded), content
5 extraction rights may be increased, etc.

Such “networking rights management” can allow for a
combination of rights management resources of plural devices
and/or other systems in diverse logical and/or physical
relationships, resulting in either greater or differing rights through
10 the enhanced resources provided by connectivity with one or more
“remote” rights authorities. Further, while providing for increased
and/or differing rights management capability and/or rights, such a
connectivity based rights management arrangement can support
multi-locational content availability, by providing for seamless
15 integration of remotely available content, for example, content
stored in remote, Internet world wide web-based, database
supported content repositories, with locally available content on
one or more DVD discs 100.

In this instance, a user may experience not only increased or
20 differing rights but may be able to use to both local DVD content

and supplementing content (i.e., content that is more current from a time standpoint, more costly, more diverse, or complementary in some other fashion, etc.). In such an instance, a DVD appliance 50 and/or a user of a DVD appliance (or other device or system 5 connected to such appliance) may have the same rights, differing, and/or different rights applied to locally and remotely available content, and portions of local and remotely available content may themselves be subject to differing or different rights when used by a user and/or appliance. This arrangement can support an overall, 10 profound increase in user content opportunities that are seamlessly integrated and efficiently available to users in a single content searching and/or usage activity.

Such a rights augmenting remote authority 1000 may be directly coupled to a DVD appliance 50 and/or other device by 15 modem (see item 1006 in Figure 12) and/or directly or indirectly coupled through the use of an I/O interface, such as a serial 1394 compatible controller (e.g., by communicating between a 1394 enabled DVD appliance and a local personal computer that functions as a smart synchronous or asynchronous information 20 communications interface to such one or more remote authorities,

including a local PC 60 or NC 1022 that serves as a local rights management authority augmenting and/or supplying the rights management in a DVD appliance) and/or by other digital communication means such as wired and/or wireless network connections.

Rights provided to, purchased, or otherwise acquired by a participant and/or participant DVD appliance 50 or other system can be exchanged among such peer-to-peer relating devices and/or other systems so long as they participate in a permanently or temporarily connected network. 1020. In such a case, rights may be bartered, sold, for currency, otherwise exchanged for value, and/or loaned so long as such devices and/or other systems participate in a rights management system, for example, such as the Virtual Distribution Environment described in Ginter, et al., and employ rights transfer and other rights management capabilities described therein. For example, this aspect of the present invention allows parties to exchange games or movies in which they have purchased rights. Continuing the example, an individual might buy some of a neighbor's usage rights to watch a movie, or transfer to another party credit received from a game

publisher for the successful superdistribution of the game to several acquaintances, where such credit is transferred (exchanged) to a friend to buy some of the friend's rights to play a different game a certain number of times, etc.

5 **Example Virtual Rights Process**

Figures 15A-15C shows an example of a process in which rights management components of two or more appliances or other devices establish a virtual rights machine environment associated with an event, operation and/or other action. The process may be initiated in a number of ways. In one example, an appliance user (and/or computer software acting on behalf of a user, group of users, and/or automated system for performing actions) performs an action with a first appliance (e.g., requesting the appliance to display the contents of a secure container, extract a portion of a content element, run a protected computer program, authorize a work flow process step, initiate an operation on a machine tool, play a song, etc.) that results in the activation of a rights management component associated with such first appliance (Figure 15A, block 1500). In other examples, the process may get started in response to an automatically generated event (e.g., based

on a time of day or the like), a random or pseudo-random event,
and/or a combination of such events with a user-initiated event.

Once the process begins, a rights management component
such as a secure node 72 (for example, an SPE and/or HPE as
5 disclosed in Ginter et al.) determines which rights associated with
such first appliance, if any, the user has available with respect to
such an action (Figure 15A, block 1502). The rights management
component also determines the coordinating and/or cooperating
rights associated with such an action available to the user located
10 in whole or in part on other appliances (Figure 15A, block 1502).

In one example, these steps may be performed by securely
delivering a request to a rights authority server 1000 that identifies
the first appliance, the nature of the proposed action, and other
information required or desired by such a rights authority server.

15 Such other information may include, for example:

- the date and time of the request,
- the identity of the user,
- the nature of the network connection,

- the acceptable latency of a response, etc.), and/or
- any other information.

In response to such a request, the rights authority server 1000 may return a list (or other appropriate structure) to the first
5 appliance. This list may, for example, contain the identities of other appliances that do, or may, have rights and/or rights related information relevant to such a proposed action.

In another embodiment, the first appliance may communicate (e.g., poll) a network with requests to other
10 appliances that do, or may, have rights and/or rights related information relevant to such proposed action. Polling may be desirable in cases where the number of appliances is relatively small and/or changes infrequently. Polling may also be useful, for example, in cases where functions of a rights authority server 1000
15 are distributed across several appliances.

The rights management component associated with the first appliance may then, in this example, check the security level(s) (and/or types) of devices and/or users of other appliances that do, or may, have rights and/or rights related information relevant to

such an action (Figure 15A, block 1506). This step may, for example, be performed in accordance with the security level(s) and/or device type management techniques disclosed in Sibert and Van Wie, and the user rights, secure name services and secure
5 communications techniques disclosed in Ginter et al. Device and/or user security level determination may be based, for example, in whole or in part on device and/or user class.

The rights management component may then make a decision as to whether each of the other appliance devices and/or
10 users have a sufficient security level to cooperate in forming the set of rights and/or rights related information associated with such an action (Figure 15A, block 1508). As each appliance is evaluated, some devices and/or users may have sufficient security levels, and others may not. In this example, if a sufficient security
15 level is not available ("No" exit to decision block 1508), the rights management component may create an audit record (for example, an audit record of the form disclosed in Ginter et al.) (Figure 15A, block 1510), and may end the process (Figure 15A, block 1512). Such audit record may be for either immediate transmission to a
20 responsible authority and/or for local storage and later

transmission, for example. The audit recording step may include, as one example, incrementing a counter that records security level failures (such as the counters associated with summary services in Ginter et al.)

- 5 If the devices and/or users provide the requisite security level (“Yes” exit to block 1508), the rights management component in this example may make a further determination based on the device and/or user class(es) and/or other configuration and/or characteristics (Figure 15B, block 1514).
- 10 Such determination may be based on any number of factors such as for example:
- the device is accessible only through a network interface that has insufficient throughput;
 - devices in such a class typically have insufficient
- 15 resources to perform the action, or relevant portion of the action, at all or with acceptable performance, quality, or other characteristics;

- the user class is inappropriate due to various conditions (e.g., age, security clearance, citizenship, jurisdiction, or any other class-based or other user characteristic); and/or
- other factors.

5 In one example, decision block 1514 may be performed in part by presenting a choice to the user that the user declines.

If processes within the rights management component determines that such device and/or user class(es) are inappropriate (“No” exit to block 1514), the rights management
10 component may write an audit record if required or desired (Figure 15B, block 1516) and the process may end (Figure 15B, block 1518).

If, on the other hand, the rights management component determines that the device and/or user classes are appropriate to
15 proceed (“Yes” exit to block 1514), the rights management component may determine the rights and resources available for performing the action on the first appliance and the other appliances acting together (Figure 15B, block 1520). This step may be performed, for example, using any or all of the method

processing techniques disclosed in Ginter et al. For example, method functions may include event processing capabilities that formulate a request to each relevant appliance that describes, in whole or in part, information related to the action, or portion of the action, potentially suitable for processing, in whole or in part, by such appliance. In this example, such requests, and associated responses, may be managed using the reciprocal method techniques disclosed in Ginter et al. If such interaction requires additional information, or results in ambiguity, the rights management component may, for example, communicate with the user and allow them to make a choice, such as making a choice among various available, functionally different options, and/or the rights management component may engage in a negotiation (for example, using the negotiation techniques disclosed in Ginter et al.) concerning resources, rights and/or rights related information.

The rights management component next determines whether there are sufficient rights and/or resources available to perform the requested action (Figure 15B, decision block 1522). If there are insufficient rights and/or resources available to perform the action (“No” exit to block 1522), the rights management component may

write an audit record (Figure 15B, block 1524), and end the process (Figure 15B, block 1526).

In this example, if sufficient rights and/or resources are available (“Yes” exit to block 1522), the rights management component may make a decision regarding whether additional events should be processed in order to complete the overall action (Figure 15B, block 1528). For example, it may not be desirable to perform only part of the overall action if the necessary rights and/or resources are not available to complete the action. If more events are necessary and/or desired (“Yes” exit to block 1528), the rights management component may repeat blocks 1520, 1522 (and potentially perform blocks 1524, 1526) for each such event.

If sufficient rights and/or resources are available for each of the events (“No” exit to block 1528), the rights management component may, if desired or required, present a user with a choice concerning the available alternatives for rights and/or resources for performing the action (Figure 15B, block 1530). Alternatively and/or in addition, the rights management component may rely on user preference information (and/or defaults) to “automatically” make such a determination on behalf

of the user (for example, based on the overall cost, performance, quality, etc.). In another embodiment, the user's class, or classes, may be used to filter or otherwise aid in selecting among available options. In still another embodiment, artificial intelligence
5 (including, for example, expert systems techniques) may be used to aid in the selection among alternatives. In another embodiment, a mixture of any or all of the foregoing (and/or other) techniques may be used in the selection process.

If there are no acceptable alternatives for rights and/or
10 resources, or because of other negative aspects of the selection process (e.g., a user presses a "Cancel" button in a graphical user interface, a user interaction process exceeds the available time to make such a selection, etc.), ("No" exit to block 1530) the rights management component may write an audit record (Figure 15B,
15 block 1532), and end the process (Figure 15B, block 1534).

On the other hand, if a selection process identifies one or more acceptable sets of rights and/or resources for performing the action and the decision to proceed is affirmative ("Yes" exit to block 1530), the rights management component may perform the
20 proposed action using the first appliance alone or in combination

with any additional appliances (e.g., a rights authority 1000, or any other connected appliance) based on the selected rights and/or resources (Figure 15C, block 1536). Such cooperative implementation of the proposed actions may include for example:

- 5 • performing some or all of the action with the first appliance;
- performing some or all of the action with one or more appliances other than the first appliance (e.g., a rights authority 1000 and/or some other appliance);
- 10 • performing part of the action with the first appliance and part of the action with one or more other appliances; or
- any combination of these.

For example, this step may be performed using the event processing techniques disclosed in Ginter et al.

- 15 As one illustrative example, the first appliance may have all of the resources necessary to perform a particular task (e.g., read certain information from an optical disk), but may lack the rights necessary to do so. In such an instance, the first appliance may

obtain the additional rights it requires to perform the task through the steps described above. In another illustrative example, the first appliance may have all of the rights required to perform a particular task, but it may not have the resources to do so. For
5 example, the first appliance may not have sufficient hardware and/or software resources available to it for accessing, processing or otherwise using information in certain ways. In this example, step 1536 may be performed in whole or in part by some other appliance or appliances based in whole or in part on rights
10 supplied by the first appliance. In still another example, the first appliance may lack both rights and resources necessary to perform a certain action, and may rely on one or more additional appliances to supply such resources and rights.

In this example, the rights management component may,
15 upon completion of the action, write one or more audit records (Figure 15C, block 1538), and the process may end (Figure 15C, block 1540).

* * * * *

An arrangement has been described which adequately satisfies current entertainment industry requirements for a low cost, mass-produceable digital video disk or other high capacity disc copy protection scheme but which also provides enhanced, 5 extensible rights management capabilities for more advanced and/or secure platforms and for cooperative rights management between devices of lessor, greater, and/or differing rights resources. While the invention has been described in connection with what is presently considered to be the most practical and 10 preferred embodiment, it is to be understood that the invention is not to be limited to the disclosed embodiment, but on the contrary, is intended to cover various modifications and equivalent arrangements included within the spirit and scope of the invention.

We Claim:

1. An electronic appliance including:

a disk use arrangement for at least one of (a) reading information from, and (b) writing information to, a digital versatile disk optical storage medium; and

a secure node coupled to the disk use arrangement, the secure node providing at least one rights management process.

2. An electronic appliance including:

a disk use arrangement for at least one of (a) reading information from, and (b) writing information to, a digital versatile disk optical storage medium; and

at least one processing arrangement coupled to the disk use arrangement, the processing arrangement selecting at least some control information associated with information recorded on the storage medium based at least in part on the class of the appliance and/or the user of the appliance.

3. A system as in claim 2 wherein the processing arrangement selects a subset of control information used on another appliance and/or class of appliance.
4. A system as in claim 2 wherein the processing arrangement selects different control information from the information selected by another appliance and/or class of appliance.
5. A system as in claim 2 wherein at least some of the control information comprises an analog signal.
6. A system as in claim 2 wherein at least some of the control information comprises digitally encoded information.
7. In an appliance capable of using digital versatile disks, a method including the following steps:

at least one of (a) reading information from, and (b) writing information to, a digital versatile disk optical storage medium; and

selecting at least some control information associated with information recorded on the storage medium based at least in part on the class of the appliance and/or the user of the appliance.

8. A method as in claim 7 wherein the selecting step includes the step of selecting a subset of control information used on another appliance and/or class of appliance.

9. A method as in claim 7 wherein the selecting step includes the step of selecting, from control information stored on the storage medium, a different set of control information than the control information selected by another appliance and/or class of appliance.

10. An electronic appliance including:

a disk use arrangement for reading information from a digital versatile disk optical storage medium; and

at least one processing arrangement coupled to the disk use arrangement, the processing arrangement protecting information read from the storage medium.

11. An appliance as in claim 10 wherein the processing arrangement includes a rights management arrangement that applies at least one rights management technique to the read information.

12. An appliance as in claim 10 wherein the appliance further includes at least one port compliant at least in part with the IEEE 1394-1995 high speed serial bus standard, and the processing arrangement couples the protected information to the port.

13. In an electronic appliance, a method including the following steps:

reading information from a digital versatile disk optical storage medium; and

protecting the information read from the optical storage medium.

14. A method as in claim 13 wherein the protecting step includes the step of applying at least one rights management technique to the read information.

15. A method as in claim 13 further including the step of sending the protected information to an IEEE 1394 port.

16. An electronic appliance including:

a disk use arrangement for using information stored,
or to be stored, on a digital versatile disk optical storage medium;
and

at least one protecting arrangement coupled to the
disk use arrangement and also coupled to receive at least one
analog signal, the protecting arrangement creating protected
digital information based at least in part on the analog signal.

17. In an electronic appliance, a method including the
following steps:

receiving at least one analog signal; and

creating protected digital content based at least in part
on the analog signal for storage on a digital versatile disk.

18. In an electronic appliance, a method including the
following steps:

reading at least one analog signal from a digital
versatile disk;

creating protected digital content based at least in part
on the analog signal; and

outputting the protected digital content.

19. An electronic appliance including:

a disk use arrangement for using information stored,
or to be stored, on a digital versatile disk optical storage medium;
and

at least one rights management arrangement coupled
to the disk use arrangement, the rights management arrangement
treating the storage medium and/or information obtained from the
storage medium differently depending on the geographical and/or
jurisdictional context of the appliance.

20. In an electronic appliance, a method including the
steps of:

reading information from at least one digital versatile
disk; and

performing at least one rights management operation based at least in part on the geographical and/or jurisdictional context of the appliance.

21. An electronic appliance including:

a disk use arrangement for using at least one secure container stored on a digital versatile disk optical storage medium; and

at least one rights management arrangement coupled to the disk use arrangement, the rights management arrangement processing the secure container.

22. In an electronic appliance, a method including the following steps:

reading at least one secure container from at least one digital versatile disk; and

performing at least one rights management operation on the secure container.

23. An electronic appliance including:

at least one rights management arrangement for generating and/or modifying at least one secure container for storage onto a digital versatile disk optical storage medium.

24. In an electronic appliance, a method including the step of performing at least one rights management operation on at least one secure container for storage onto a digital versatile disk optical storage medium.

25. A digital versatile disk use system and/or method characterized in that the system and/or method uses at least one secure container.

26. A digital versatile disk use system and/or method characterized in that the system and/or method uses at least one

secure container of the type disclosed in PCT Publication No. WO 96/27155.

27. An electronic appliance including:

a disk use arrangement for writing information onto and/or reading information from a digital versatile disk optical storage medium; and

a secure arrangement that securely manages information on the storage medium such that at least a first portion of the information may be used on at least a first class of appliance while at least a second portion of the information may be used on at least a second class of appliance

28. In an electronic appliance, a method including the following steps:

reading information from and/or writing information to at least one digital versatile disk optical storage medium;

using at least a first portion of the information on at least a first class of appliance; and

using at least a second portion of the information on at least a second class of appliance.

29. A system including first and second classes of electronic appliances each including a secure processing arrangement, the first appliance class secure arrangement securely managing and/or using at least a first portion of the information, the second appliance class secure arrangement securely managing and/or using at least a second portion of the information.

30. A system as in claim 29 wherein the first and second information portions are different, and the second appliance class secure arrangement does not use the first information portion.

31. A system as in claim 29 wherein the first appliance class does not use the second information portion.

32. In a system including first and second classes of electronic appliances each including a secure arrangement, a method comprising:

(a) securely managing and/or using at least a first portion of the information with the first appliance class secure arrangement, and

(b) securely managing and/or using at least a second portion of the information with the second appliance class secure arrangement.

33. A method as in claim 32 wherein the first and second information portions are different, and step (b) does not use the first information portion.

34. A method as in claim 32 wherein step (a) does not use the second information portion.

35. An electronic appliance including:

a disk use arrangement for writing information onto and/or reading information from a digital versatile disk optical storage medium; and

a secure arrangement that securely stores and/or transmits information associated with at least one of payment, auditing, controlling and/or otherwise managing content recorded on the storage medium.

36. In an electronic appliance, a method including the following steps:

reading information from and/or writing information to at least one digital versatile disk optical storage medium; and

securely storing and/or transmitting information associated with at least one of payment, auditing, controlling and/or otherwise managing content recorded on the storage medium.

37. An electronic appliance including:

a disk use arrangement for writing information onto and/or reading information from a digital versatile disk optical storage medium;

a cryptographic engine coupled to the disk use arrangement, the engine using at least one cryptographic key; and

a secure arrangement that securely updates and/or replaces at least one cryptographic key used by the cryptographic engine to at least in part modify the scope of information usable by the appliance.

38. A method of operating an electronic appliance including:

writing information onto and/or reading information from a digital versatile disk optical storage medium;

using at least one cryptographic key in conjunction with said information; and

securely updating and/or replacing at least one cryptographic key used by the cryptographic engine to at least in part modify the scope of information useable by the appliance.

39. A digital versatile disk appliance characterized in that at least one cryptographic key used by the appliance is securely updated and/or replaced to at least in part modify the scope of information that can be used by the appliance.

40. An appliance as in claim 39 further characterized in that the key updating and/or replacing is based on class of appliance.

41. An electronic appliance having a class associated therewith, characterized in that at least one cryptographic key set used by the appliance class is selected to help ensure security of information released from at least one digital versatile disk.

42. A digital camera for generating at least one image to be written onto a digital versatile disk optical storage medium, characterized in that the camera includes at least one information protecting arrangement that at least in part protects the image so that the information is persistently protected through subsequent processes such as editing, production, writing onto a digital versatile disk, and/or reading from a digital versatile disk.

43. A digital camera for generating image information that can be written onto a digital versatile disk optical storage medium, a method comprising:

capturing at least one image with a digital camera; and

protecting information provided by the digital camera so that the information is selectively persistently protected through subsequent processes such as distribution, editing and/or production, writing onto the digital versatile disk optical storage medium, and/or reading from the digital versatile disk optical storage medium.

44. In an electronic appliance including a disk use arrangement, a method comprising:

reading information from at least one digital versatile disk optical storage medium; and

persistently protecting at least some of the read information through at least one subsequent editing and/or production process.

45. In an electronic appliance, a method including the following steps:

reading information from and/or writing information to at least one digital versatile disk optical storage medium; and

securely managing information on the storage medium, including the step of using at least a first portion of the information on at least a first class of appliance, and using at least a second portion of the information on at least a second class of appliance.

46. A method of providing copy protection and/or use rights management of at least one digital property content and/or secure container to be stored and/or distributed on a digital versatile disk medium, comprising the step(s) of:

providing a set of use control(s) within a cryptographically encapsulated data structure having a predetermined format, the data structure format defining at least one secure software container for providing use rights information for digital property content to be stored on the digital versatile disk medium.

47. A method as in claim 46 further including the step of using at least one digital property content stored on an optical disk in accordance with the use controls, including the step of using a prescribed secure cryptographic key or set of cryptographic keys for using rights information.

48. A method as in claim 46 further including the step of decrypting control rules and/or other selected encrypted

information content encapsulated in the software container using at least one set of cryptographic keys.

49. A method as in claim 46 further including the step of applying decrypted control rules to regulate use in accordance with control information contained within said control rules, so as to facilitate management of a diverse set of use and distribution rights which may be specific to different users and/or optical disk appliances.

50. A method of providing rights management of digital property stored on digital versatile disk according to claim 46 wherein said secure container data structure comprises:

one or more content objects comprising digital property content; and

one or more control objects comprising a set of control rules defining copy protection, use and distribution rights to digital property content stored on the optical disk.

51. A method of providing rights management of digital property stored on a digital versatile disk according to claim 46, wherein a content object further comprises one or more reference pointers to digital property content stored elsewhere on the digital versatile disk.

52. A method of providing rights management of digital property stored on a digital versatile disk according to claim 46, wherein a control object further comprises one or more reference pointers to control information stored elsewhere on the digital versatile disk.

53. A method of providing rights management of digital property stored on digital versatile disk according to claim 46, wherein digital information stored on said optical disk includes one or more metadata blocks comprising further information used in conjunction with the control rules to use digital property content stored elsewhere on the optical disk.

54. A method of providing rights management of digital property stored on digital versatile disk according to claim 46, wherein a metablock may be either of a protected type or of an unprotected type.

55. An arrangement for implementing a rights management system for controlling copy protection, use and/or distribution rights to multi-media digital property content stored or otherwise contained on a digital versatile disk, comprising:

an encrypted data structure defining a secure information container stored on an optical disk medium, the encrypted data structure including and/or referencing at least one content object and at least one control object associated with the content object, said content object comprising digital property content and said control object comprising rules defining use rights to the digital property content.

56. An arrangement for implementing a rights management system for digital versatile disks according to claim 55, wherein a content object further comprises one or more reference pointers to digital property content stored elsewhere on the digital versatile disk.

57. An arrangement for implementing a rights management system for digital versatile disks according to claim 55, wherein a control object further comprises one or more reference pointers to control information stored elsewhere on the digital versatile disk.

58. An arrangement for implementing a rights management system for digital versatile disks according to claim 55, wherein an control object further comprises information for controlling various operations of an optical disk appliance or computer.

59. An arrangement for implementing a rights management system for digital versatile disks according to claim 55, wherein a control object further comprises information for controlling various operations of an optical disk appliance or computer.

60. An arrangement for implementing a rights management system for digital versatile disks according to claim 55, wherein a control object further comprises a rule specifying decoding and/or enforcement of CGMA encoded copy protection rules associated with the digital content property.

61. An arrangement for implementing a rights management system for digital versatile disks according to claim 55, wherein a control object further comprises a rule specifying at least one content scrambling system compatible encoding/decoding of digital property content.

62. An arrangement for implementing a rights management system for digital versatile disks according to claim 55, wherein said optical disk contains a block of stored information comprising encrypted keys used for decryption of said encrypted data structure.

63. An arrangement for implementing a rights management system for digital versatile disks according to claim 55, wherein said optical disk contains a block of stored information comprising hidden keys used for decryption of said encrypted keys.

64. An arrangement for implementing a rights management system for digital versatile disks according to claim 55, wherein a content object further comprises one or more reference pointers to digital property content stored on a separate storage medium.

65. A rights management system for providing copy protection, use and/or distribution rights management for multi-media digital property content stored or otherwise contained on a digital versatile disk for access by an optical disk player device that uses digital property content stored on said optical disk medium, wherein said appliance includes a microprocessor controller for decrypting and using control rules and other selected encrypted information content encapsulated in the secure container by using a prescribed cryptographic key and applying said decrypted control rules to regulate use in accordance with control information contained within said control rules, so as to facilitate management of a diverse set of use and/or distribution rights which may be specific to different users and/or optical disk appliances, the system including:

an optical disk medium having stored thereon an encrypted data structure defining a secure information container, the encrypted data structure comprising and/or referencing at least one content object and at least one control object, said content object comprising digital property content, said control object

comprising rules defining use rights associated with the digital property.

66. A method for providing copy protection, use and distribution rights management of multi-media digital property stored on and/or distributed via digital versatile disk, said optical disk medium having stored thereon an encrypted data structure defining a secure container for housing rights and/or copy protection information pertaining to digital property content stored on the optical disk, wherein an optical disk player appliance for using digital property content stored on an optical disk must utilize a prescribed secure cryptographic key or set of keys to use the secure container, said data structure comprising one or more content objects comprising digital property content and one or more control objects comprising a set of rules defining use rights to digital property, comprising the steps of:

(a) decrypting control rules and other selected encrypted information content encapsulated in the secure container using one or more cryptographic keys; and

(b) applying decrypted control rules to regulate use and/or distribution of digital property content stored on the optical disk in accordance with control information contained within the control rules, so as to provide customized use and/or distribution rights that are specific to different optical disk user platforms and/or optical disk appliances.

67. A rights management system for providing copy protection, use and/or distribution rights management of digital property stored or otherwise contained on a digital versatile disk, comprising:

a secure container means provided on an optical disk medium for cryptographically encapsulating digital property content stored on the optical disk, said container means comprising a content object means for containing digital property content and a control object means for containing control rules for regulating use and/or distribution of said digital property content stored on the optical disk.

68. The rights management system of claim 67 wherein an optical disk player appliance for using information stored on an optical disk comprises a secure node means for using said secure container means provided on an optical disk and implementing said control rules to control operation of said player appliance to regulate use of said digital property content.

69. In a system including plural electronic appliances at least temporarily connected to one another, a rights authority broker that determines what appliances are connected and specifies at least one rights management context depending on said determination.

70. An electronic appliance at least temporarily connected to a rights authority broker, the electronic appliance receiving at least one rights context from the rights authority broker when the device is connected to the rights authority broker.

71. A first electronic appliance at least temporarily connected to a second electronic appliance, the first

electronic appliance selecting between at least first and second rights management contexts depending at least in part on whether the first appliance is connected to the second electronic appliance.

72. In a system including first and second electronic appliances that may be selectively coupled to communicate with one another, an arrangement for defining at least one different rights management control based at least in part on whether the first and second electronic appliances are connected.

73. A method of defining at least one rights management context comprising:

(a) determining whether a first electronic appliance is present; and

(b) defining at least one rights management control set based at least in part on the determining step (a).

74. A method of defining at least one rights management context including:

(a) coupling an optical disk storing information to an electronic appliance that can be selectively connected to a rights management broker;

(b) determining whether the electronic appliance is currently coupled to a rights management broker; and

(c) conditioning at least one aspect of use of at least some of the information stored on the optical disk based on whether the electronic appliance is coupled to the rights management broker.

75. A method as in claim 74 wherein step (c) includes the step of obtaining at least one rights management context from the rights management broker.

76. A method as in claim 74 wherein step (c) includes the step of obtaining at least one combined control set from the rights management broker.

77. A method of defining at least one rights management context including:

(a) coupling an optical disk storing information to an electronic appliance;

(b) using at least some of the information stored on the optical disk based on a first rights management context;

(c) coupling the electronic appliance to a rights management broker; and

(d) concurrently with step (c), using at least some of the information stored on the optical disk based on a second rights management context different from the first rights management context

78. An electronic appliance include a secure node and an optical disk reader, the electronic appliance applying different rights management contexts to protected information stored on an optical disk coupled to the optical disk reader depending at least in part on whether the electronic appliance is coupled to at least one additional secure node.

79. An electronic appliance including:

an optical disk reading and/or writing arrangement;

a secure node coupled to the optical disk reading and/or writing arrangement, the secure node performing at least one rights management related function with respect to at least some information read by the optical disk reading and/or writing arrangement; and

at least one serial bus port coupled to the secure node, the serial bus port for providing any or all of the functions, structures, protocols and/or methods of IEEE 1394-1995.

80. A digital versatile disk appliance including:

means for watermarking content; and

serial bus means for communicating the watermarked content,

wherein the serial bus means complies with IEEE 1394-1995.

81. An optical disk reading and/or writing device including:

at least one secure node capable of watermarking content

and/or processing watermarked content; and

an IEEE 1394-1995 serial bus port.

82. An optical disk using device comprising:

a secure processing unit; and

an IEEE 1394-1995 serial bus port.

83. A device as in claim 82 wherein the secure processing
unit includes a channel manager.

84. A device as in claim 82 wherein the secure processing
unit executes a rights operating system in whole or in part.

85. A device as in claim 82 wherein the secure processing
unit includes a tamper-resistant barrier.

86. A device as in claim 82 wherein the secure processing
unit includes an encryption/decryption engine.

87. A rights cooperation method comprising:

(a) connecting an appliance to at least one further appliance;

(b) determining whether the first and/or further appliance and/or user(s) of said first and/or further appliance have appropriate rights and/or resources for performing an action; and

(c) selectively performing the action based at least in part on the determination.

88. A rights cooperation method comprising:

(a) connecting an appliance to at least one further appliance;

(b) determining whether the first and/or further appliance and/or user(s) of said first and/or further appliance have appropriate security for performing an action; and

(c) cooperating between the first and further appliance to selectively perform the action.

89. A cooperative rights management arrangement comprising:

a communications arrangement that allows at least first and second appliances to communicate; and

an arrangement that processes at least one event based at least in part on assessing and/or pooling rights and/or resources between the first and second appliances.

90. An optical disk using system and/or method including at least some of the elements shown in Figure 1A.

91. An optical disk using system and/or method including at least some of the elements shown in Figure 1B.

92. An optical disk using system and/or method including at least some of the elements shown in Figure 1C.

93. An optical disk using system and/or method including at least some of the elements shown in Figure 2A.

94. An optical disk using system and/or method including at least some of the elements shown in Figure 2B.

95. An optical disk using system and/or method including at least some of the elements shown in Figure 3.

96. An optical disk using system and/or method using at least some of the elements shown in Figure 3A.

97. An optical disk using system and/or method using at least some of the control set elements shown in Figure 3B.

98. An optical disk using system and/or method using at least some of the elements shown in Figure 4A.

99. An optical disk using system and/or method using at least some of the elements shown in Figure 4B.

100. An optical disk using system and/or method using at least some of the elements shown in Figure 5.

101. An optical disk using system and/or method using at least some of the elements shown in Figure 6.

102. An optical disk using system and/or method using at least some of the elements shown in Figure 7.

103. An optical disk using system and/or method using at least some of the elements shown in Figure 8.

104. An optical disk using system and/or method using at least some of the elements shown in Figure 9.

105. An optical disk using system and/or method using at least some of the elements shown in Figure 10.

106. An optical disk using system and/or method using at least some of the elements shown in Figure 11.

107. An optical disk using system and/or method including at least some of the elements shown in Figure 12.

108. An optical disk using system and/or method including at least some of the elements shown in Figure 13.

109. An optical disk using system and/or method including at least some of the elements shown in Figure 14.

110. A system and/or method including some or all of the elements shown in Figures 15A-15C.

111. A system and/or method as in any one of the preceding claims, further including, in combination, any element described in any one of the following prior patent specifications:

PCT Publication No. WO 96/27155;

European Patent No. EP 329681;

PCT Application No. PCT/US96/14262;

U.S. Patent Application Serial No. 08/689,606; and/or

U.S. Patent Application Serial No. 08/689,754.

112. A system or process as in any of the preceding claims wherein the phrase "high capacity optical disk" is substituted for "digital versatile disk."

113. A method of clearing or otherwise processing information resulting at least in part from one or more digital versatile disk appliances and/or methods as defined in any of the preceding claims.

114. A system and/or method for defining rules for use in one or more digital versatile disk appliances and/or methods as defined in any of the preceding claims.

115. A system and/or method for defining rules and associated content for use in one or more digital versatile disk appliances and/or methods as defined in any of the preceding claims.

116. A system and/or method for producing an optical disk for use with one or more digital versatile disk appliances and/or methods as defined in any of the preceding claims.

117. A system and/or method for clearing audit information from one or more appliances and/or methods as defined in any of the preceding claims.

118. In an network including at least one electronic appliance that reads information from and/or writes information to at least one digital versatile disk optical storage medium, and securely communicates information associated with at least one of

payment, auditing, usage, access, controlling and/or otherwise managing content recorded on the storage medium, a method of processing said communicated information including the step of generating at least one payment request and/or order based at least in part on the information.

119. A method of defining at least one control set for storage on a high capacity optical disk that can storage images, audio, text and/or other information, the high capacity optical disk for use by any of plural different electronic appliance types, the method including the step of specifying at least one control that provides different conditions and/or consequences depending upon at least one of the following:

electronic appliance class;

electronic appliance security;

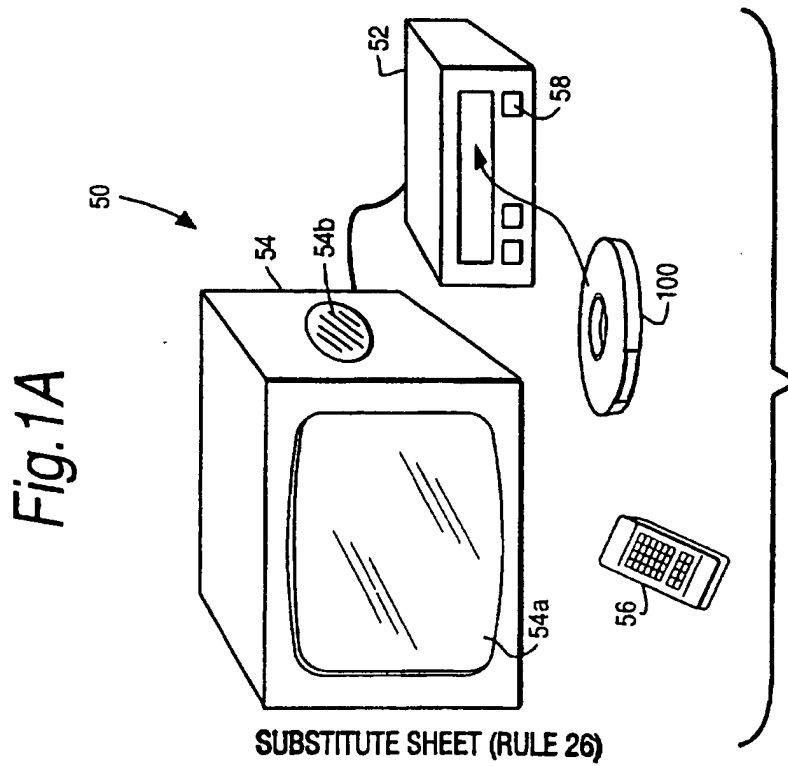
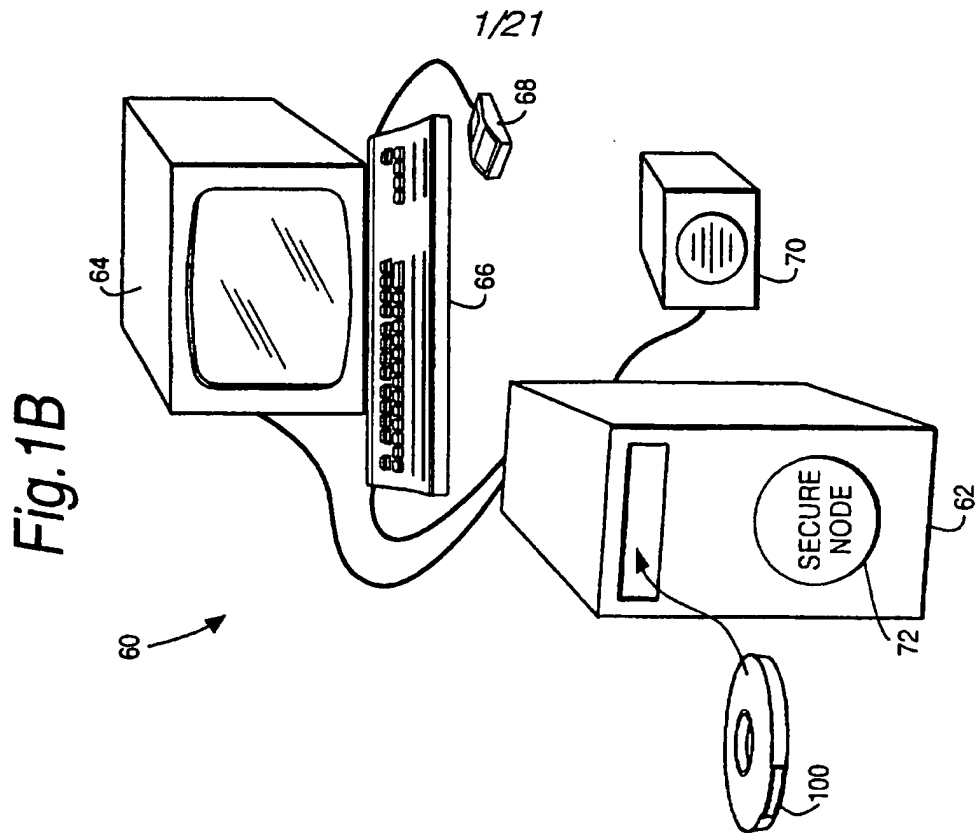
electronic appliance user class;

electronic appliance connectivity;

electronic appliance resources;

electronic appliance access to resources; and

rights management cooperation between plural electronic
appliances.



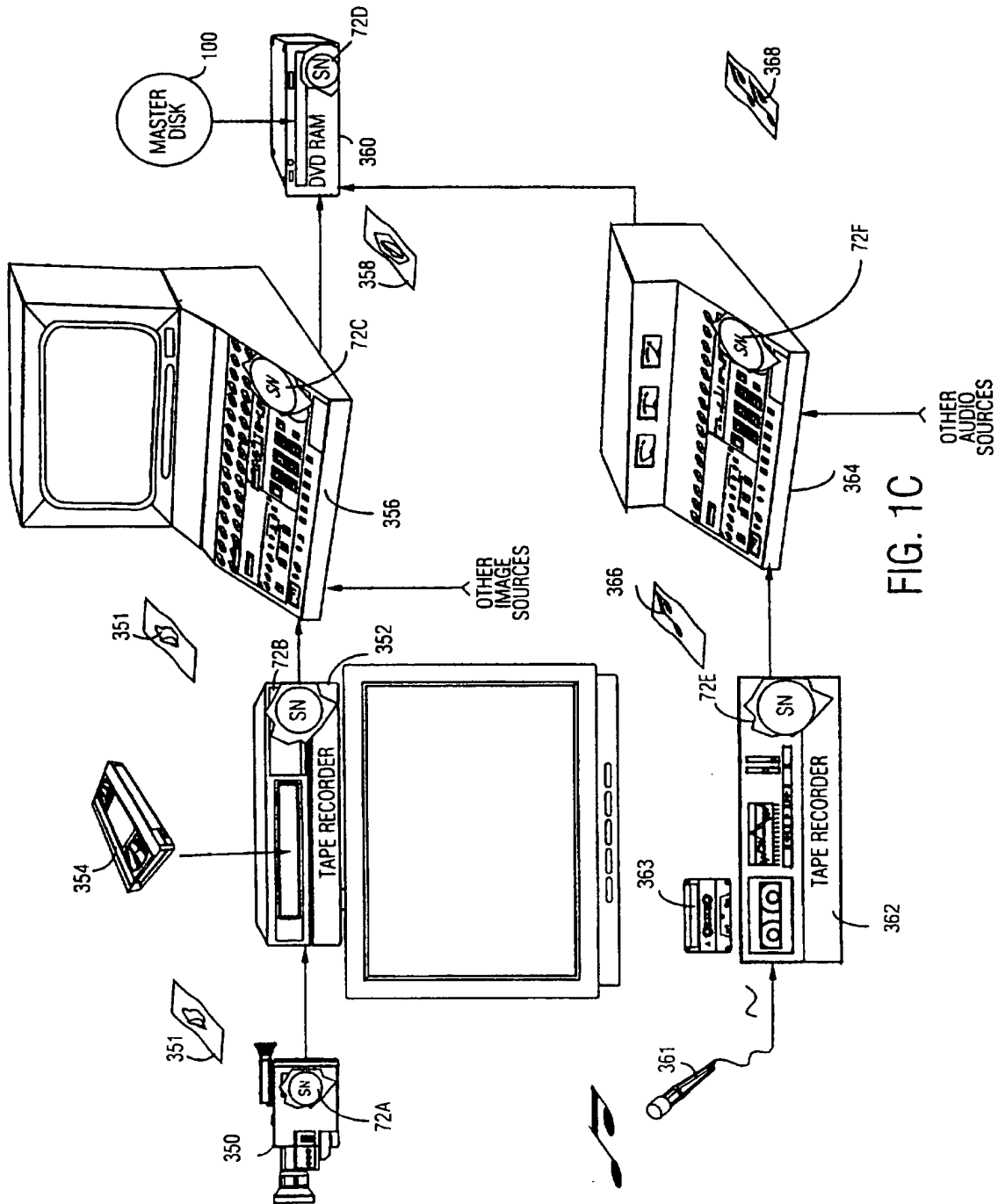


FIG. 1C

3/21

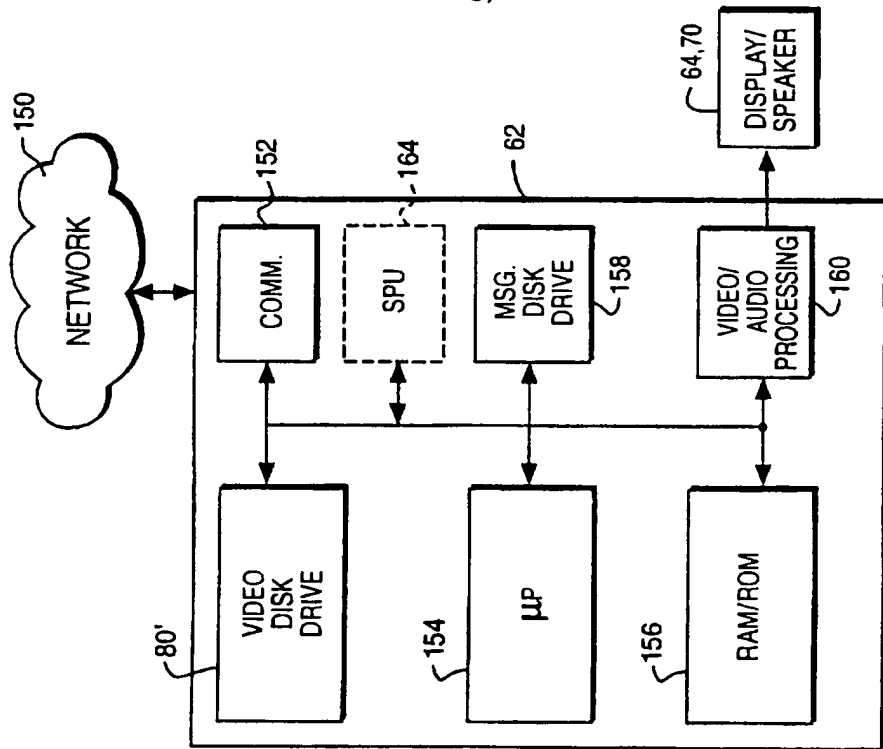


Fig.2B

EXAMPLE SECURE NODE ARCHITECTURE

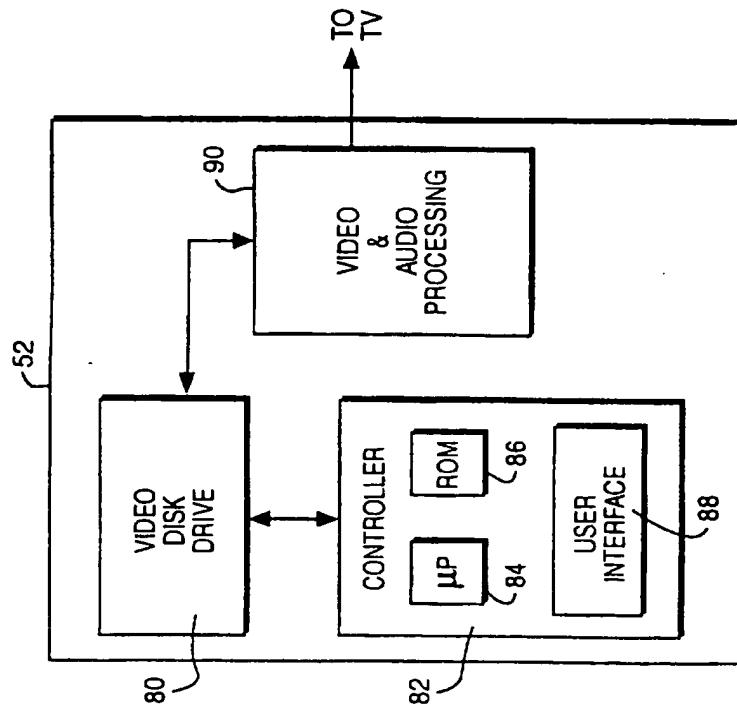


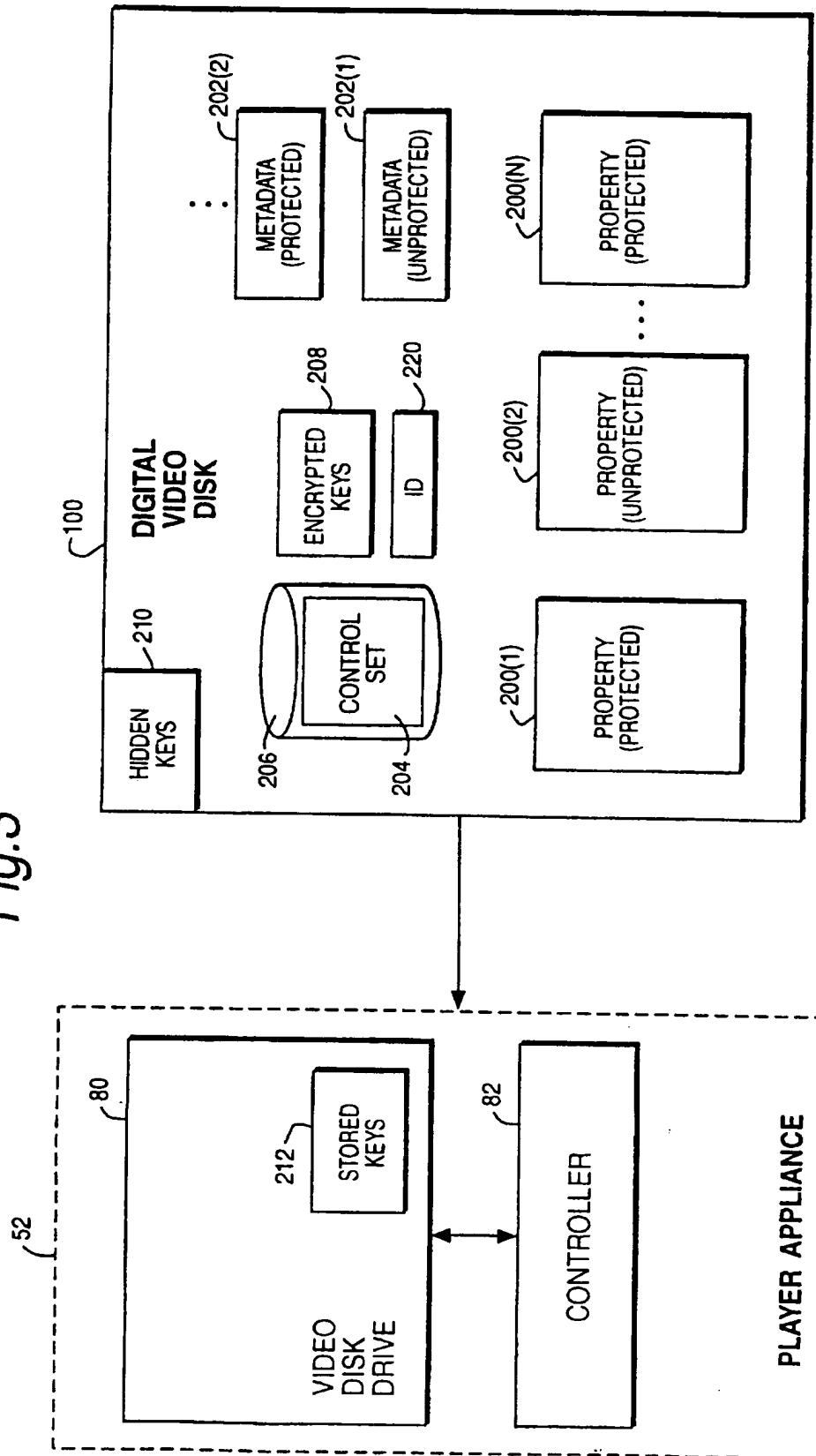
Fig.2A

EXAMPLE PLAYER ARCHITECTURE

SUBSTITUTE SHEET (RULE 26)

4/21

Fig.3



SUBSTITUTE SHEET (RULE 26)

5/21

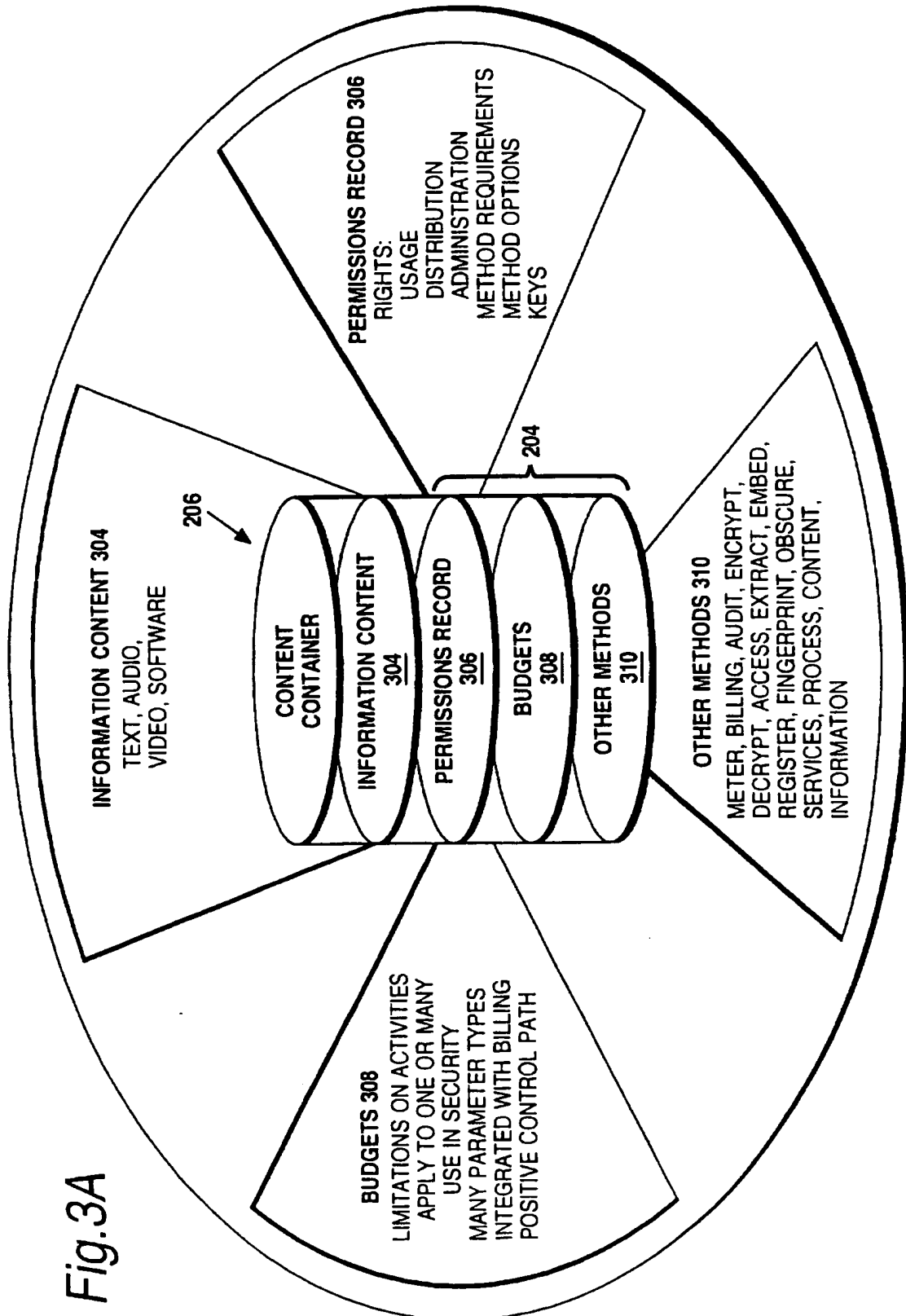


Fig.3A

SUBSTITUTE SHEET (RULE 26)

6/21

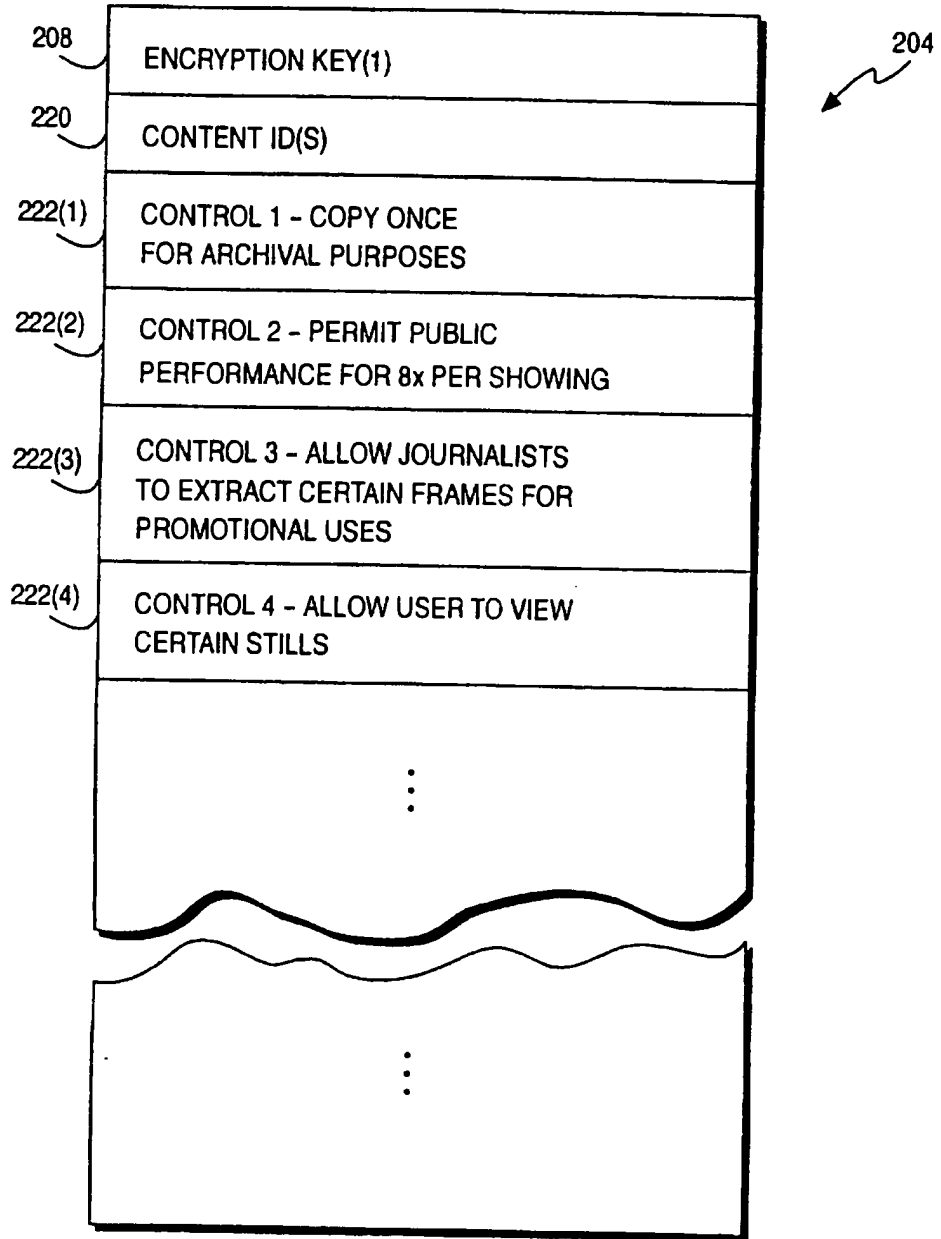
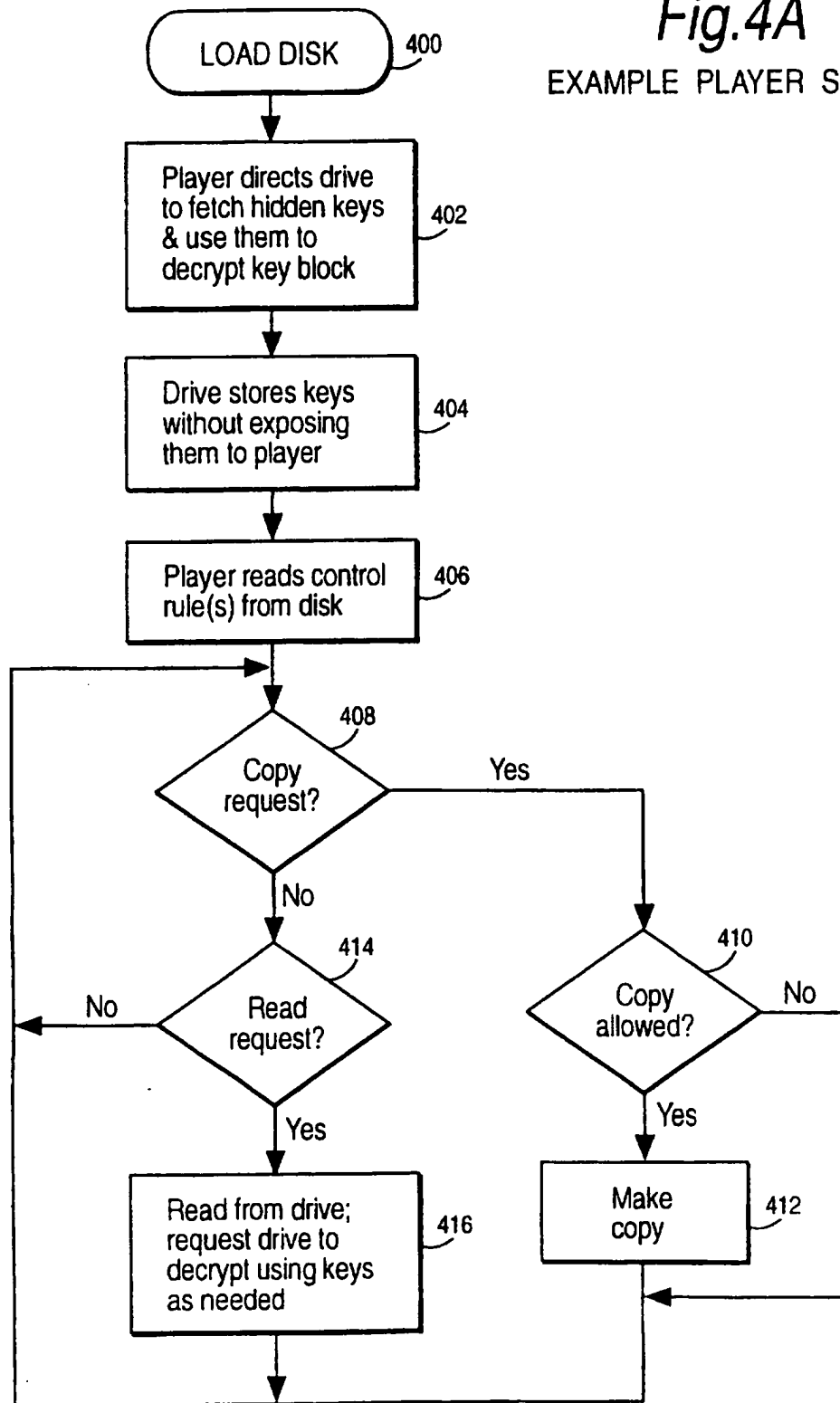


Fig.3B

EXAMPLE CONTROL SET
SUBSTITUTE SHEET (RULE 26)

7/21 -

Fig.4A
EXAMPLE PLAYER STEPS

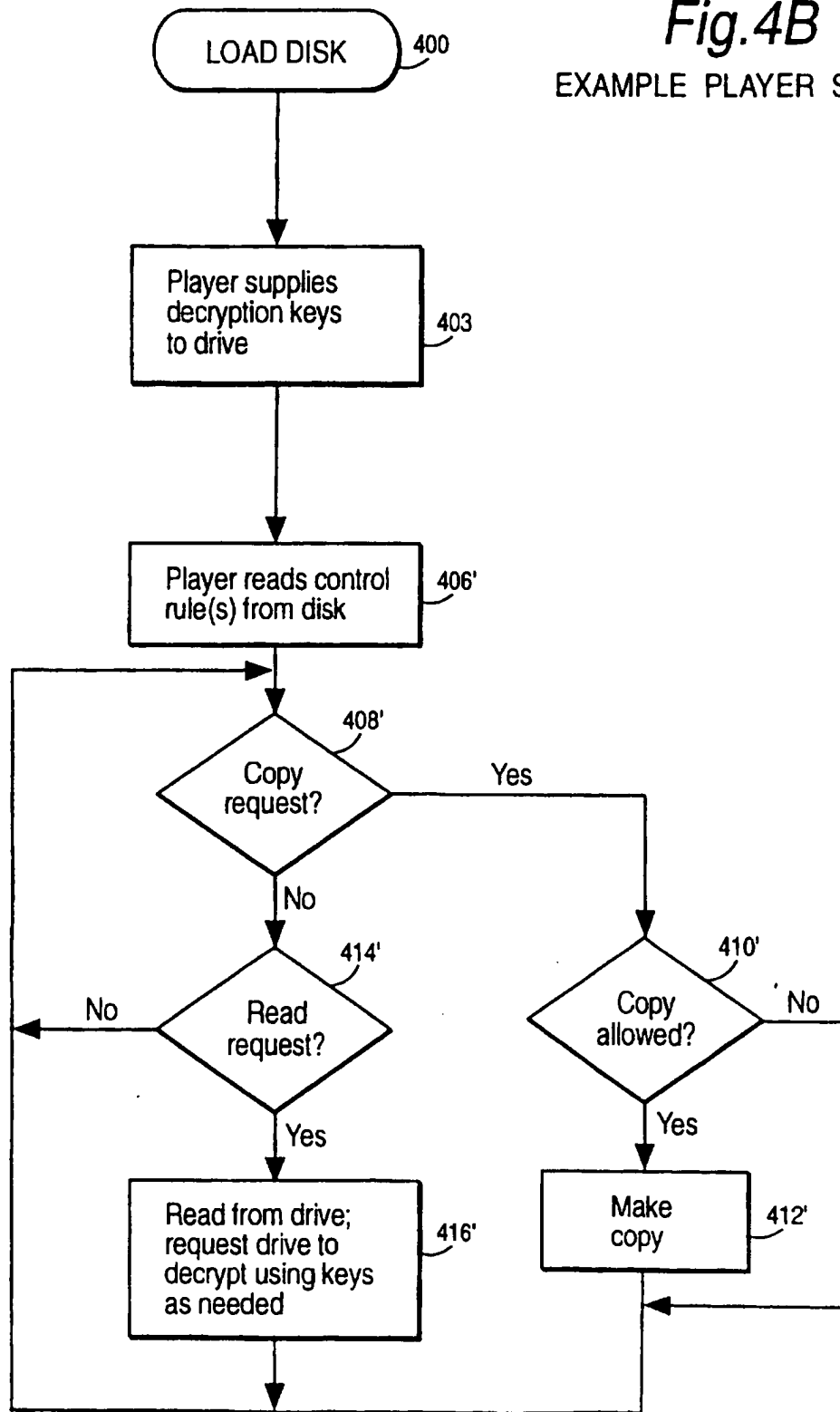


SUBSTITUTE SHEET (RULE 26)

8/21 -

Fig.4B

EXAMPLE PLAYER STEPS



SUBSTITUTE SHEET (RULE 26)

9/21

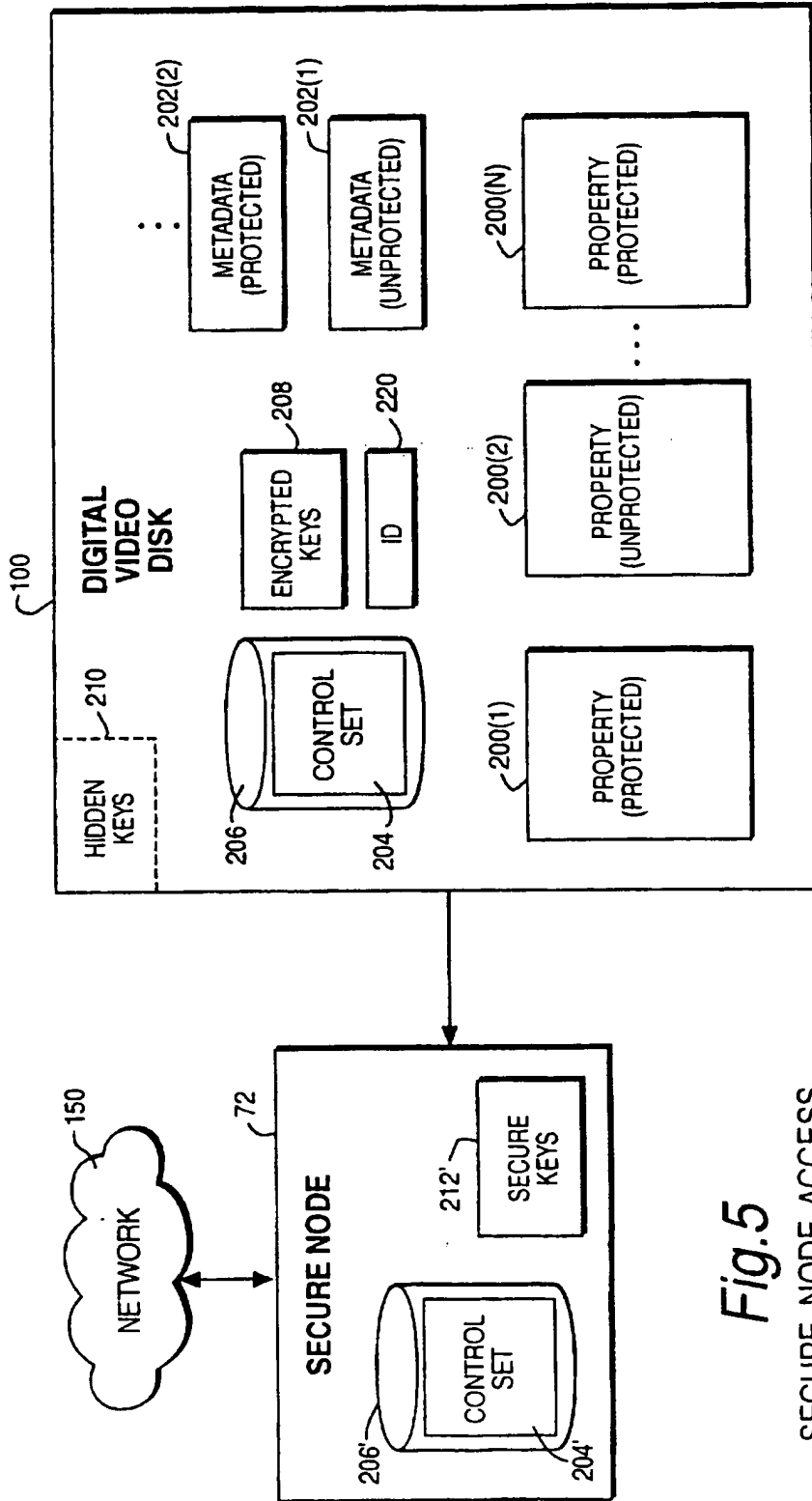


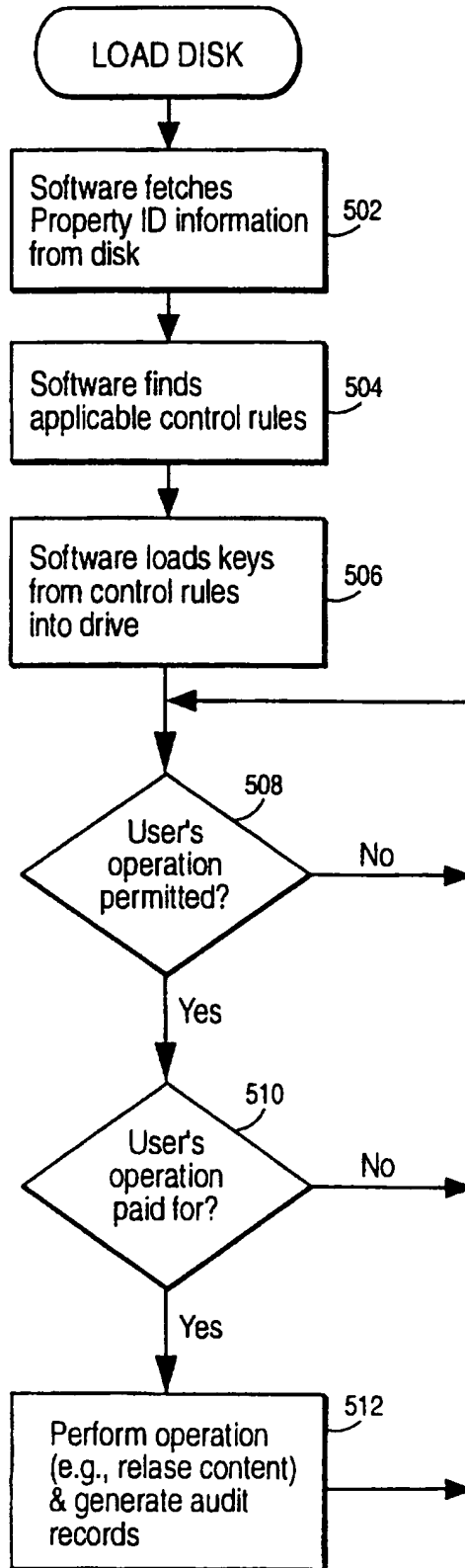
Fig.5

SECURE NODE ACCESS

SUBSTITUTE SHEET (RULE 26)

10/21

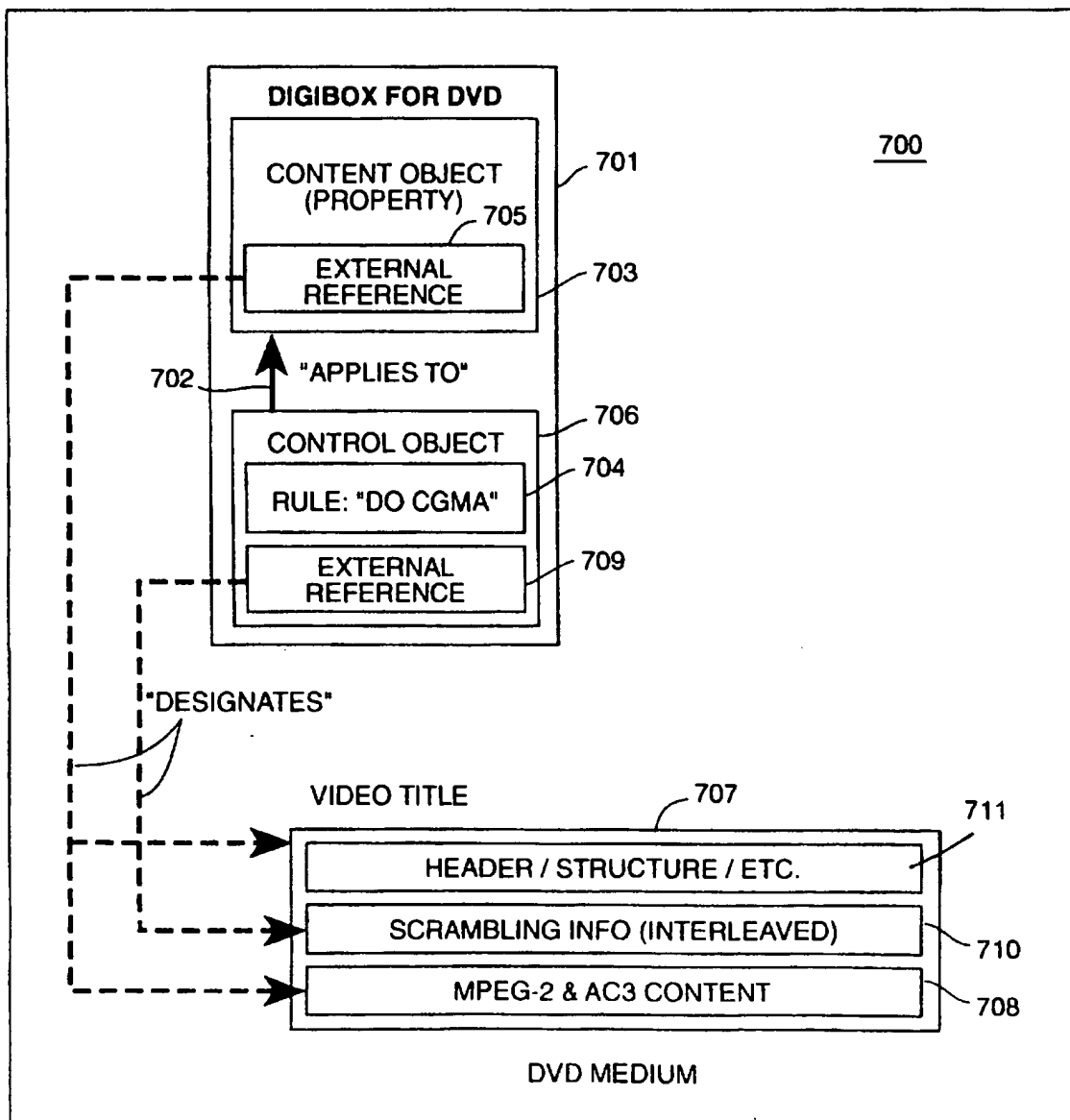
Fig.6



SUBSTITUTE SHEET (RULE 26)

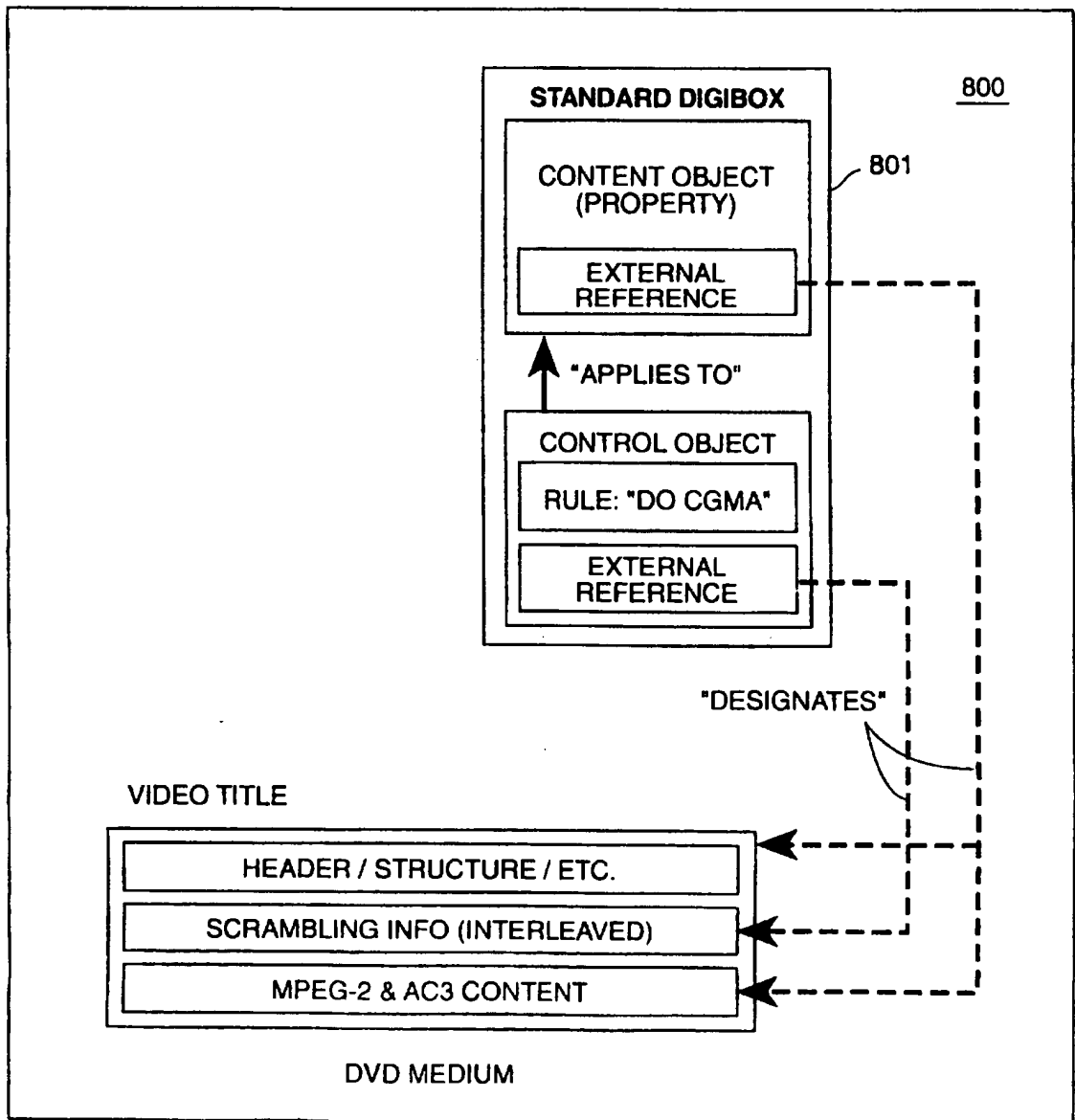
11/21

FIG. 7



12/21

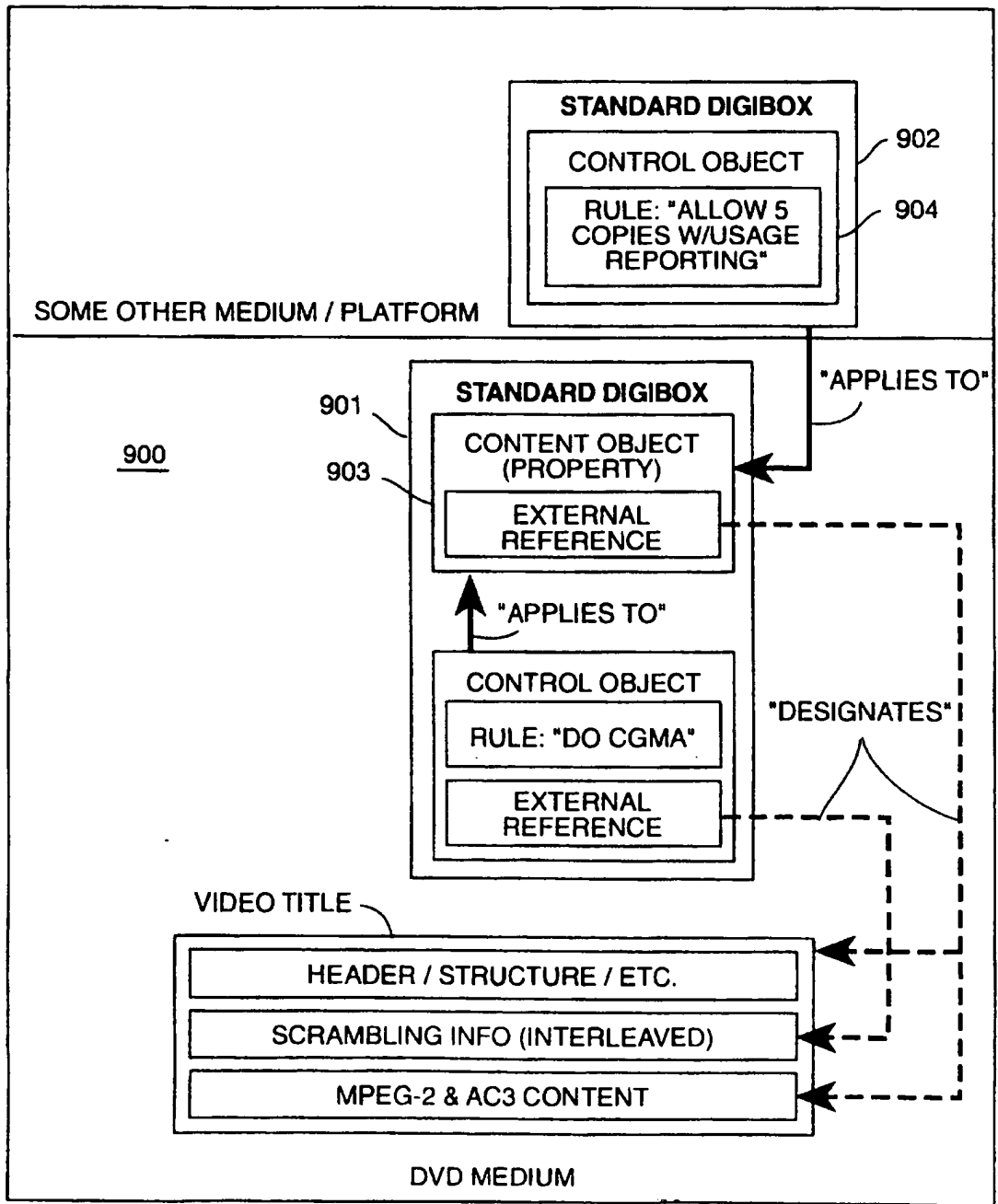
FIG. 8



SUBSTITUTE SHEET (RULE 26)

13/21

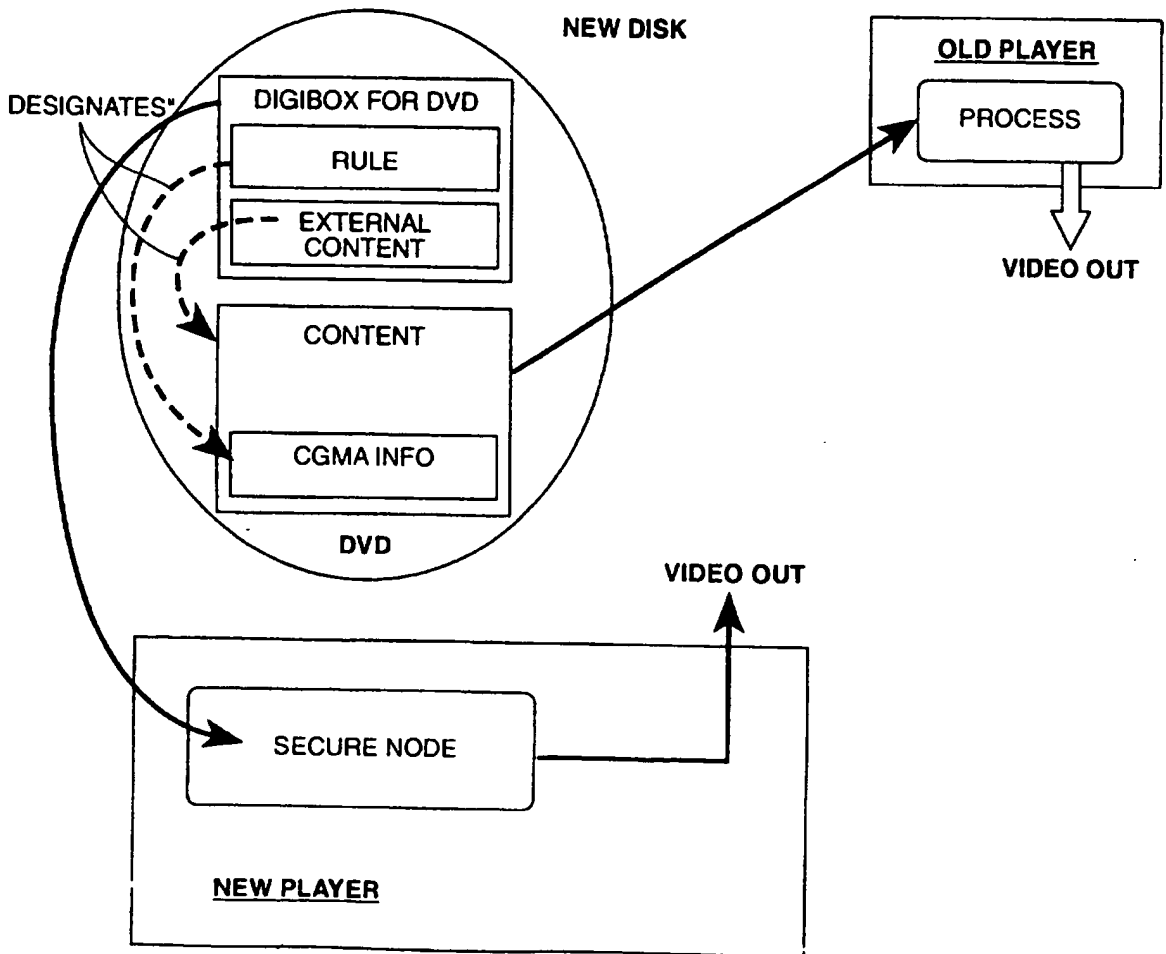
FIG. 9



SUBSTITUTE SHEET (RULE 26)

14/21

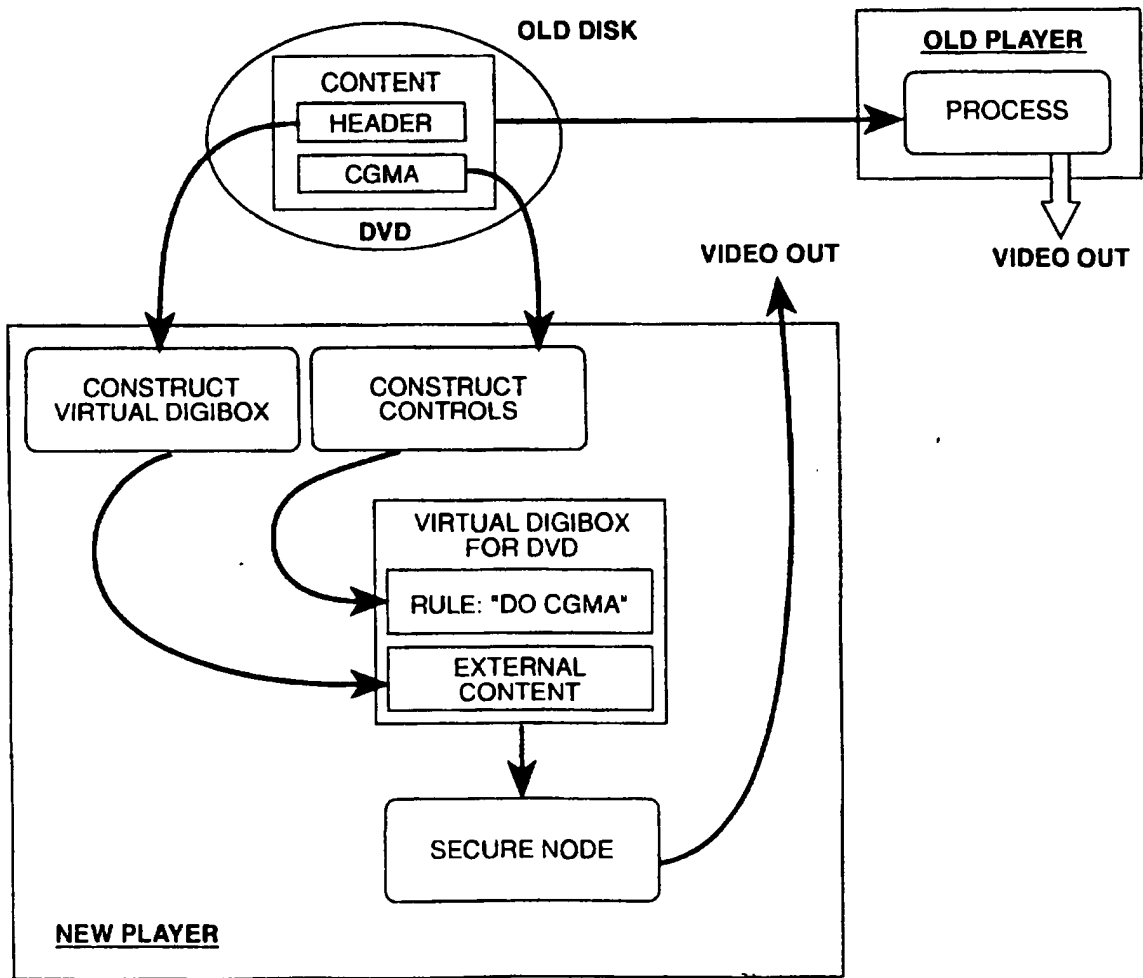
FIG. 10



SUBSTITUTE SHEET (RULE 26)

15/21

FIG. 11



SUBSTITUTE SHEET (RULE 26)

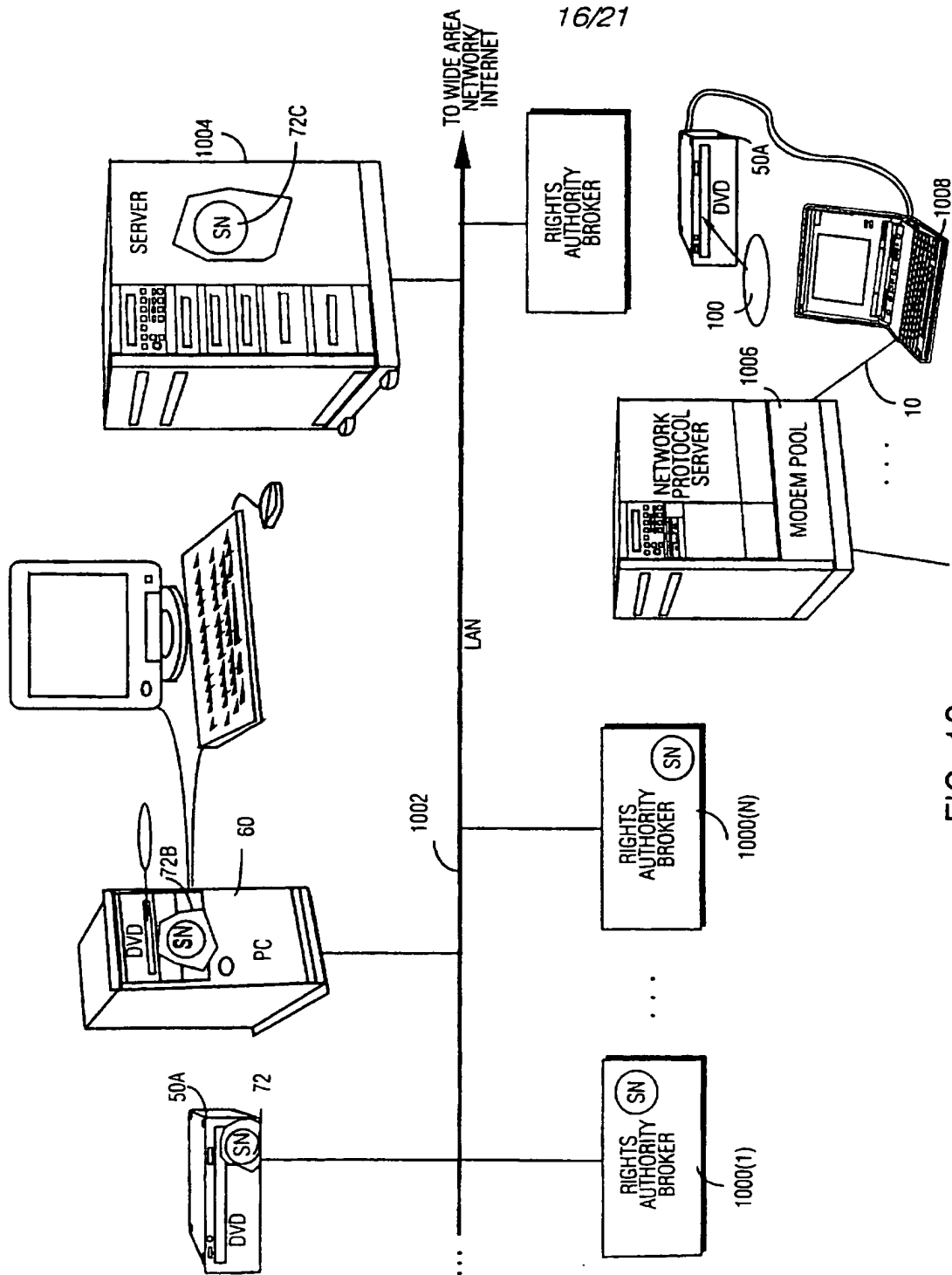


FIG. 12

SUBSTITUTE SHEET (RULE 26)

17/21

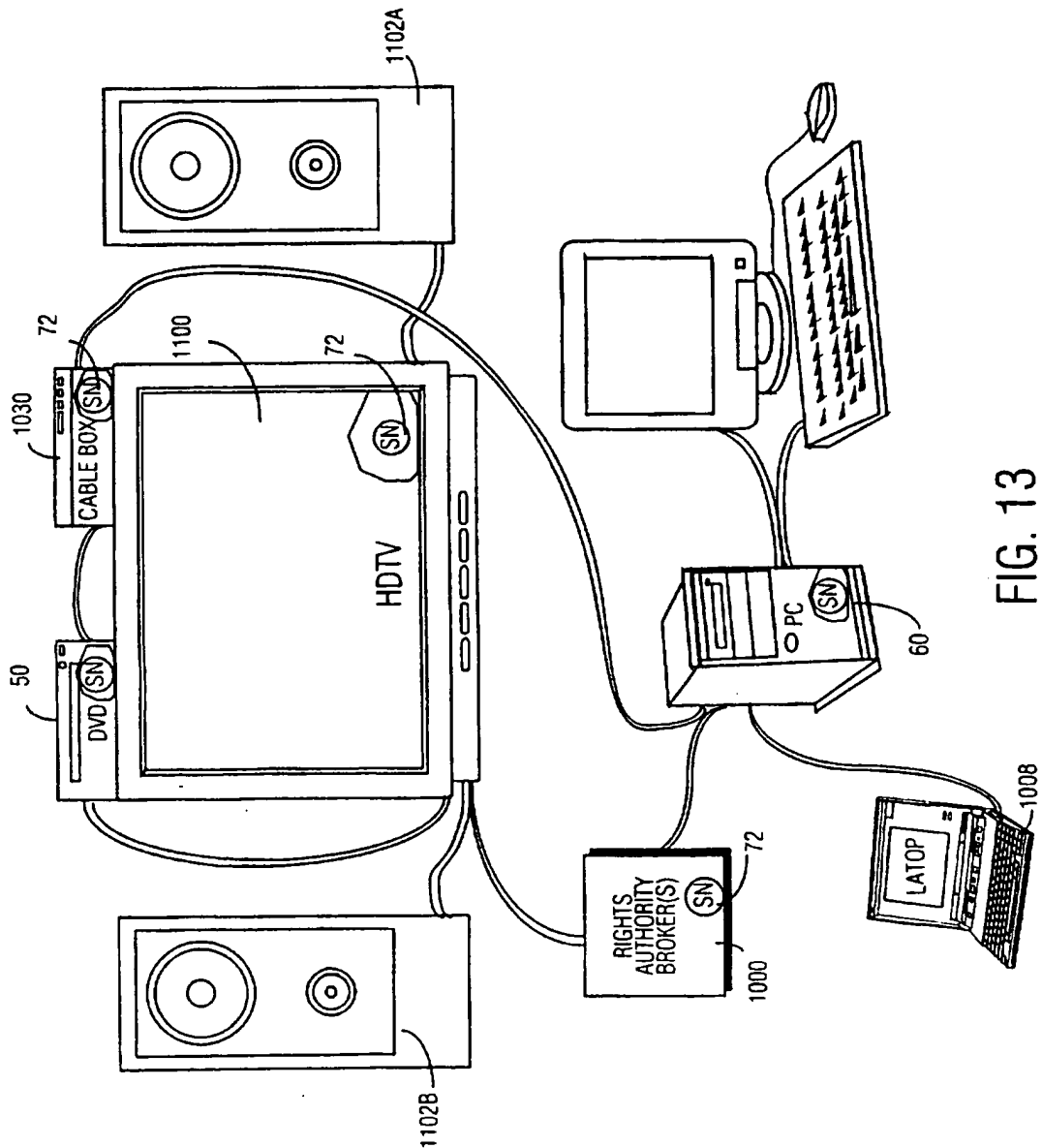


FIG. 13

SUBSTITUTE SHEET (RULE 26)

19/21

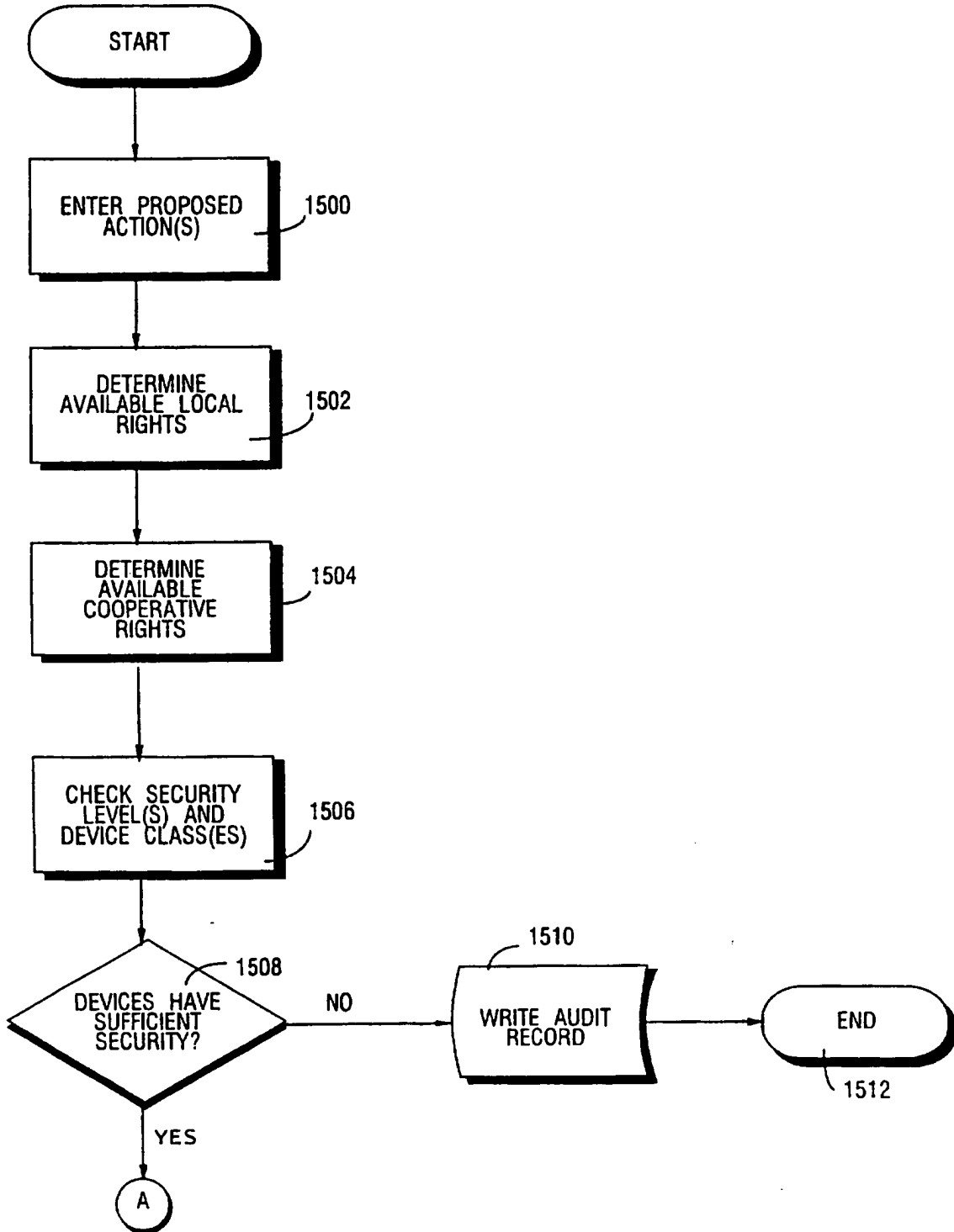


FIG.15A

SUBSTITUTE SHEET (RULE 26)

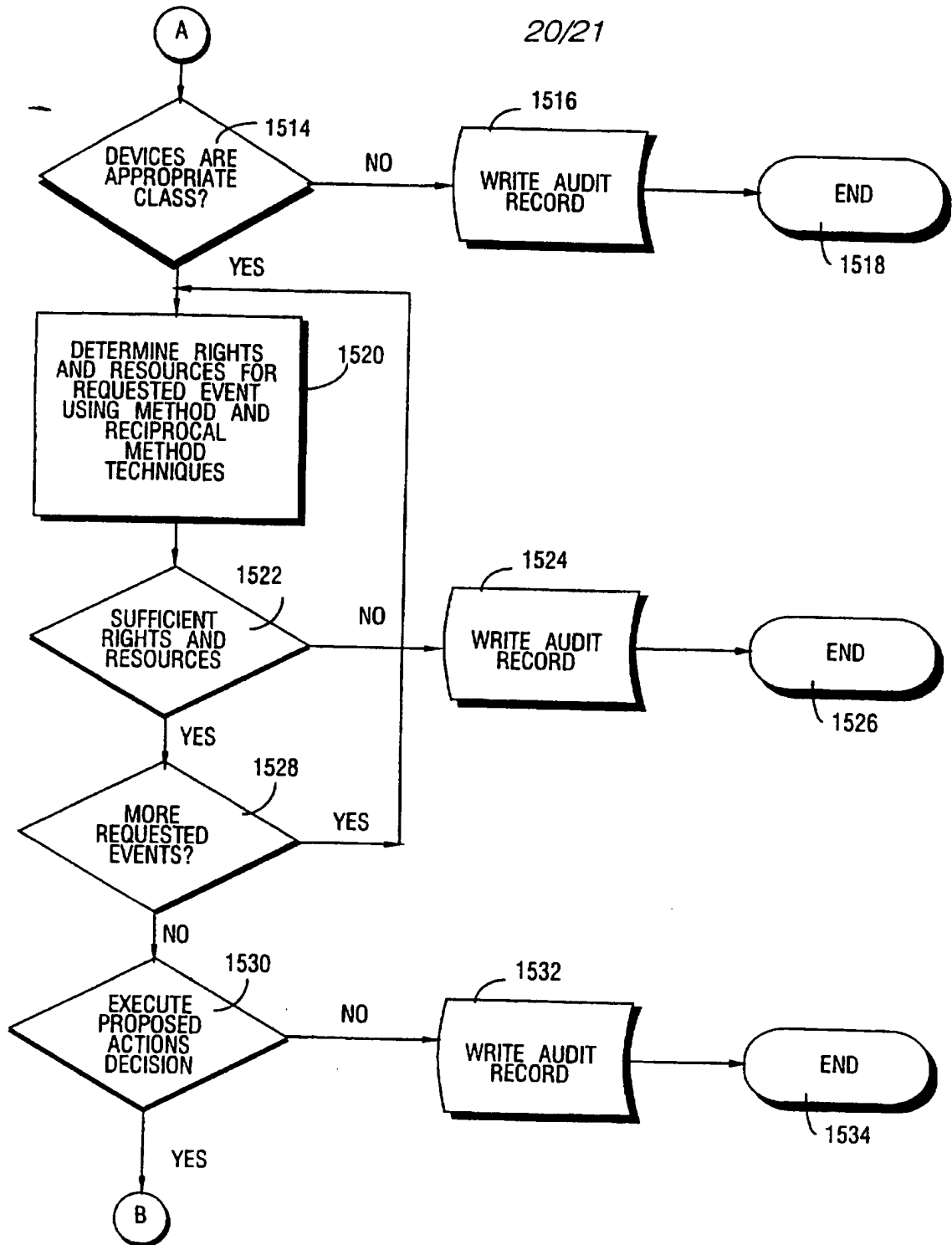
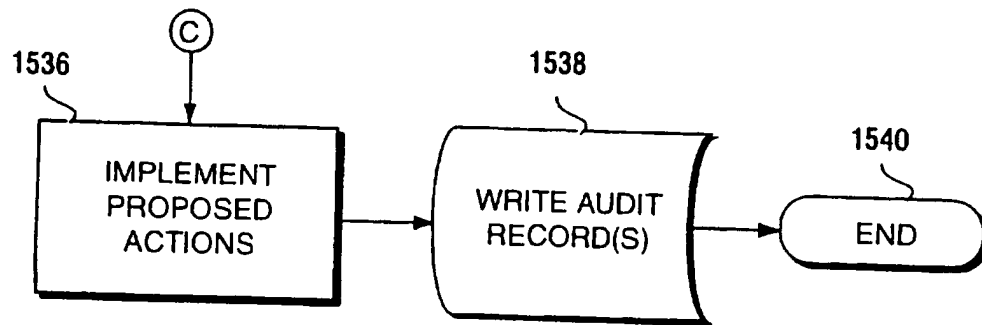


FIG. 15B
SUBSTITUTE SHEET (RULE 26)

FIG. 15C



SUBSTITUTE SHEET (RULE 26)



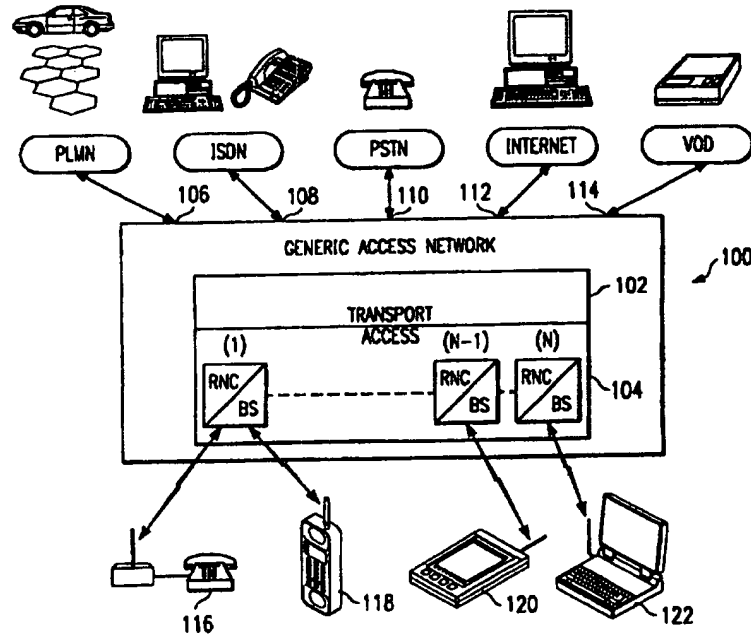
INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

<p>(51) International Patent Classification⁶ : H04L 9/08, H04Q 7/32</p>	<p>A1</p>	<p>(11) International Publication Number: WO 98/10561 (43) International Publication Date: 12 March 1998 (12.03.98)</p>
<p>(21) International Application Number: PCT/SE97/01407 (22) International Filing Date: 26 August 1997 (26.08.97) (30) Priority Data: 08/708,796 9 September 1996 (09.09.96) US (71) Applicant: TELEFONAKTIEBOLAGET LM ERICSSON (publ) [SE/SE]; S-126 25 Stockholm (SE). (72) Inventor: RUNE, Johan; Motionsvägen 5, S-181 30 Lidingö (SE). (74) Agents: BANDELIN, Hans et al.; Telefonaktiebolaget LM Ericsson, Patent and Trademark Dept., S-126 25 Stockholm (SE).</p>		<p>(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, HU, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, TJ, TM, TR, TT, UA, UG, UZ, VN, YU, ARIPO patent (GH, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).</p> <p>Published <i>With international search report.</i></p>

(54) Title: METHOD AND APPARATUS FOR ENCRYPTING RADIO TRAFFIC IN A TELECOMMUNICATIONS NETWORK

(57) Abstract

A generic communications network (100) provides an encrypted communications interface between service networks (130, 132, 134) and their subscribers. When communications are initiated between a subscribing communications terminal (118) and the generic network (100), the terminal (118) compares a stored network identifier associated with a stored public key, with a unique identifier broadcast by the generic network (100). If a match is found, the terminal (118) generates a random secret key, encrypts the secret key with the stored public key, and transmits the encrypted secret key. The generic communications network (100) decrypts the secret key using a private key associated with the public key. The secret key is used thereafter by the terminal (118) and the generic network (100) to encrypt and decrypt the ensuing radio traffic. Consequently, the network (100) can maintain secure communications with the terminal (118) without ever knowing the terminal's identity.



FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakistan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LJ	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

-1-

**METHOD AND APPARATUS FOR ENCRYPTING RADIO TRAFFIC
IN A TELECOMMUNICATIONS NETWORK**

BACKGROUND OF THE INVENTION

5 Technical Field of the Invention

The present invention relates generally to the field of wireless radio communications and, in particular, to a method and apparatus for encrypting radio traffic between terminals and a mobile communications network.

Description of Related Art

10 The need for increased mobility and versatility in telecommunications networks requires the networks to cover larger geographical areas and provide a broader range of telecommunications services to subscribers. These telecommunications services include teleservices and bearer services. A teleservice provides the necessary hardware and software for a subscriber to communicate with
15 another subscriber (e.g., terminal, etc.). A bearer service provides the capacity required to transmit appropriate signals between two access points (e.g., ports) that provide an interface with a network. Telecommunications services can be provided to subscribers by a number of service networks, such as, for example, public land mobile telecommunications networks (PLMNs), public switched telephone networks
20 (PSTNs), integrated services digital networks (ISDNs), the so-called "Internet" access networks, video on demand (VOD) networks, and other proprietary service networks.

 In response to the need for increased mobility and versatility, a new mobile radio telecommunications network is being developed, which has a generic interface
25 through which a service network subscriber can be connected with that service network regardless of the subscriber's geographic location. This generic mobile radio network is referred to as the "Generic Access Network" (GAN). In order to more readily understand the present invention, which deals primarily with encrypting communications traffic between terminals and a GAN, a brief description
30 of such a GAN is provided below with respect to FIGURE 1.

-2-

FIGURE 1 is a perspective view of an exemplary GAN connected to a plurality of service networks and service network subscribers. The GAN (10) illustrated by FIGURE 1 includes an access network interconnected with a transport network. The access network includes a plurality of base stations (e.g., BS1 and BS2). Each base station includes a radio transmitter and receiver that provides communications coverage for a respective geographical area (e.g., a so-called cell, C1 and C2). The base stations are connected to a radio network controller (RNC) 12. Although not shown explicitly, certain of the base stations can be connected to RNC 12 (e.g., BS1 and BS2), and certain other of the base stations can be connected to one or more other RNCs. A plurality of the RNCs can be interconnected to provide a communications path therebetween. The RNCs distribute signals to and from the connected base stations.

A plurality of service networks (e.g., VOD network, PLMN, PSTN, Internet) are connected through respective access input ports (14, 16, 18, 20, 22, 24 and 26) to the access network of GAN 10. Each service network uses its own standard signaling protocol to communicate between its internal signaling nodes. For example, the Global System for Mobile communications (GSM), which is a digital cellular PLMN that has been fielded throughout Europe, uses the Multiple Application Part (MAP) signaling protocol. As illustrated by FIGURE 1, the RNCs in the access network are connected through at least one of the access input ports to a service network. As shown, RNC 12 is connected through access ports 20 and 24, respectively, to the PLMN and PSTN service networks.

Mobile terminals 28 and 30 are located within the radio coverage area of GAN 10, and can establish a connection with each of the base stations (e.g., BS2) in the access network. These mobile terminals can be, for example, a cellular phone, mobile radiotelephone, personal computer (notebook, laptop, etc.) possibly connected to a digital cellular phone, or mobile television receiver (for VOD). Signal transport between a mobile terminal and a selected service network takes place over specified signal carriers. For example, signals are transported between the cellular phone (28) and the PLMN service network over signal carriers SC1 and SC2.

-3-

The mobile terminals (e.g., 28 and 30) include an access section and service network section. The access section of a mobile terminal is a logical part of the access network and handles the signaling required to establish the signal carrier (e.g., SC2 and SC4) between the mobile terminals and RNC 12. The service network section of a mobile terminal is a logical part of the service network to which that terminal's user subscribes. The service network section of a mobile terminal receives and transmits signals, in accordance with the specified standards of its related service network, via the established signal carriers SC1 and SC2 (or SC4). The radio interface portion of the signal carrier SC2 or SC4 (between the mobile terminal and base station) can be time division multiple access (TDMA), code division multiple access (CDMA), or any other type of multiple access interface.

The service network subscribers can access their respective service network through the GAN. The GAN provides a signal carrier interface that allows a message to be transported transparently over a signal carrier (e.g., SC1 and SC2) between the service network part of a mobile terminal and its service network. The GAN accomplishes this function by matching the characteristics of the signaling connections and traffic connections of all of the service networks that connect to it. Consequently, the GAN can extend the coverage of existing service networks and also increase the subscribers' degree of mobility.

A unique characteristic of a GAN is that it has no subscribers of its own. The mobile users of the GAN are permanent subscribers to their own service networks, but they are only temporary users of the GAN. Consequently, a GAN does not know (or need to know) the identity of these users. However, a problem arises in attempting to encrypt radio traffic between the mobile terminals and the GAN.

Radio traffic (e.g., speech information or data) between mobile terminals and base stations is typically encrypted to ensure that the information being passed remains confidential. Although some service networks (e.g., GSM) encrypt traffic, most other service networks do not. Consequently, a GAN should be capable of encrypting traffic for those service networks that do not have that capability.

-4-

However, since a GAN does not know the identity of its users (the service network subscribers), it must be capable of encrypting radio traffic using encryption keys that are created without knowing a subscribing terminal's identity or authenticity. Unfortunately, most existing mobile communications networks use encryption techniques that generate encryption keys by using authentication parameters. In other words, to encrypt radio traffic in a conventional mobile communications network, the user terminal's identity must be known.

SUMMARY OF THE INVENTION

It is an object of the present invention to encrypt communications between a mobile terminal and a communications network without requiring the network to know the identity of the terminal.

It is also an object of the present invention to encrypt communications between a plurality of mobile terminals and a communications network without requiring the network to maintain individual encryption keys for each of the terminals.

It is another object of the present invention to encrypt communications between a mobile terminal and a communications network without requiring the terminal to permanently store a secret encryption key.

It is yet another object of the present invention to minimize call setup time, minimize transmission delays, and maximize data throughput, while encrypting communications between a mobile terminal and a communications network.

In accordance with one aspect of the present invention, a method is provided for encrypting communications between a communications network and a communications terminal, by storing a public key associated with the network at the terminal, generating a secret key at the terminal, encrypting the secret key with the stored public key at the terminal, transmitting the encrypted secret key from the terminal, receiving the encrypted secret key at the network, decrypting the received encrypted secret key with a private key, where the private key is associated with the public key, and encrypting the ensuing traffic with the secret key. If a public key has not been stored at the terminal, then the terminal transmits a request to the

-5-

network for a public key. As such, the network is not required to know the identity of the terminal in order to maintain encrypted communications with the terminal.

In accordance with another aspect of the present invention, the foregoing and other objects are achieved by a method and an apparatus for encrypting traffic
5 between a communications network and a communications terminal by broadcasting a (asymmetric) public key from the network. The public key is received by the terminal. The network maintains a private key that can be used to decrypt information encrypted with the public key. The terminal generates and stores a naturally occurring random number as a secret session (symmetric) key, encrypts the
10 symmetric session key with the public key, and transmits the encrypted session key to the network. The network decrypts the session key with the private key, and both the network and terminal encrypt the ensuing communications with the secret session key. Again, the communications network is not required to know the identity of the terminal in order to maintain encrypted communications with the terminal.

15

BRIEF DESCRIPTION OF THE DRAWINGS

A more complete understanding of the method and apparatus of the present invention may be had by reference to the following detailed description when taken in conjunction with the accompanying drawings wherein:

20 FIGURE 1 is a perspective view of an exemplary generic access network connected to a plurality of service networks and service network subscribers;

FIGURE 2 is a top level schematic block diagram of a generic access network in which a method of encrypting radio traffic between service networks and service network subscribers can be implemented, in accordance with a preferred
25 embodiment of the present invention;

FIGURE 3 is a schematic block diagram of the access network illustrated in FIGURE 2;

FIGURE 4 is a sequence diagram that illustrates a method that can be used to encrypt radio communications between a generic access network and a terminal,
30 in accordance with a preferred embodiment of the present invention; and

-6-

FIGURE 5 is a block diagram of a method that can be used to certify the authenticity of a public key and the owner of the key with a digital signature, in accordance with a preferred embodiment of the present invention.

5 DETAILED DESCRIPTION OF THE DRAWINGS

The preferred embodiment of the present invention and its advantages are best understood by referring to FIGURES 1-5 of the drawings, like numerals being used for like and corresponding parts of the various drawings.

Essentially, in accordance with a preferred embodiment of the present
10 invention, a mobile terminal stores at least one public key, along with a unique identification character of at least one GAN associated with that public key, in a memory location. A GAN continuously broadcasts its unique identification character in all cells connected to that GAN. When contact is initiated between the terminal and that GAN, the terminal compares the received identifier with the stored
15 identifier(s), and if a match can be made, the terminal generates a random secret key, encrypts the secret key with the public key associated with that GAN's identifier, and transmits the encrypted secret key. The GAN decrypts the secret key using a private key associated with the public key. The secret key is used thereafter by the terminal and the GAN to encrypt and decrypt the ensuing radio traffic.
20 Notably, the GAN can maintain secure communications with the terminal without ever knowing the terminal's identity. Furthermore, since the GAN does not need to know the identity of such a terminal, the GAN is not required to maintain a database of individual terminal encryption keys. Additionally, the terminal is not required to store its own secret key, because it can generate a new secret key for
25 each communications session.

FIGURE 2 is a top level schematic block diagram of a generic access network in which a method of encrypting radio traffic between service networks and service network subscribers can be implemented, in accordance with a preferred embodiment of the present invention. A GAN 100 is shown, which includes a
30 transport network 102 interconnected with an access network 104. A plurality of service networks (e.g., PLMN, ISDN, PSTN, INTERNET, VOD) are connected

-7-

through respective access ports (e.g., 106, 108, 110, 112, 114) to transport network 102 and access network 104. Access network 104 includes a plurality of RNCs and associated base stations (e.g., RNC(1)-RNC(N)). The plurality of RNCs and associated base stations are connected by a respective radio interface to a plurality of mobile transceivers (terminals) 116, 118, 120 and 122. A user of each mobile terminal is a subscriber to at least one of the service networks PLMN, etc. The mobile terminals can communicate with their respective service networks in the manner described above with respect to FIGURE 1. More specifically, the RNCs control communications between the terminals and their respective service networks. Notably, although a plurality of mobile terminals (116, etc.) are shown in FIGURE 2, this is for illustrative purposes only. One or more fixed radio terminals may also be connected to GAN 100 and are thus capable of communicating with at least one of the service networks.

FIGURE 3 is a schematic block diagram of access network 104 illustrated in FIGURE 2. Access network 104 includes a plurality of RNCs (e.g., RNC(1)-RNC(N)). However, although a plurality of RNCs is shown for this embodiment, the present invention can be implemented with only one RNC. At least one service network (e.g., 130, 132, 134) is connected through at least one respective access port (e.g., AP1, AP(N-1), AP(N)) to at least one RNC. At least one base station (e.g., BS(1), BS(N)) is connected to a respective RNC (e.g., RNC(1), RNC(N)). Although a plurality of base stations is shown, the present invention can be implemented with only one base station.

A mobile terminal (e.g., cellular phone 118) is connected by a radio interface to base station BS(1). It should be readily understood that one terminal (118) is shown for illustrative purposes only and that one or more additional terminals could be shown. The RNCs (e.g., RNC(1)-RNC(N)) are interconnected by communications lines (136, 138) for communications therebetween. Consequently, terminal 118 can establish communications with any of the service networks (e.g., 130, 132, 134) through access network 104 and GAN 100 (FIGURE 2). Notably, the coverage provided for each service network can be enlarged by switching to a different access port of access network 104. In other words, terminal 118 can

-8-

communicate with service network 132 through RNC(1), interconnecting line 136, and RNC(N-1). Alternatively, if service network 132 is switched to access port AP(1), terminal 118 can communicate with service network 132 through RNC(1).

5 FIGURE 4 is a sequence diagram that illustrates a method that can be used to encrypt radio communications between a generic access network and a terminal, in accordance with a preferred embodiment of the present invention. The method 200 of encrypting communications can begin at the GAN or the terminal. For example, in this embodiment, at step 204, the GAN (e.g., 10) continuously broadcasts a unique identification character in all cells connected to that GAN. The
10 terminal (e.g., 118) contains a non-volatile memory located in a GAN section of the terminal. The terminal stores at least one public key in the non-volatile memory. Along with each public key, the terminal also stores a respective expiration date for the key, and a GAN identification character that identifies a specific GAN associated with that key. In other words, each public key stored in the terminal's memory is
15 thereby associated with a specific GAN. The terminal initiates contact by registering with a GAN (but not necessarily setting up a call). A processor in the terminal compares the received GAN identifier with the stored identifiers, and if a match can be made (and the key has not expired), the processor retrieves the stored public key associated with the identified GAN. However, in the event that no such match is
20 found, the terminal sends a request for the GAN to transmit a public key. The transmitted public key (and its expiration date) is stored in the terminal and can be used to encrypt a secret key in the current and ensuing communication sessions.

 At step 206, the terminal generates a (symmetric) secret key (described in detail below). At step 208, the terminal uses the retrieved public key to encrypt the
25 secret key. At step 210, the terminal transmits the encrypted secret key to the identified GAN. At step 212, the GAN decrypts the secret key, which, at step 214, is used by the GAN and the terminal for encrypting traffic during the ensuing communications session (described in detail below).

 Alternatively, at the end of a session with a GAN, the terminal stores the
30 public key used for that session. When the terminal or a GAN begins a new communications session, the terminal retrieves the public key stored from the last

-9-

session with a GAN, and uses that public key to encrypt a secret key to be used for the ensuing session. If the use of that stored public key is unsuccessful, the terminal then sends a request to the GAN for a new public key. This technique advantageously increases network throughput, because a network channel is not tied up transmitting a public key. However, if a public key has not been stored from a past session with a particular GAN, the terminal can still receive the public key by requesting it from the GAN and using it to encrypt a secret key that will be used for the ensuing session. In any event, by storing the relatively large (bit-wise) public keys in the terminal, as opposed to transmitting them from the GAN, radio transmission delays can be reduced significantly, a substantial amount of network transmission time can be saved, and data throughput will be increased.

FIGURE 4 also illustrates a method that can be used to encrypt radio communications between a generic access network and a mobile terminal, in accordance with another embodiment of the present invention. For example, when communications are desired between a service network and a terminal (e.g., PLMN and terminal 118), the service network or terminal can initiate communications with a call setup message. At step 202, as the initial connection between the GAN and the terminal is established, the service network can request that the ensuing traffic will be encrypted. If so, at step 204, still during the initial call setup process, the terminal receives a public key which is continuously broadcast from one or more base stations (e.g., BS(1)-BS(N)).

In this embodiment, all of the RNCs maintain at least one public key/private key pair (the same pair in every RNC) in a memory storage location. The public key that was broadcast by the GAN is received by the terminal (118) that has initiated contact with that GAN. Preferably, both the call setup procedure and the procedure to transfer the public key is performed by an RNC, which is connected through an access port to the service network of interest (e.g., RNC(1) to AP(1) to PLMN 130). Alternatively, a base station (e.g., BS1) can be configured to maintain public/private key pairs and broadcast or otherwise transfer a public key to a terminal.

-10-

The RNC can broadcast the public key in all cells in the RNC's coverage area. Consequently, since the GAN broadcasts the public key instead of having the terminal request the key from the GAN, the terminal can register with the GAN much faster, and a call can be set up in a substantially shorter period of time. Alternatively, instead of broadcasting the public key in a plurality of cells, the RNC can transfer the public key directly through the base station that has established contact with the terminal. However, the method of broadcasting the public key in a plurality of cells before call setup advantageously decreases the load on the GAN's dedicated traffic channels.

For all embodiments, as long as the terminal is registered with the GAN, the same public key can be used for all subsequent communications with that GAN, because the same key is stored at the GAN and also at the terminal. Alternatively, the public key can be changed periodically in accordance with a predetermined scheme or algorithm, or even at the whim of the GAN operator. If an operator desires to change public keys periodically, storing each public key's expiration date at the terminal facilitates their use in that regard. Furthermore, in the preferred embodiment, when the public key is changed, it can be broadcast by the GAN for a predetermined period of time, to minimize the number of terminal requests for a new public key.

As described earlier, at step 202, the GAN can maintain one or more asymmetric public key/private key pairs. In that event, a so-called "RSA Algorithm" can be used to create the public key/private key pairs. The RSA Algorithm combines the difficulty of factoring a prime number with the ease of generating large prime numbers (using a probabilistic algorithm) to split an encryption key into a public part and a private part.

Specifically, assuming that the letters P and Q represent prime numbers, the letter M represents an unencrypted message, and the letter C represents the encrypted form of M, the RSA Algorithm can be expressed as follows:

$$M^E \text{ mod } PQ = > C \text{ (encrypted message M)} \quad (1)$$

$$C^D \text{ mod } PQ = > M \text{ (decrypted message C)} \quad (2)$$

-11-

where the term $(DE-1)$ is a multiple of $(P-1)(Q-1)$. In this embodiment, the exponent E is set to 3. The private and public keys are each composed of two numbers. For example, the numbers represented by (PQ, D) make up the private key, and the numbers represented by (PQ, E) make up the public key. Since the same value for E is used consistently, only the PQ portion of the number need be sent on request or broadcast and used for the public key (e.g., at step 204). By knowing the private key, any message encrypted with the public key can be decrypted.

Returning to FIGURE 4, at step 206, the terminal (118) receives and/or stores the asymmetric public key. The terminal generates a random symmetric secret key. The random secret key, which is used to encrypt communications preferably for the complete session, can be generated in at least one of four ways. Using one method, the terminal takes several samples from measurements of the strength of the received signal, concatenates the lower order bits of the several samples, and processes the result to produce a random number. Since the lower order bits of the received signal are well within the noise level of the received signal, a naturally occurring, truly random number is generated. A second random number generating method is to use the random noise signal created at the input of an A/D converter connected to a microphone. Again, using this method, a naturally occurring, truly random number can be generated for the secret key. A third random number generating method is for the terminal to take samples from phase measurements of the received signal, concatenate the lower order bits of the samples, and process the result to produce a random number. A fourth random number generating method is for the terminal to take samples from the encoding section of the speech codec, concatenate the lower order bits of the samples, and process the result to produce the random number.

Alternatively, a random number generated at the terminal can be used as a seed for a pseudorandom number generator. The seed is encrypted with the public key from the GAN, and transmitted to the GAN. The seed is used simultaneously in the GAN and the terminal to produce a pseudorandom number. The

-12-

pseudorandom number thus generated can be used by the GAN and the terminal as the secret key for the ensuing communications session.

The session key can be changed periodically to a different number in the pseudorandom number sequence. For example, the session key can be changed for a number of reasons, such as after a predetermined amount of data has been encrypted, or after traffic has been encrypted for a predetermined amount of time. The terminal or the GAN can initiate a change of the secret key, or the key can be changed according to a predetermined scheme or algorithm. For example, a request to change the secret session key can be implemented by transmitting a "session key change request" message, or by setting a "session key change request" bit in the header of a transmitted message.

Additionally, shorter session keys can be generated and less complicated encryption algorithms can be used with the pseudorandom number generation method described above. Consequently, a substantial amount of processing power can be saved in the GAN and especially in the terminal. The terminal can be configured to select the length of the session key to be used, in order to address trade offs between security and computational requirements. For example, the terminal's processor can select the length of a secret session key by generating a session key at that length, or by specifying the number of bits to be used from the output of the pseudorandom number generator. Alternatively, the terminal can specify the range of the output of the pseudorandom number generator to set a predetermined length.

Other alternative methods may be used to generate a pseudorandom number for a secret session key. For example, using a "Lagged Fibonacci" type of pseudorandom number generator, the n^{th} number in the pseudorandom number sequence, N_n , can be calculated as follows:

$$N_n = (N_{n-k} - N_{n-l}) \bmod M \quad (3)$$

where k and l are the so-called lags, and M defines the range of the pseudorandom numbers to be generated. For optimum results, the largest lag should be between 1000 and 10000. If a relatively long key is desired, a plurality of the pseudorandom numbers produced by equation 3 can be concatenated to produce a longer key. If

-13-

the pseudorandom numbers produced by equation 3 are to be floating point numbers between 0 and 1, M can be set to 1. The bit patterns of such floating point pseudorandom numbers can be used as symmetric encryption keys.

Another pseudorandom number generator that can be used to create a secret session key is based on an algorithm that produces pseudorandom numbers uniformly distributed between 0 and 1. Specifically, the seeds X_0 , Y_0 and Z_0 of the pseudorandom numbers N_n are initially set to integer values between 1 and 30000. The pseudorandom numbers N_n are then calculated as follows:

$$X_n = 171 * (X_{n-1} \bmod 177) - (2 * X_{n-1} / 177) \quad (4)$$

$$Y_n = 172 * (Y_{n-1} \bmod 176) - (35 * Y_{n-1} / 176) \quad (5)$$

$$Z_n = 170 * (Z_{n-1} \bmod 178) - (63 * Z_{n-1} / 178) \quad (6)$$

If any of the values of X_n , Y_n or Z_n are less than zero, respectively, then X_n is set equal to $X_n + 30269$, Y_n is set equal to $Y_n + 30307$, or Z_n is set equal to $Z_n + 30323$. The pseudorandom numbers N_n are then equal to $((X_n / 30269 + Y_n / 30307 + Z_n / 30323) \bmod 1)$, where X_n , Y_n and Z_n are floating point numbers, and "amod" means that these numbers can be fractions. The floating point numbers generated with this algorithm form bit patterns that are suitable for use as symmetric encryption keys. The length of such keys can be extended by concatenating a plurality of the pseudorandom numbers generated.

Returning to the method illustrated by FIGURE 4, at step 208, preferably using the above-described RSA Algorithm, the terminal encrypts the secret symmetric key with the public key. For example, assume that the secret symmetric key generated at the terminal is represented by the letters SK. Using equation 1 of the RSA Algorithm, the secret key is encrypted as follows:

$$M^E \bmod PQ = > C$$

where (PQ, E) represents the public key, M is equal to SK, and C is the encrypted version of SK. The exponent E is set to 3.

In the preferred embodiment, the terminal places the encrypted secret key into a message format, which includes a header and message field. The header provides control information associated with the encrypted secret key that follows in the message field. A bit in the header can be set to indicate that the message field

-14-

that follows the header is encrypted. In other words, only the secret key field of the message is encrypted. The header of the message is transmitted in the clear. Consequently, a substantial amount of network processing time can be saved at the RNC, since the header indicates whether the subsequent message field is encrypted, and if so, only that portion of the message is to be decrypted.

5 At step 210, the terminal (118) transmits the encrypted secret key (C) to the GAN via the contacted base station (e.g., BS(1)). In the preferred embodiment, this secret key is used for the ensuing communications. Alternatively, at any time during the ensuing communications session, the terminal can generate a new secret key, encrypt it with the public key, and transmit the new encrypted secret key to the GAN. The security of the session is thereby increased, because by reducing the amount of time that a particular secret key is used for a session, the likelihood that the secret key will be broken by an unauthorized user is also reduced.

10 At step 212, the RNC (e.g., RNC(1)) receives the encrypted secret key (C) from the base station, and decrypts the secret key using the private key part of the RSA Algorithm. For example, using equation 2 (above) of the RSA Algorithm, the received encrypted secret key (C) is decrypted as follows:

$$C^D \text{ mod } PQ = > M$$

where (PQ, D) represents the private key, and M is equal to SK (secret key).

20 At step 214, the ensuing radio traffic between the RNC and the terminal is encrypted and decrypted with the secret key, which is now known to both the RNC and the terminal. A known symmetric encryption algorithm can be used to encrypt and decrypt the ensuing radio traffic with the secret key, such as, for example, a one, two or three pass Data Encryption Standard (DES) algorithm, or a Fast Encipherment Algorithm (FEAL).

25 As yet another encryption alternative, instead of using the RSA Algorithm to create a public/private key pair, a so-called Diffie-Hellman "exponential key exchange" algorithm can be used to let the terminal and the GAN agree on a secret session key. In using this encryption scheme, two numbers (α , q) are stored at the GAN. At the beginning of a communications session, the RNC transmits the two numbers directly (or broadcasts the numbers) to the terminal. The numbers α and

30

-15-

q are required to meet the following criteria: q is a large prime number that defines the finite (Galios) field $GF(q) = 1, 2, \dots, q-1$; and α is a fixed primitive element of $GF(q)$. In other words, the exponents (x) of $(\alpha^x \text{ mod } q)$ produce all of the elements 1,2,..., q-1 of $GF(q)$. In order to generate an agreed to secret session key, the two numbers (α , q) are transmitted directly (or broadcast) from the GAN to the terminal. 5 Alternatively, the two numbers can be already resident in the terminal's non-volatile memory. The terminal (118) generates the random number $X_T(1 < X_T < q-1)$, and computes the value of $Y_T = \alpha^{X_T} \text{ mod } q$. The GAN (e.g., the RNC or base station) generates the random number $X_G(1 < X_G < q-1)$, and computes the value of $Y_G = \alpha^{X_G} \text{ mod } q$. The random numbers can be generated at the terminal using the methods described above with respect to generating naturally occurring, truly random numbers. 10

Y_T and Y_G are transferred unencrypted to the respective GAN and terminal. Upon receipt of the number Y_G , the terminal calculates the value of $K_S = Y_G^{X_T} \text{ mod } q = \alpha^{X_G X_T} \text{ mod } q$. Upon receipt of the number Y_T , the GAN calculates the value of $K_S = Y_T^{X_G} \text{ mod } q = \alpha^{X_T X_G} \text{ mod } q$. The number X_T is kept secret at the terminal, the number X_G is kept secret at the GAN, but the value of K_S is now known at both the terminal and the GAN. The number K_S is therefore used by both as the communications session encryption key. An unauthorized user would not know either X_T or X_G and would have to compute the key K_S from Y_T and Y_G , which is a prohibitive computational process. A significant security advantage of using the exponential key exchange algorithm is that the GAN is not required to maintain secret private key data on a permanent basis. 20

In summary, when a communications session is first initiated between a GAN and a terminal, the terminal receives an asymmetric public key that has been continuously broadcast by the GAN, retrieved from the terminal's internal memory, or requested from the GAN. The GAN maintains a private key that can be used to decrypt information encrypted with the public key. The terminal generates and stores a naturally occurring random number as a secret session (symmetric) key, encrypts the symmetric session key with the public key, and transmits the encrypted session key to the GAN. The GAN decrypts the session key with the private key, 30

-16-

and both the GAN and terminal encrypt the ensuing communications with the secret session key. A primary technical advantage of transferring a public key from a GAN to a terminal at the onset of communications is that the GAN is not required to know the identity of the terminal in order to have encrypted communications with the terminal. However, a problem can arise if an unauthorized user attempts to impersonate a GAN and transmits a public key to the terminal. In that event, as described below, the terminal can be configured to authenticate the received public key and the identity of the GAN.

For example, when a public key is to be transferred from a GAN to a terminal, the key can be transferred with a public key "certificate". This certificate provides proof that the associated public key and the owner of that key are authentic. A "trusted" third party can issue the public key along with the certificate, which includes a "digital signature" that authenticates the third party's identity and the public key. The certificate can also contain the GAN's identity and the expiration date of the certificate, if any.

In one aspect of the invention, the GAN transmits the certificate and public key to the terminal. In that case, the public key of the third party is pre-stored (a priori) at the subscribing terminals.

FIGURE 5 is a block diagram of a method that can be used to certify the authenticity of a public key and the owner of the key with a digital signature, in accordance with the present invention. The method (300) of digitally signing a public key certificate and verifying its authenticity begins at step 302. At step 302, a "certificate" containing unencrypted information about the owner of the public key to be transferred to a terminal is prepared by a trusted third party. The unencrypted information also includes the public key and the expiration date of the certificate. At step 304, the resulting "unsigned" certificate is processed with an irreversible algorithm (e.g., a hashing algorithm) to produce a message digest at step 306, which is a digested or shortened version of the information included on the certificate. At step 308, the digest information is encrypted with a private key of a different public/private key pair. Preferably, an RSA algorithm similar to equations 1 and 2 above is used to derive this key pair. At step 310, a digitally signed public key

-17-

certificate is thereby produced that contains the originally unencrypted information (including the public key to be used for the communications session) and the digest information, which is now encrypted with the certificate issuer's private key. The digitally signed public key certificate is then transferred to the terminal that has initiated contact with the GAN.

5 At step 312, upon receiving the digitally signed certificate, the terminal's processor analyzes the unencrypted and encrypted portions of the document. At step 314, the unencrypted information is processed using an algorithm identical to the hashing algorithm used at step 304. At step 316, a second digested version of the unencrypted information is produced at the terminal. At step 318, the terminal's processor retrieves the pre-stored certificate issuer's public key from memory, and using an RSA algorithm, decrypts the encrypted digest information from the certificate. Another version of the unencrypted digested information is thereby produced at step 320. At step 322, the terminal compares the two versions of the unencrypted digested information, and if the compared information is identical, the certificate's signature and the session public key are assumed to be authentic. That certified public key can now be used by the terminal to encrypt the secret session key.

10 Although a preferred embodiment of the method and apparatus of the present invention has been illustrated in the accompanying Drawings and described in the foregoing Detailed Description, it will be understood that the invention is not limited to the embodiments disclosed, but is capable of numerous rearrangements, modifications and substitutions without departing from the spirit of the invention as set forth and defined by the following claims.

25

-18-

WHAT IS CLAIMED IS:

1. A method for encrypting communications traffic between a mobile communications network and a communications terminal, comprising the steps of:
storing a public key and a first identifier associated with said mobile
communications network at said communications terminal;
5 comparing said first identifier stored at said communications terminal with
a second identifier received from said mobile communications network and
producing a first predetermined result;
generating a secret key at said communications terminal;
10 encrypting said secret key with said stored public key at said communications
terminal; and
transmitting said encrypted secret key from said communications terminal.
2. The method according to Claim 1, further comprising the steps of:
15 receiving said encrypted secret key at said mobile communications network;
decrypting said received encrypted secret key with a private key, said private
key associated with said public key; and
encrypting said communications traffic with said secret key.
- 20 3. The method according to Claim 1, wherein the step of storing a
public key comprises the step of a priori pre-storing the public key.
4. The method according to Claim 1, further comprising the step of
transmitting said public key from said mobile communications network upon
25 receiving a public key request from said communications terminal.
5. The method according to Claim 4, wherein the step of transmitting
said public key further comprises the step of transmitting information to authenticate
said public key.

30

-19-

6. The method according to Claim 4, further comprising the step of transmitting said request from said communications terminal upon said comparing step producing a second predetermined result.

5 7. The method according to Claim 1, wherein the steps of receiving and decrypting said encrypted secret key are performed at a radio base station in said mobile communications network.

8. The method according to Claim 1, wherein the step of decrypting said received encrypted secret key is performed at a radio network controller in said mobile communications network.

9. The method according to Claim 1, wherein said mobile communications network comprises a generic communications network.

15 10. The method according to Claim 1, wherein said communications terminal comprises a mobile terminal.

11. The method according to Claim 1, wherein said communications terminal comprises a fixed terminal.

12. The method according to Claim 1, wherein said communications terminal comprises an unidentified communications terminal.

25 13. The method according to Claim 1, wherein said mobile communications network comprises a cellular phone network.

14. The method according to Claim 1, further comprising the steps of: connecting a plurality of service networks to said mobile communications network, a user of said communications terminal being a subscriber to at least one of said plurality of service networks; and

30

-20-

providing a communications path between said communications terminal and said at least one of said plurality of service networks.

5 15. The method according to Claim 1, wherein said private key and said public key are associated by an RSA Algorithm.

16. The method according to Claim 1, wherein said secret key comprises a symmetric encryption key.

10 17. The method according to Claim 1, wherein the step of generating a secret key comprises the step of generating a naturally occurring random number.

18. The method according to Claim 1, wherein the step of generating a secret key comprises the steps of:
15 detecting a received signal in digital form at said communications terminal;
and
extracting at least one low order bit from said detected received signal.

19. The method according to Claim 1, wherein the step of generating a secret key comprises the steps of:
20 detecting a signal at an output of a microphone A/D converter; and
extracting at least one low order bit from said detected output signal.

20. The method according to Claim 1, wherein the step of generating a secret key comprises the steps of:
25 detecting a signal at an output of a speech codec; and
extracting at least one low order bit from said detected output signal.

30 21. The method according to Claim 1, wherein the step of generating a secret key comprises the steps of:

-21-

generating a seed for a pseudorandom number; and
generating a pseudorandom number from said seed.

22. The method according to Claim 1, wherein a length of said secret key
5 is predetermined at said communications terminal.

23. The method according to Claim 1, wherein said secret key further
comprises a plurality of concatenated numbers.

10 24. The method according to Claim 1, wherein the step of storing said
public key and said first identifier further comprises storing an expiration date
associated with said public key.

15 25. The method according to Claim 24, wherein said communications
terminal transmits a public key request to said mobile communications network if
said public key has expired.

20 26. The method according to Claim 1, further comprising the steps of:
changing said public key at said mobile communications network; and
storing said changed public key at said communications terminal.

25 27. The method according to Claim 26, wherein the step of changing said
public key further comprises the step of broadcasting said changed public key from
said mobile communications network for a predetermined period of time.

28. A method for encrypting traffic between a generic
communications network and a first communications terminal, comprising the steps
of:

30 broadcasting a public key from said generic communications network to a
plurality of communications terminals, said plurality of communications terminals
including said first communications terminal;

-22-

generating a secret key at said first communications terminal;
encrypting said secret key with said public key at said first communications terminal;
5 transmitting said encrypted secret key from said first communications terminal;
receiving said encrypted secret key at said generic communications network;
decrypting said received encrypted secret key with a private key, said private key associated with said public key; and
10 encrypting said traffic with said secret key.

29. The method according to Claim 28, wherein the broadcasting step further comprises the steps of:

transferring said public key from a radio network controller to at least one base station in said generic communications network; and
15 transmitting said public key from said at least one base station.

30. The method according to Claim 28, wherein said broadcasting step comprises the step of transmitting said public key from a plurality of base stations in said generic communications network.
20

31. The method according to Claim 28, wherein said first communications terminal comprises an unidentified communications terminal.

32. The method according to Claim 28, wherein the step of broadcasting said public key further comprises the step of broadcasting information to authenticate said public key.
25

33. The method according to Claim 28, wherein the step of broadcasting said public key further comprises the step of transmitting, on request, information to authenticate said public key.
30

-23-

34. A method for encrypting communications traffic between a mobile communications network and a communications terminal, comprising the steps of:

storing two numbers associated with a Diffie-Hellman exponential key exchange algorithm and a first identifier associated with said mobile communications network at said communications terminal;

5 comparing said first identifier stored at said communications terminal with a second identifier received from said mobile communications network and producing a first predetermined result;

generating a first random number at said communications terminal;

10 generating a second random number at said mobile communications network;

and

using said first and second random numbers as inputs to said Diffie-Hellman exponential key exchange algorithm, generating a third number to be used as a secret key by said communications terminal and said mobile communications network.

15

35. The method according to Claim 34, wherein the step of storing two numbers comprises the step of a priori pre-storing said two numbers.

20 36. The method according to Claim 34, further comprising the step of transmitting said two numbers from said mobile communications network upon receiving a request for said two numbers from said communications terminal.

25 37. The method according to Claim 36, further comprising the step of transmitting said request from said communications terminal upon said comparing step producing a second predetermined result.

30 38. The method according to Claim 34, wherein the step of storing said two numbers and said first identifier further comprises storing an expiration date associated with said two numbers.

-24-

39. The method according to Claim 38, wherein said communications terminal transmits a request for two new numbers associated with said Diffie-Hellman exponential key exchange algorithm if said two numbers has expired.

5 40. The method according to Claim 34, further comprising the steps of:
changing said two numbers associated with a Diffie-Hellman exponential key
exchange algorithm at said mobile communications network; and
storing said changed two numbers at said communications terminal.

10 41. The method according to Claim 40, wherein the step of changing said
two numbers further comprises the step of broadcasting said changed two numbers
from said mobile communications network for a predetermined period of time.

15 42. A method for encrypting traffic between a generic communications
network and a first communications terminal, comprising the steps of:
broadcasting two numbers associated with an exponential key exchange
algorithm from said generic communications network to a plurality of
communications terminals, said plurality of communications terminals including said
first communications terminal;
20 generating a first random number at said first communications terminal;
generating a second random number at said generic communications network;
using said first and second random numbers as inputs to said exponential key
exchange algorithm, generating a third number to be used as a secret key by said
first communications terminal and said generic communications network;
25 and encrypting said traffic with said secret key.

43. A system for use in encrypting traffic between a generic
communications network and a communications terminal, comprising:
an access network included in said generic communications network; and
30 access network means coupled to said communications terminal and
associated with said access network, for storing a public encryption key associated

-25-

with said generic communications network, generating a secret key, encrypting said secret key with said stored public encryption key, and transmitting said encrypted secret key to said generic communications network.

5 44. A system for use in encrypting traffic between a generic communications network and a communications terminal, comprising:

first network means for storing a private encryption key, distributing a public encryption key, and decrypting an encrypted secret session key;

10 second network means connected to said first network means, for broadcasting said distributed public encryption key, said first and second network means associated with an access network of said generic communications network; and

15 access network means coupled to said communications terminal and associated with said access network of said generic communications network, for receiving said broadcast public encryption key, generating a secret key, encrypting said secret key with said received public encryption key, and transmitting said encrypted secret key to said generic communications network.

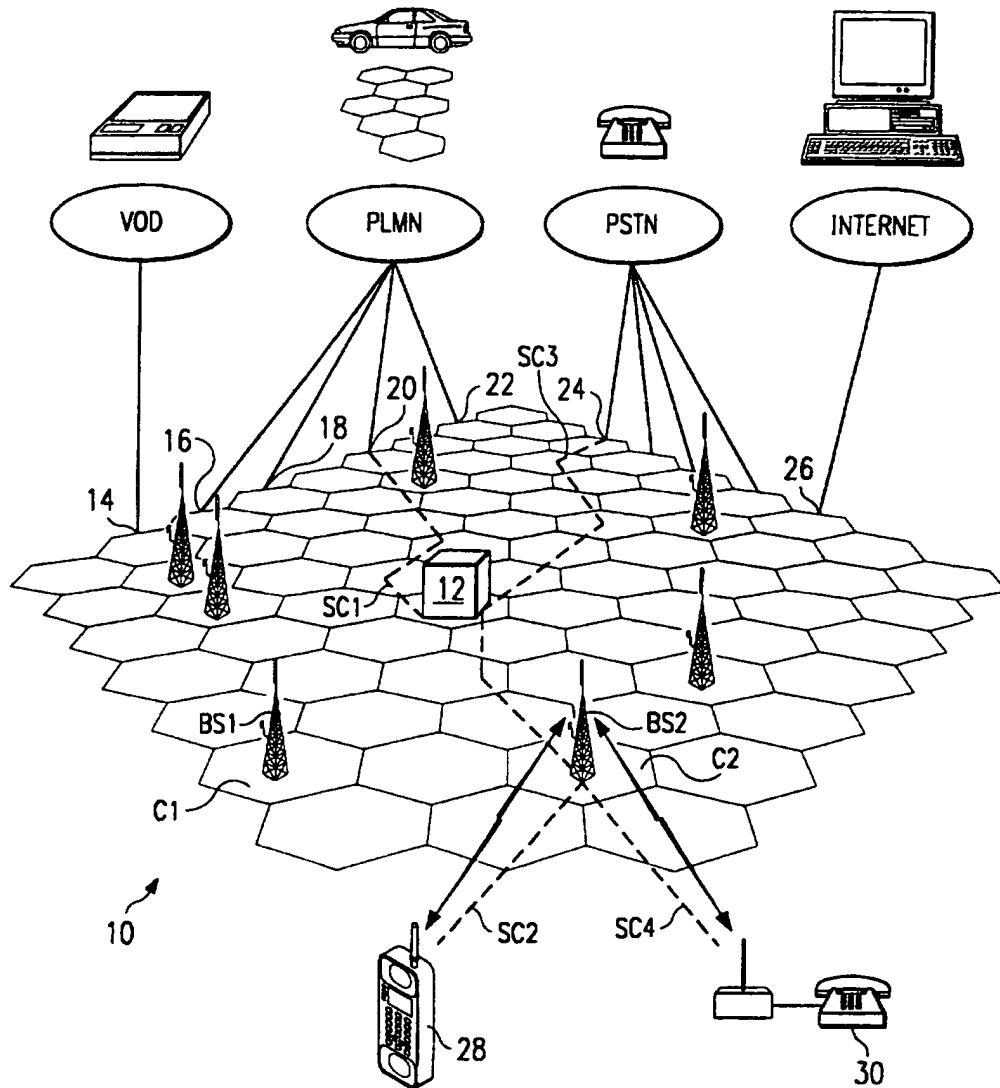


FIG. 1

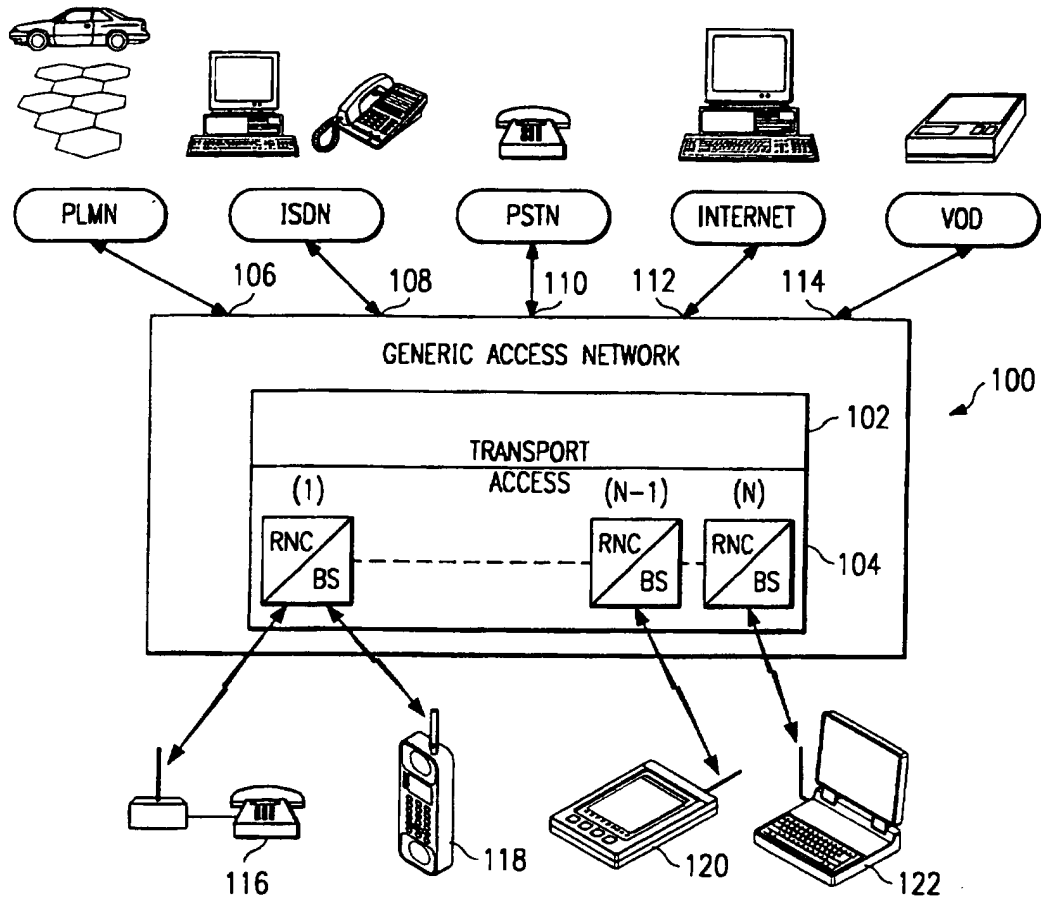


FIG. 2

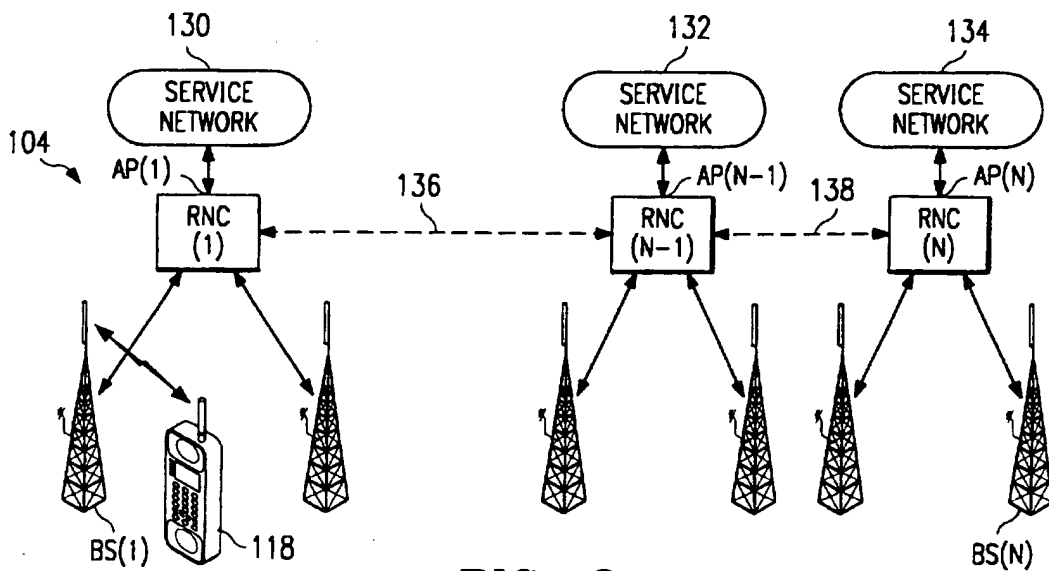


FIG. 3

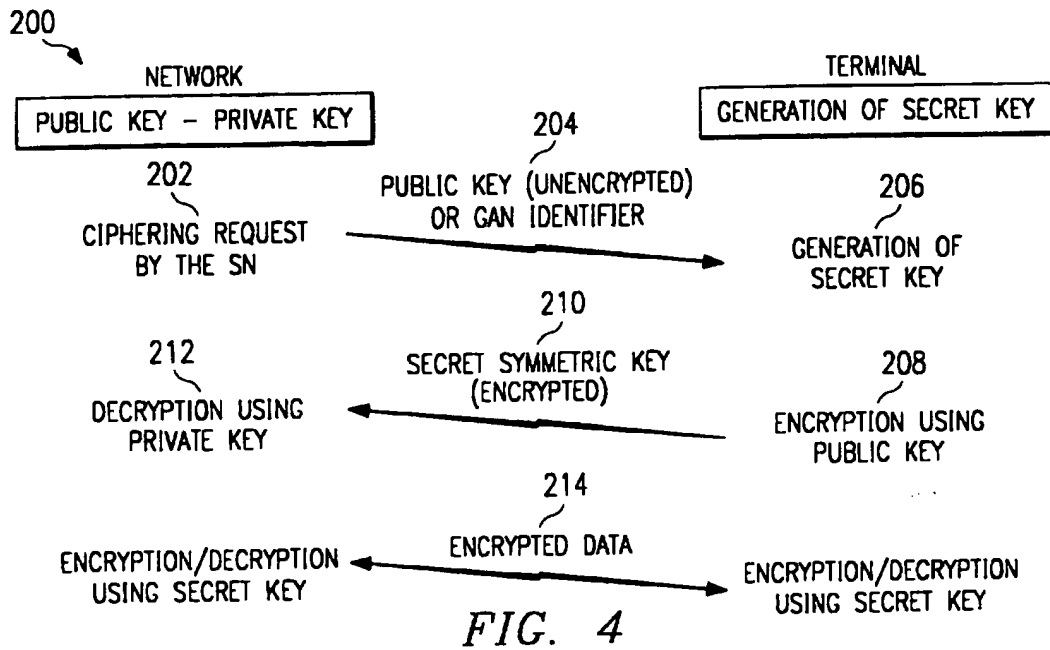


FIG. 4

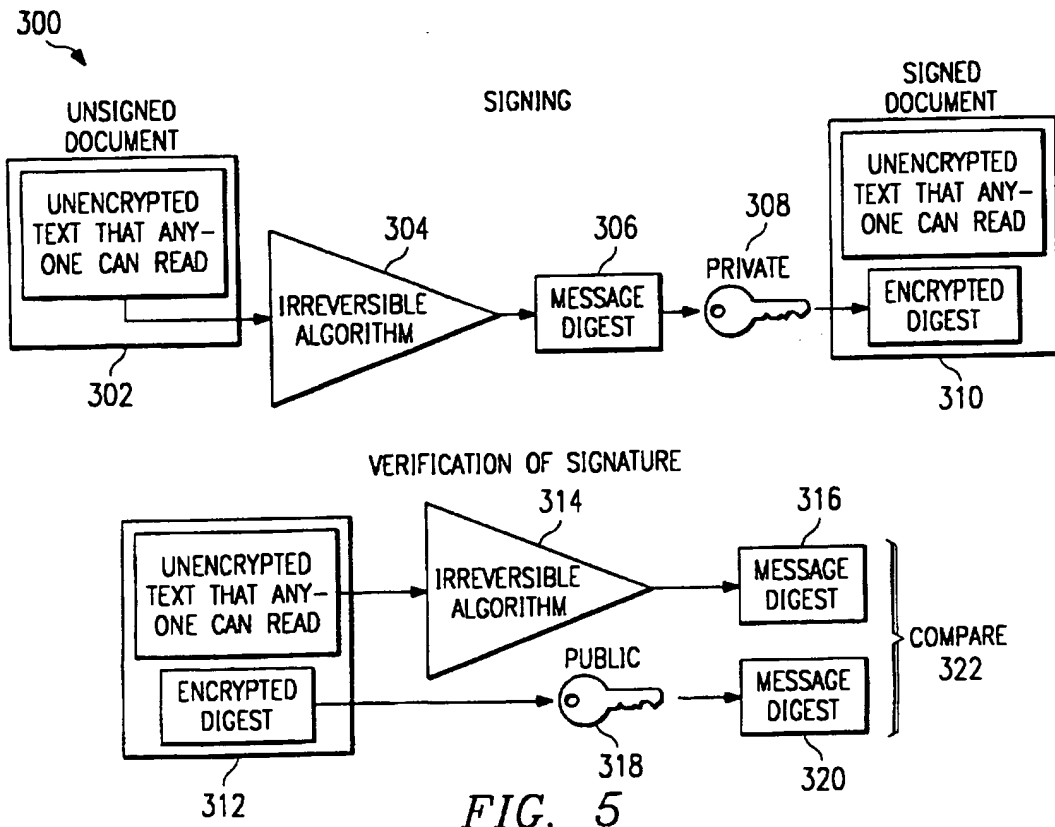


FIG. 5

INTERNATIONAL SEARCH REPORT

International Application No

PCT/SE 97/01407

A. CLASSIFICATION OF SUBJECT MATTER
 IPC 6 H04L9/08 H04Q7/32

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
 IPC 6 H04L H04Q

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 5 222 140 A (BELLER ET AL.) 22 June 1993 see column 4, line 57 - column 5, line 3 see column 5, line 13 - line 37 ---	1, 2, 7, 10, 28, 29, 34, 42-44
A	GB 2 297 016 A (KOKUSAI DENSHIN DENWA) 17 July 1996 see page 19, line 11 - page 21, line 2; figure 7 --- -/--	28, 44

Further documents are listed in the continuation of box C.

Patent family members are listed in annex.

* Special categories of cited documents :

- "A" document defining the general state of the art which is not considered to be of particular relevance
- "E" earlier document but published on or after the international filing date
- "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- "O" document referring to an oral disclosure, use, exhibition or other means
- "P" document published prior to the international filing date but later than the priority date claimed

- "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
- "&" document member of the same patent family

Date of the actual completion of the international search

19 November 1997

Date of mailing of the international search report

02/12/1997

Name and mailing address of the ISA
 European Patent Office, P.B. 5818 Patentlaan 2
 NL - 2280 HV Rijswijk
 Tel. (+31-70) 340-2040, Tx. 31 651 epo.nl,
 Fax: (+31-70) 340-3016

Authorized officer

 Holper, G

INTERNATIONAL SEARCH REPORT

International Application No

PCT/SE 97/01407

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	<p>MEVEL F ET AL: "Distributed communication services in the Masix system" CONFERENCE PROCEEDINGS OF THE 1996 IEEE FIFTEENTH ANNUAL INTERNATIONAL PHOENIX CONFERENCE ON COMPUTERS AND COMMUNICATIONS (CAT. NO.96CH35917), CONFERENCE PROCEEDINGS OF THE 1996 IEEE FIFTEENTH ANNUAL INTERNATIONAL PHOENIX CONFERENCE ON COMPUTERS AND, ISBN 0-7803-3255-5, 1996, NEW YORK, NY, USA, IEEE, USA, pages 172-178, XP000594787 see page 174, right-hand column, line 25 - line 29 see page 176, right-hand column, line 26 - page 177, left-hand column, last line; figure 5</p>	28,43,44
A	<p style="text-align: center;">---</p> <p>PATENT ABSTRACTS OF JAPAN vol. 95, no. 008 & JP 07 203540 A (N T T IDOU TSUUSHINMOU KK), 4 August 1995, see abstract</p>	1,34
A	<p style="text-align: center;">---</p> <p>EP 0 067 977 A (SIEMENS) 29 December 1982 see abstract; figure 2</p> <p style="text-align: center;">-----</p>	24

1

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No

PCT/SE 97/01407

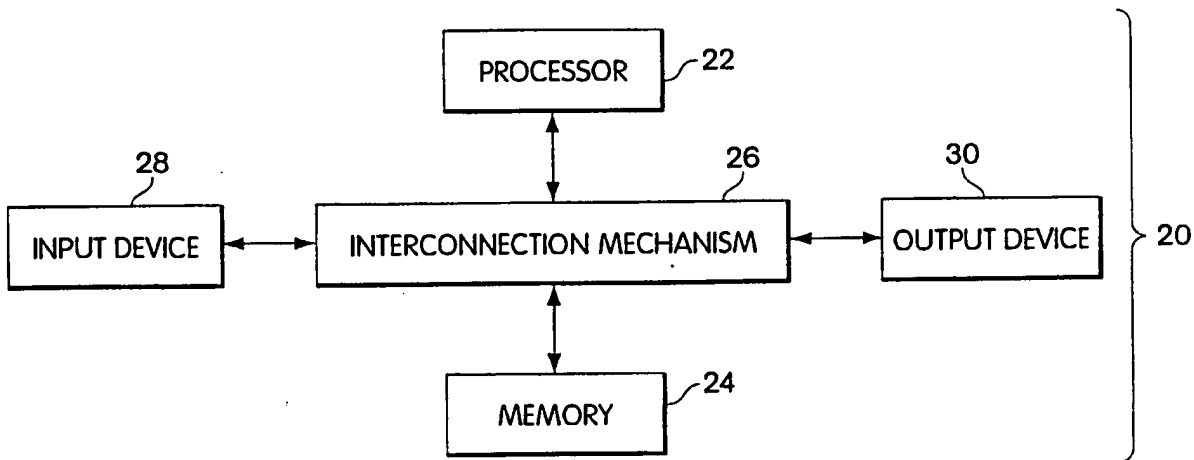
Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 5222140 A	22-06-93	NONE	
GB 2297016 A	17-07-96	JP 8195741 A	30-07-96
EP 67977 A	29-12-82	DE 3123167 C	24-02-83



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

<p>(51) International Patent Classification ⁶ : H04L 9/00</p>	<p>A1</p>	<p>(11) International Publication Number: WO 98/11690 (43) International Publication Date: 19 March 1998 (19.03.98)</p>
<p>(21) International Application Number: PCT/US97/16223 (22) International Filing Date: 12 September 1997 (12.09.97) (30) Priority Data: 60/025,991 12 September 1996 (12.09.96) US 08/887,723 3 July 1997 (03.07.97) US (71)(72) Applicant and Inventor: GLOVER, John, J. [US/US]; 26 Amaranth Avenue, Medford, MA 02155 (US). (74) Agent: GORDON, Peter, J.; Wolf, Greenfield & Sacks, P.C., 600 Atlantic Avenue, Boston, MA 02210 (US).</p>	<p>(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, HU, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, UZ, VN, YU, ZW, ARIPO patent (GH, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).</p> <p>Published <i>With international search report. Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.</i></p>	

(54) Title: SELF-DECRYPTING DIGITAL INFORMATION SYSTEM AND METHOD



(57) Abstract

The claimed data protection device (20) includes a processor (22) connected to a memory system (24) through an interconnection mechanism (26). An input device (28) is also connected to the processor (22) and memory system (24) through the interconnection mechanism (26). The interconnection mechanism (26) is typically a combination of one or more buses and one or more switches. The output device (30) may be a display, and the input device (28) may be a keyboard and/or mouse or other cursor control device.

* (Referred to in PCT Gazette No. 32/1998, Section II)

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

SELF-DECRYPTING DIGITAL INFORMATION SYSTEM AND METHOD**Field of the Invention**

The present invention is related to mechanisms for protecting digital information from being copied. In particular, the present invention is related to mechanisms which permit authorized execution of computer program code or access to other digital information which is
5 encrypted or otherwise encoded.

Background of the Invention

A serious problem which faces the electronic publishing and software industries is the ease with which digital information can be copied without authorization from the publisher.
10 Digital information also may be used or modified without authorization. For example, computer software may be reverse engineered or attacked by computer viruses.

There are many mechanisms available which may be used to limit or prevent access to digital information. Such mechanisms often either restrict the ability of the user to make back-up copies or involve the use of special purpose hardware to limit access to the digital information.
15 For example, some mechanisms restrict the use of digital information to a particular machine. See, for example, U.S. Patent 4,817,140. Other mechanisms require the digital information to be stored on a particular recording medium in order to be used. See, for example, U.S. Patent 5,412,718. Yet other mechanisms allow only a certain number of uses of the digital information. See for example, U.S. Patent 4,888,798. Many of these access control mechanisms cause
20 distribution to be more costly.

Several other patents describe a variety of systems for encryption, compression, licensing and royalty control and software distribution such as: U.S. Pat. No. 4,405,829, U.S. Pat. No. 4,864,616, U.S. Pat. No. 4,888,800, U.S. Pat. No. 4,999,806, U.S. Pat. No. 5,021,997, U.S. Patent No. 5,027,396, U.S. Pat. No. 5,033,084, U.S. Pat. No. 5,081,675, U.S. Pat. No.
25 5,155,847, U.S. Pat. No. 5,166,886, U.S. Pat. No. 5,191,611, U.S. Pat. No. 5,220,606, U.S. Pat. No. 5,222,133, U.S. Pat. No. 5,272,755, U.S. Pat. No. 5,287,407, U.S. Pat. No. 5,313,521, U.S. Pat. No. 5,325,433, U.S. Pat. No. 5,327,563, U.S. Pat. No. 5,337,357, U.S. Pat. No. 5,351,293, U.S. Pat. No. 5,341,429, U.S. Pat. No. 5,351,297, U.S. Pat. No. 5,361,359, U.S. Pat. No. 5,379,433, U.S. Pat. No. 5,392,351, U.S. Pat. No. 5,394,469, U.S. Pat. No. 5,414,850, U.S. Pat.
30 No. 5,473,687, U.S. Pat. No. 5,490,216, U.S. Pat. No. 5,497,423, U.S. Pat. No. 5,509,074, U.S.

Pat. No. 5,511,123, U.S. Pat. No. 5,524,072, U.S. Pat. No. 5,532,920, U.S. Pat. No. 5,555,304, U.S. Pat. No. 5,557,346, U.S. Pat. No. 5,557,765, U.S. Pat. No. 5,592,549, U.S. Pat. No. 5,615,264, U.S. Pat. No. 5,625,692, and U.S. Pat. No. 5,638,445.

Computer programs or other digital information also may be encrypted in order to prevent an individual from making a useful copy of the information or from reverse engineering a program. Even with such encryption, however, a computer program must be decrypted in order for a computer to load and execute the program. Similarly, other digital information must be decrypted before it can be accessed and used. Generally, digital information is decrypted to disk, and not to main memory of the computer which is more protected by the operating system, because decryption to main memory results in a significant loss of memory resources. If the purpose for using encryption is to prevent users from copying the digital information, then decryption of the information to accessible memory for use defeats this purpose.

One way to protect digital information using encryption has been made available by International Business Machines (IBM) and is called a "CRYPTOLOPE" information container. This technology is believed to be related to U.S. Patent Nos. 5,563,946 and 5,598,470 (to Cooper et al.), and published European patent applications 0679977, 0679978, 0679979 and 0681233. The CRYPTOLOPE system requires a user to have a "helper application" and a key. The CRYPTOLOPE information container is generated by IBM. The content provider submits data to IBM, which in turn encrypts and packages the data in a CRYPTOLOPE information container. The helper application is a form of memory resident program, called a terminate and stay resident (TSR) program, which is a form of input/output (I/O) device driver installed in the operating system and which monitors requests from the operating system for files on specified drives and directories. Because the TSR program must know the directory, and/or file name to be accessed, that information also is available to other programs. Other programs could use that information to manipulate the operation of the TSR program in order to have access to decrypted contents of the information container. The encrypted information container includes an executable stub which is executed whenever the application is run without the installed TSR program or from a drive not monitored by the TSR program to prevent unpredictable activity from executing encrypted code. This stub may be used to install decryption and cause the application be executed a second time, or to communicate with the TSR program to instruct the TSR program to monitor the drive. It may be preferable from the point of view of the content provider however to maintain an encryption process and keys independently of any third party.

Multimedia content, such as a movie or hypertext presentation also may be stored on a digital versatile disk (DVD), sometimes called a digital video disk, compact disk read-only memory (CD-ROM), rewriteable compact disks (CD-RW) or other medium in an encrypted digital format for use with special-purpose devices. For example, concern about illegal copying
5 of content from digital video disks or other digital media has resulted in a limited amount of content being available for such devices. This problem has caused representatives of both multimedia providers and digital video disk manufacturers to negotiate an agreement on an encryption format for information stored on DVDs. This copy protection scheme is licensed through an organization called the CSS Interim Licensing organization. However, in this
10 arrangement, the content provider is limited to using the agreed upon encryption format and a device manufacturer is limited to using a predetermined decryption system.

Encryption has also been used to protect and hide computer viruses. Such viruses are typically polymorphic, i.e., they change every time they infect a new program, and are encrypted. The virus includes a decryption program that executes to decrypt the virus every time the
15 infected program is run. Such viruses are described, for example, in "Computer Virus-Antivirus Coevolution" by Carey Nachenberg, Communications of the ACM, Vol. 40, No. 1, (Jan. 1997), p. 46 et seq. Such viruses include decryption keys within them since, clearly, their execution is not carried out by the user and a user would not be asked for authorization keys to permit execution of the viruses. Additionally, such viruses are typically only executed once at the start
20 of execution of an infected program and permanently return control to the infected program after execution.

Summary of the Invention

Some of these problems with digital information protection systems may be overcome
25 by providing a mechanism which allows a content provider to encrypt digital information without requiring either a hardware or platform manufacturer or a content consumer to provide support for the specific form of corresponding decryption. This mechanism can be provided in a manner which allows the digital information to be copied easily for back-up purposes and to be transferred easily for distribution, but which should not permit copying of the digital information
30 in decrypted form. In particular, the encrypted digital information is stored as an executable computer program which includes a decryption program that decrypts the encrypted information

to provide the desired digital information, upon successful completion of an authorization procedure by the user.

In one embodiment, the decryption program is executed as a process within a given operating system and decrypts the digital information within the memory area assigned to that process. This memory area is protected by the operating system from copying or access by other processes. Even if access to the memory area could be obtained, for example through the operating system, when the digital information is a very large application program or a large data file, a copy of the entire decrypted digital information is not likely to exist in the memory area in complete form.

By encrypting information in this manner, a platform provider merely provides a computer system with an operating system that has adequate security to define a protected memory area for a process and adequate functionality to execute a decryption program. The content provider in turn may use any desired encryption program. In addition, by having a process decrypt information within a protected memory area provided by the operating system, the decrypted information does not pass through any device driver, memory resident program or other known logical entity in the computer system whose behavior may be controlled to provide unauthorized access to the data. The ability to reverse engineer or attack a computer program with a computer virus also may be reduced.

In another embodiment, the decryption program is part of a dynamically loaded device driver that responds to requests for data from the file containing the encrypted data. When the digital information product is first executed, this device driver is extracted from the file and is loaded into the operating system. The executed digital information product then informs the loaded device driver of the location of the hidden information in the file, any keys or other passwords, and the name of a phantom directory and file to be called that only the digital information product and the device driver know about. The name of this directory may be generated randomly. Each segment of hidden information in the digital information product may be assigned its own unique file name in the phantom directory. The digital information product then makes a call to the operating system to execute one of the files in the phantom directory. The loaded driver traps these calls to the operating system, accesses the original file, decrypts the desired information and outputs the desired information to the operating system.

In combination with other mechanisms that track distribution, enforce royalty payments and control access to decryption keys, the present invention provides an improved method for

identifying and detecting sources of unauthorized copies. Suitable authorization procedures also enable the digital information to be distributed for a limited number of uses and/or users, thus enabling per-use fees to be charged for the digital information.

Accordingly, one aspect of the invention is a digital information product including a
5 computer-readable medium with digital information stored thereon. The digital information includes computer program logic having a first portion of executable computer program logic and a second portion of digital information. The first portion of executable program logic, when executed, defines a mechanism for responding to requests for digital information from an operating system of a computer. This mechanism, when used to access the second portion of the
10 encrypted digital information, decrypts the encrypted digital information, and provides the encrypted digital information to the operating system.

In the foregoing aspect of the invention, the digital information may be executable computer program logic. Hence, one aspect of the invention is a computer program product, including a computer readable medium with computer program logic stored thereon. The
15 computer program logic includes a first portion of executable computer program logic and a second portion of encrypted computer program logic. The first portion of executable computer program logic, when executed, defines a mechanism for responding to requests for computer program logic from an operating system of a computer. This mechanism accesses the second portion of encrypted computer program logic, decrypts the encrypted computer program logic,
20 and provides the decrypted computer program logic to the operating system.

Another aspect of the present invention is a computer program product, a computer system and a process which produce a computer program or digital information product in accordance with other aspects of the invention, using executable program code for the first and second portions of the desired computer program product.

25 Another aspect of the present invention is a computer program product including a self-decrypting encrypted executable computer program. The product includes a computer readable medium having computer program logic stored thereon. The computer program logic defines first, second and third modules, wherein the third module defines the encrypted executable computer program. The first module, when executed by a computer, defines a mechanism for
30 loading the second module into memory of the computer. The second module, when executed by a computer, defines a mechanism for communicating with an operating system of the computer to receive requests for program code from the encrypted executable computer program from the

third module, and for processing the requests to access and decrypt the encrypted executable computer program and for providing the decrypted executable code from the third module to the operating system.

Another aspect of the invention is a process for executing encrypted executable
5 computer programs on a computer system having a processor, memory and operating system. The process involves receiving computer program logic having a first module defining a start up routine, a second module, and a third module containing the encrypted executable computer program. The first module of the received computer program logic is executed using the processor. When the first module is executed, the second module is caused to be loaded into the
10 memory of the computer system. Requests are generated from the operating system for data from the encrypted executable computer program and are received by the second module. The second module accesses and decrypts the encrypted executable computer program in response to these requests and returns the decrypted executable computer program to the operating system.

These and other aspects, advantages and features of the present invention and its
15 embodiments will be more apparent given the following detailed description.

Brief Description of the Drawing

In the drawing,

Fig. 1 is a block diagram of a typical computer system with which the present invention
20 may be implemented;

Fig. 2 is a block diagram of a memory system in the computer system of Fig. 1;

Fig. 3 is a diagram of a computer program or digital information product which may be recorded on a computer readable and writable medium, such as a magnetic disc;

Fig. 4 is a flowchart describing how the computer program or digital information
25 product of Fig. 3 is used;

Fig. 5 is a flowchart describing operation of an example unwrap procedure as shown in Fig. 3 in one embodiment of the invention;

Fig. 6 is a flowchart describing operation of an example device driver as shown in Fig. 3 in one embodiment of the invention;

Fig. 7 is a block diagram of a computer system in the process of executing a computer
30 program product in accordance with one embodiment of the invention;

Fig. 8 is a flowchart describing operation of an example unwrap procedure in another embodiment of the invention; and

Fig. 9 is a flowchart describing how a computer program product such as shown in Fig. 3 is constructed.

5

Detailed Description

The present invention will be more completely understood through the following detailed description which should be read in conjunction with the attached drawing in which similar reference numbers indicate similar structures.

10 Embodiments of the present invention may be implemented using a general purpose digital computer or may be implemented for use with a digital computer or digital processing circuit. A typical computer system 20 is shown in Fig. 1, and includes a processor 22 connected to a memory system 24 via an interconnection mechanism 26. An input device 28 also is connected to the processor and memory system via the interconnection mechanism, as is an
15 output device 30. The interconnection mechanism 26 is typically a combination of one or more buses and one or more switches. The output device 30 may be a display and the input device may be a keyboard and/or a mouse or other cursor control device.

It should be understood that one or more output devices 30 may be connected to the computer system. Example output devices include a cathode ray tube (CRT) display, liquid
20 crystal display (LCD), television signal encoder for connection to a television or video tape recorder, printers, communication devices, such as a modem, and audio output. It also should be understood that one or more input devices 28 may be connected to the computer system. Example input devices include a keyboard, keypad, trackball, mouse, pen and tablet, communication device, audio or video input and scanner. It should be understood that the
25 invention is not limited to the particular input or output devices used in combination with the computer system or to those described herein.

The computer system 20 may be a general purpose computer system, which is programmable using a high level computer programming language, such as "C++," "Pascal,"
"VisualBasic." The computer system also may be implemented using specially programmed,
30 special purpose hardware. In a general purpose computer system, the processor is typically a commercially available processor, such as the Pentium processor from Intel Corporation. Many other processors are also available. Such a processor executes a program called an operating

system, such as Windows 95 or Windows NT 4.0, both available from Microsoft Corporation, which controls the execution of other computer programs and provides scheduling, debugging, input output control, accounting compilation, storage assignment, data management and memory management, and communication control and related services. Other examples of operating systems include: MacOS System 7 from Apple Computer, OS/2 from IBM, VMS from Digital Equipment Corporation, MS-DOS from Microsoft Corporation, UNIX from AT&T, and IRIX from Silicon Graphics, Inc.

The computer system 20 also may be a special purpose computer system such as a digital versatile disk or digital video disk (DVD) player. In a DVD player, there is typically a decoder controlled by some general processor which decodes an incoming stream of data from a DVD. In some instances, the DVD player includes a highly integrated DVD decoder engine. Such devices generally have a simple operating system which may be modified to include the capabilities described and used herein in connection with the typical operating systems in a general purpose computer. In particular, some operating systems are designed to be small enough for installation in an embedded system such as a DVD player, including the WindowsCE operating system from Microsoft Corporation and the JavaOS operating system from SunSoft Corporation. The operating system allows a content provider to provide its own programs that define some of the content, which is particularly useful for interactive multimedia. This capability also can be used to provide encryption and decryption, in accordance with the invention.

The processor and operating system define a computer platform for which application programs in a programming language such as an assembly language or a high level programming language are written. It should be understood that the invention is not limited to a particular computer platform, operating system, processor, or programming language. Additionally, the computer system 20 may be a multi-processor computer system or may include multiple computers connected over a computer network.

An example memory system 24 will now be described in more detail in connection with Fig. 2. A memory system typically includes a computer readable and writable non-volatile recording medium 40, of which a magnetic disk, a flash memory, rewriteable compact disk (CD-RW) and tape are examples. The recording medium 40 also may be a read only medium such as a compact disc-read only memory (CD-ROM) or DVD. A magnetic disk may be removable, such as a "floppy disk" or "optical disk," and/or permanent, such as a "hard drive." The disk,

which is shown in Fig. 2, has a number of tracks, as indicated at 42, in which signals are stored, in binary form, i.e., a form interpreted as a sequence of 1's and 0's, as shown at 44. Such signals may define an application program to be executed by the microprocessor, or information stored on the disk to be processed by the application program. Typically, in the operation of a general purpose computer, the processor 22 causes data to be read from the non-volatile recording medium 40 into an integrated circuit memory element 46, which is typically a volatile random access memory, such as a dynamic random access memory (DRAM) or static random access memory (SRAM). The integrated circuit memory element 46 allows for faster access to the information by the processor than disk 40, and is typically called the system or host memory.

10 The processor generally causes the data to be manipulated within the integrated circuit memory 46 and may copy the data to the disk 40, if modified, when processing is completed. A variety of mechanisms are known for managing data movement between the disk 40 and the integrated circuit memory 46, and the invention is not limited thereto. It should also be understood that the invention is not limited to a particular memory system.

15 The file system of a computer generally is the mechanism by which an operating system manages manipulation of data between primary and secondary storage, using files. A file is a named logical construct which is defined and implemented by the operating system to map the name and a sequence of logical records of data to physical storage media. An operating system may specifically support various record types or may leave them undefined to be interpreted or controlled by application programs. A file is referred to by its name by application programs and is accessed through the operating system using commands defined by the operating system. An operating system provides basic file operations provided by for creating a file, opening a file, writing a file, reading a file and closing a file.

25 In order to create a file, the operating system first identifies space in the storage media which is controlled by the file system. An entry for the new file is then made in a directory which includes entries indicating the names of the available files and their locations in the file system. Creation of a file may include allocating certain available space to the file. Opening a file returns a handle to the application program which it uses to access the file. Closing a file invalidates the handle.

30 In order to write data to a file, an application program issues a command to the operating system which specifies both an indicator of the file, such as a file name, handle or other descriptor, and the information to be written to the file. Given the indicator of the file, the

operating system searches the directory to find the location of the file. The directory entry stores a pointer, called the write pointer, to the current end of the file. Using this pointer, the physical location of the next available block of storage is computed and the information is written to that block. The write pointer is updated in the directory to indicate the new end of the file.

5 In order to read data from a file, an application program issues a command to the operating system specifying the indicator of the file and the memory locations assigned to the application where the next block of data should be placed. The operating system searches its directory for the associated entry given the indicator of the file. The directory may provide a pointer to a next block of data to be read, or the application may program or specify some offset
10 from the beginning of the file to be used.

A primary advantage of using a file system is that, for an application program, the file is a logical construct which can be created, opened, written to, read from and closed without any concern for the physical storage used by the operating system.

The operating system also allows for the definition of another logical construct called a
15 process. A process is a program in execution. Each process, depending on the operating system, generally has a process identifier and is represented in an operating system by a data structure which includes information associated with the process, such as the state of the process, a program counter indicating the address of the next instruction to be executed for the process, other registers used by process and memory management information including base and bounds
20 registers. Other information also may be provided. The base and bounds registers specified for a process contain values representing the largest and smallest addresses that can be generated and accessed by an individual program. Where an operating system is the sole entity able to modify these memory management registers, adequate protection from access to the memory locations of one process from another process is provided. As a result, this memory management information
25 is used by the operating system to provide a protected memory area for the process. A process generally uses the file system of the operating system to access files.

The present invention involves storing encrypted digital information, such an audio, video, text or an executable computer program, on a computer readable medium such that it can be copied easily for back-up purposes and transferred easily for distribution, but also such that it
30 cannot be copied readily in decrypted form during use. In particular, the digital information is stored as a computer program that decrypts itself while it is used to provide the digital information, e.g., to provide executable operation code to the operating system of a computer, as

the digital information is needed. Any kind of encryption or decryption may be used and also may include authorization mechanisms and data compression and decompression. In one embodiment of the present invention, decrypted digital information exists only in memory accessible to the operating system and processes authorized by the operating system. When the digital information is a large application program, a copy of the entire decrypted application program is not likely to exist in the main memory at any given time, further reducing the likelihood that a useful copy of decrypted code could be made. The decryption operation also is performed only if some predetermined authorization procedure is completed successfully.

One embodiment of the invention, in which the decryption program is a form of dynamically loaded device driver, will first be described. Fig. 3 illustrates the structure of digital information as stored in accordance with one embodiment of the present invention, which may be stored on a computer readable medium such as a magnetic disc or compact disc read only memory (CD-ROM) to form a computer program product. The digital information includes a first portion 50, herein called an unwrap procedure or application, which is generally unencrypted executable program code. The purpose of the unwrap procedure is to identify the locations of the other portions of the digital information, and may perform other operations such as verification. In particular, the unwrap procedure identifies and extracts a program which will communicate with the operating system, herein called a virtual device driver 52. The unwrap procedure may include decryption and decompression procedures to enable it to decrypt/decompress the driver, and/or other content of this file. The program 52 need not be a device driver. The virtual device driver 52 typically follows the unwrap procedure 50 in the file container, the digital information. The virtual device driver, when executed, decrypts and decodes the desired digital information such as an executable computer program code from hidden information 54, which may be either encrypted and/or encoded (compressed). It is the decrypted hidden information which is the desired digital information to be accessed. This hidden information may be any kind of digital data, such as audio, video, text, and computer program code including linked libraries or other device drivers.

In this embodiment of the computer program product, labels delineate the boundaries between the device driver and the hidden files. These labels may or may not be encrypted. A first label 56 indicates the beginning of the code for the virtual device driver 52. A second label 58 indicates the end of the virtual device driver code. Another label 60 indicates the beginning of the hidden information and a label 62 indicates the end of that application. There may be one

or more blocks of such hidden information, each of which can be given a different name. It may be advantageous to use the name of the block of information in its begin and end tags. This computer program product thus contains and is both executable computer program code and one or more blocks of digital information. A table of locations specifying the location of each
5 portion of the product could be used instead of labels. Such a table could be stored in a predetermined location and also may be encrypted.

The overall process performed using this computer program product in one embodiment of the invention will now be described in connection with Fig. 4. This embodiment may be implemented for use with the Windows95 operating system and is described in more detail in
10 connection with Figs. 5-7. An embodiment which may be implemented for use on the WindowsNT 4.0 operating system is described in more detail below in connection with Fig. 8. In both of these described embodiments, the digital information is an executable computer program which is read by the operating system as data from this file and is executed. The same principle of operation would apply if the data were merely audio, video, text or other information
15 to be conveyed by a user. In the embodiment of Fig. 4, the computer program is first loaded into memory in step 70, and the unwrap procedure 50 is executed by the operating system, as any typical executable computer program is executed. The unwrap procedure may perform authorization, for example by checking for a required password or authentication code, and may receive any data needed for decryption or decompression, for example keys or passwords, in step
20 72. Suitable authorization procedures may provide the ability to distribute software for single use. The unwrap procedure locates the virtual device driver 52 within the computer program in step 74, and then locates the hidden application in step 76. The virtual device driver 52 is then extracted by the unwrap procedure from the computer program, copied to another memory location and loaded for use by the operating system in step 78. An advantage of an operating
25 system like Windows95 is that it allows such device drivers to be loaded dynamically without restarting the computer.

The executed unwrap procedure 50, in step 80, informs the loaded virtual device driver 52 of the location of the hidden information in the file, any keys or other passwords, and a name of a phantom directory and file to be called that only the unwrap procedure and the virtual device
30 driver know about. The name of this phantom directory may be generated randomly. Each segment information hidden in the digital information product may be assigned its own unique file name in the phantom directory.

After the loaded virtual device driver 52 receives all communications from the unwrap procedure, it opens the original application file for read only access in step 82. The unwrap procedure then makes a call to the operating system in step 84 to execute the file in the phantom directory for which the name was transmitted to the loaded virtual device driver. One function of the loaded virtual device driver 52 is to trap all calls from the operating system to access files in step 86. Any calls made by the operating system to access files in the phantom directory are processed by the virtual device driver, whereas calls to access files in other directories are allowed to proceed to their original destination. In response to each call from the operating system, the virtual device driver obtains the bytes of data requested by the operating system from the original computer program file in step 88. These bytes of data are then decrypted or decompressed in step 90 and returned to the operating system. When processing is complete, the phantom application is unloaded from the operating system in step 92, and may be deleted from the memory.

A more detailed description of the process of Fig. 4 will now be described in connection with Figs. 5-7. Fig. 5 is a flowchart describing the operation of one embodiment of the unwrap procedure in more detail. The first step performed by this procedure is identifying the operating system being used, in step 100. This step is useful because different methods may be used with different operating systems. All code that may be used to run in various operating systems may be placed in this unwrap procedure. This procedure also may contain the decompression/decryption code, for example or any other computer program code to be executed.

The executed application then opens the original executable file as a data file and searches for the begin and end tags of the device driver and hidden files in step 102. The device driver code is copied into memory and loaded into the operating system in step 104. The unwrap procedure then informs the device driver of the name of the original application file, offsets of the hidden files and the name of a phantom directory, which is typically randomly generated (step 106). This communication may be performed using a "DeviceIOControl" function call in the Windows95 operating system. The unwrap procedure then makes a call to the operating system to execute the hidden file in the phantom directory, in step 108.

The operation of one embodiment of a device driver will now be described in connection with Fig. 6. After the device driver is loaded into the operating system, it hooks into a position between the operating system and a file system driver (FSD), in step 110, to intercept

calls made by the operating system to the FSD for data from files in the phantom directory. The FSD is the code within the operating system that performs physical reading and writing of data to disk drives. The operating system makes requests to the FSD for data from files in directories on the disk drives. The driver then receives information from the unwrap procedure including the
5 name of the original file, the location of hidden files within the original file, and the name of the phantom directory created by the unwrap procedure (step 112). The device driver opens the original file as a read only data file. The device driver now traps calls, in step 114, made from the operating system for files in the phantom directory. Calls to other directories are ignored and passed on to the original destination. The device driver then reads the data from the original data
10 file, decrypts and decompresses it, and returns the decrypted/decompressed data to the operating system in step 116.

For example, if the offset for the hidden application in the original data file is 266,270 bytes and the operating system asks for 64 bytes starting at offset 0 of the hidden application in the phantom directory, the device driver reads 64 bytes from the original file starting at offset
15 266,270, decrypts/decompresses those 64 bytes, and returns the first 64 decrypted/decompressed bytes back to the operating system. From the point of view of the operating system, the 64 bytes appear to have come from the file in the phantom directory. Steps 114 and 116 are performed on demand in response to the operating system.

A block diagram of the computer system in this embodiment, with a device driver
20 loaded and in operation, will now be described in more detail in connection with Fig. 7. Fig. 7 illustrates the operating system 120, the loaded device driver 122, a file system driver 124, the original executable file 126 as it may appear on disk and the unwrap procedure 128. The executable file may in fact be on a remote computer and accessed through a network by the device driver. The unwrap procedure causes the operating system to begin execution of the
25 hidden file by issuing an instruction to execute the file in the phantom directory, as indicated at 130. This command is issued after the device driver 122 is informed of the file name of the original executable file 126, offsets of the hidden files within that file and the name of the phantom directory, as indicated at 132. The operating system then starts making calls to the phantom directory as indicated at 134. The device driver 122 traps these calls and turns them
30 into requests 136 to the file system driver to access the original executable file 126. Such requests actually are made to the operating system 120, through the device driver 122 to the file system driver 124. The file system driver 124 returns encrypted code 138 to the device driver

122. The encrypted code 138 actually passes back through the device driver 122 to the operating system 120 which in turn provides the encrypted code 138 to the device driver 122 as the reply to the request 136 for the original file. The device driver 122 then decrypts the code to provide decrypted code 140 to the operating system 120.

5 Another embodiment of the invention will now be described in connection with Fig. 8. This embodiment may be implemented using the WindowsNT 4.0 operating system, for example. In this embodiment, the device driver portion 52 of the computer program product is not used. The unwrap procedure for this embodiment begins by identifying the operating system being used similar, which is step 100 in Fig. 5. If the operating system is Windows NT 4.0, for
10 example, a different unwrap procedure for this embodiment is performed. Before describing this unwrap procedure, a brief description of some of the available operating system commands will be provided.

 Currently, under all versions of the Window operating system or operating environment from Microsoft Corporation (such as Windows 3.1, Windows 95 and Windows NT 3.51 and 4.0)
15 all executable files (.exe) or dynamic link library (.dll and .ocx) files, which are executable files with different header and loading requirements than .exe files, that are loaded into memory by the operating system must reside as a file either locally, e.g., on a disk drive or remotely, e.g., over a network or communications port. All further references herein to loading an executable will be using the Win32 function calls used in Windows 95 and NT 3.51 and 4.0 operating
20 systems. The CreateProcess() function which loads files with an .exe extension takes ten parameters:

```

BOOL CreateProcess(// Prototype from Microsoft Visual C++ Help Documentation
    LPCTSTR lpApplicationName,           // pointer to name of executable module
25  LPTSTR lpCommandLine,               // pointer to command line string
    LPSECURITY_ATTRIBUTES lpProcessAttributes, // pointer to process security attributes
    LPSECURITY_ATTRIBUTES lpThreadAttributes, // pointer to thread security attributes
    BOOL bInheritHandles,               // handle inheritance flag
    DWORD dwCreationFlags,             // creation flags
30  LPVOID lpEnvironment,               // pointer to new environment block
    LPCTSTR lpCurrentDirectory,        // pointer to current directory name
    LPSTARTUPINFO lpStartupInfo,       // pointer to STARTUPINFO
    LPPROCESS_INFORMATION lpProcessInformation // pointer to PROCESS_INFORMATION
);

```

Three of these parameters are pointers to strings that contain an application file name, command line parameters, and the current directory. The other parameters are security, environmental, and process information. The LoadLibrary() function takes one parameter that is a pointer to a string that contains the application file name:

```

5
HINSTANCE LoadLibrary(// Prototype from Microsoft Visual C++ Help Documentation
    LPCTSTR lpLibFileName    // address of filename of executable module
);

```

10 The LoadLibraryEx() function takes three parameters the first being the same as LoadLibrary(), the second parameter must be null, and the third tells the operating system whether to load the file as an executable or as a data file in order to retrieve resources such as icons or string table data from it and not load it as an executable:

```

15 HINSTANCE LoadLibraryEx(// Prototype from Microsoft Visual C++ Help Documentation
    LPCTSTR lpLibFileName,    // points to name of executable module
    HANDLE hFile,             // reserved, must be NULL
    DWORD dwFlags            // entry-point execution flag
);

```

20

The CreateFile() function is used to create and open files and to load files such as device drivers. This function also requires a pointer to a string that contains the name of a physical file:

```

HANDLE CreateFile(// Prototype from Microsoft Visual C++ Help Documentation
25 LPCTSTR lpFileName,          // pointer to name of the file
    DWORD dwDesiredAccess,      // access (read-write) mode
    DWORD dwShareMode,         // share mode
    LPSECURITY_ATTRIBUTES lpSecurityAttributes, // pointer to security descriptor
    DWORD dwCreationDistribution, // how to create
30 DWORD dwFlagsAndAttributes,  // file attributes
    HANDLE hTemplateFile       // handle to file with attributes to copy
);

```

There are other functions such as `MapViewOfFile()` and `MapViewOfFileEx()` that map areas of memory to an already opened physical file through a handle to that file. They have the following parameters:

```

5  LPVOID MapViewOfFile(// Prototype from Microsoft Visual C++ Help Documentation
    HANDLE hFileMappingObject,      // file-mapping object to map into address space
    DWORD dwDesiredAccess,          // access mode
    DWORD dwFileOffsetHigh,         // high-order 32 bits of file offset
    DWORD dwFileOffsetLow,          // low-order 32 bits of file offset
10  DWORD dwNumberOfBytesToMap      // number of bytes to map
    );

LPVOID MapViewOfFileEx(// Prototype from Microsoft Visual C++ Help Documentation
    HANDLE hFileMappingObject,      // file-mapping object to map into address space
15  DWORD dwDesiredAccess,          // access mode
    DWORD dwFileOffsetHigh,         // high-order 32 bits of file offset
    DWORD dwFileOffsetLow,          // low-order 32 bits of file offset
    DWORD dwNumberOfBytesToMap,     // number of bytes to map
    LPVOID lpBaseAddress            // suggested starting address for mapped view
20  );

```

All of the foregoing functions directly use a pointer to a string that is a physical file. The only file functions that do not directly use a physical filename are functions like `CreateNamedPipe()`, which has the following parameters:

```

25  HANDLE CreateNamedPipe(// Prototype from Microsoft Visual C++ Help Documentation
    LPCTSTR lpName,                 // pointer to pipe name
    DWORD dwOpenMode,               // pipe open mode
    DWORD dwPipeMode,               // pipe-specific modes
    DWORD nMaxInstances,            // maximum number of instances
30  DWORD nOutBufferSize,           // output buffer size, in bytes
    DWORD nInBufferSize,           // input buffer size, in bytes
    DWORD nDefaultTimeOut,          // time-out time, in milliseconds
    LPSECURITY_ATTRIBUTES lpSecurityAttributes // pointer to security attributes structure
    );
35

```

The string to which CreateNamedPipe() points using the first parameter is a string that both an existing executable and the operating system know about and does not exist physically.

Unfortunately both of the executables that "know" this private name could only be loaded using one of the other procedures that required a physical file. Currently it is not possible to load an executable using a "named pipe" name. Both of or any executables that use the name of the "named pipe" already must have been loaded into memory.

All of the foregoing functions require a physical file because all of them use "file mapping" processes. File mapping allows large executable files to appear to be loaded rapidly since they are rarely completely loaded into memory but rather are mapped into memory. The detriment to this mapping capability is that executable code must remain in physical memory in a file in unencrypted form in order to be loaded, unless there is a middle layer or file system driver that the operating system uses as a physical layer and that decrypts the executable code to the operating system on demand. The potential weakness here is that another file system driver can hook into the operating system to monitor traffic between the operating system and all file system drivers and capture decrypted executable code passing from the file system driver to the operating system. Some operating systems allow such monitoring more than others. Many anti-viral software packages use this technique to prevent computer virus attacks.

One method of loading and executing encrypted executable computer program code is to use a stub executable having two parts. The first part is the normal front end loader code that all executables have. In addition, the first part would perform any authorization which may include receiving a password from the user, then allocate enough memory to hold hidden encrypted code when it is decrypted, either in its entirety or a portion of it, copy the encrypted code into that area of protected (and preferably locked so no disk swapping occurs) memory, decrypt it once it is in memory and only in memory, and then have the operating system load the code only from memory therefore bypassing any file system drivers or TSRs so they have access to only encrypted code.

Some of the file functions listed above and similar functions on other operating systems could be modified easily by a programmer having access to source code for those operating systems, or a new operating system may be made to provide functions which allow direct loading of executable code from memory rather than physical files. For example, in the Win32 commands, a command similar to CreateProcess() command could be provided. The command should have a few extra parameters including the process identifier of the process that contains

the now decrypted executable code, the memory address of the start of the decrypted code, and the size of the decrypted code. The command could also contain a parameter specifying a "call back" function within the first process that would provide decrypted code on demand directly to the operating system through a protected buffer, therefore allowing only a portion of the
5 encrypted code to be decrypted at any one time instead of in its entirety, for better protection and less memory use. The second parameter of the LoadLibraryEx() command that now needs to be NULL could be expanded to hold a structure that contained the same information. Both of these and other similar functions could be changed or created to allow loading executable code either as an .exe, .dll, or other extensions or identifiers, such as by using a "named pipe" name that only
10 the operating system and process that holds decrypted code know about and having the operating system load from the named pipe.

Alternatively, without having such additional capabilities in the operating system, an application program can be divided into two parts. The first part is code that is common to all applications such as code for allocating memory off the heap and code that provides some
15 interaction with the user. This kind of code is generally not code that the content provider is concerned about copying. The second part is the code that the content provider believes is valuable. Typically this valuable code is a business logic code or what would be considered a middle tier of a three-tier environment. A content provider would like to protect this second part of the code, at least much more than the first part of the code. The content provider would place
20 all of the important code to be protected inside a dynamic link library and the code that is not that important would reside in the front end "stub" executable. Both of these would be combined into another executable containing the .dll in encrypted form only, along with any other files, data, information, and/or tables for holding, for example, hardware identifiers. This other executable is the final digital information product.

25 The first part of the digital information product, i.e., the executable stub, would load and execute normally like any other application. It then would perform any authorization procedures. Once the proper authorization or password was completed successfully, an unwrap procedure would be performed as will now be described in connection with Fig. 8, it would then allocate enough protected memory using a function like VirtualAlloc() as shown in step 150:

30

```
DWORD nFileSize = 0;  
DWORD nPhantomFileSize = 0;
```

```

DWORD exeOffset = 0;
DWORD nPreferredLoadAddress = GetPreCompiledLoadAddress();
CString cCommandFile = UnwrapGetNTCommandFile();
exeOffset = UnwrapGetDllOffset(cCommandFile);
5  nFileSize = UnwrapGetDllSize(cCommandFile);
   nPhantomFileSize = nFileSize + 0x3000; // add any needed extra space
   // Increase buffer size to account for page size (currently Intel page size).
   DWORD nPageSize = GetPageSize();
   nPhantomFileSize += (nPageSize -(nPhantomFileSize % nPageSize));
10 // Allocate the memory to hold the decrypted executable.
   LPVOID lpvBlock = VirtualAlloc((LPVOID) nPreferredLoadAddress,
                                   nPhantomFileSize,
                                   MEM_RESERVE | MEM_COMMIT, PAGE_READWRITE);

15 This function can request a particular address space. Preferably, this address space is the
   preferred load address space to which the .dll was linked in order to minimize any needed
   relocation and fix up code. The stub executable may lock that area of memory in step 152, for
   example by using VirtualLock() to prevent any memory writes to a swap file, depending on the
   operating system, as shown below:

20  BOOL bVLock = VirtualLock((LPVOID) nPreferredLoadAddress, nPhantomFileSize);

   The memory area still should be secure even without this preventive step since the Windows 95
   and NT operating systems do not allow any user access to swap files.

25  The encrypted code is then copied from the digital information product into the allocated
   protected memory in step 154, for example by using the following command:

   UnwrapCopyHiddenExeToMem(cCommandFile, exeOffset, nFileSize, (char *) lpvBlock);

30  Once in memory, the stub would then decrypt the code to that same portion of memory in step
   156, for example by using the following commands:

```

```
CwrapDecryptSeed(cPassword.GetBuffer(0), cPassword.GetLength());  
CwrapDecrypt((unsigned char *) lpvBlock, 0, nFileSize);
```

Any "fix up and relocation" type services would then be performed in step 158, for example by
5 using the following command:

```
UnwrapFixUpAndRelocateDll(lpvBlock);
```

Possibly, the memory protection may be changed to execute only in step 160, for example by
10 using the VirtualProtect() command as follows:

```
DWORD lpfOldProtect; // variable to get old protection  
BOOL bVProtect = VirtualProtect((LPVOID) nPreferredLoadAddress,  
                                nPhantomFileSize,  
15                                PAGE_EXECUTE,  
                                &lpfOldProtect);
```

Function calls then can be made into that area of memory that now contains the decrypted code:

```
20 UnwrapDoDllAlgorithms();
```

Some of the "fix up" operations to be performed above include placing the addresses of external
or stub.exe functions into the address place holders of the decrypted .dll or internal code, by
using commands similar to the following:

```
25 WriteAddress((char*) 0x0a406104, (DWORD) &CallBackFunction1);  
WriteAddress((char*) 0x0a406100, (DWORD) &CallBackFunction2);
```

For instance a wrapper function could be created in the outer stub.exe that received a size
30 parameter, allocated that amount of memory off of the heap, and passed back the starting address
of that block of memory. Another example would be to have encrypted algorithms within the
hidden, encrypted .dll which would be called at run time from the front end stub once decrypted

within protected memory. The dynamic link library would be compiled and linked to expect a pointer to a function that took that parameter and/or returned a value by including prototypes in the header file as follows:

```
5 void (*lpCallBackFunc1)();
void (*lpCallBackFunc2)(unsigned long);
```

Function calls to "external" functions also could be added as follows:

```
10 (*lpCallBackFunc1)();
unsigned long z = x * x;
(*lpCallBackFunc2)(z);
```

At run time the "fix up" code would take the run time address of that "wrapper function" and place it into the pointer address within the .dll block of code as follows:

```
WriteAddress((char*) 0x0a406104, (DWORD) &CallBackFunction1);
WriteAddress((char*) 0x0a406100, (DWORD) &CallBackFunction2);
```

20 This information is readily available using the .cod output files from the compiler, an example of which follows:

```
_TestSum PROC NEAR                                ; COMDAT
; Line 8
25 00000 56          push  esi
; Line 23
00001 ff 15 00 00 00
      00          call  DWORD PTR _lpCallBackFunc1
; Line 24
30 00007 8b 44 24 08  mov  eax, DWORD PTR _a$[esp]
0000b 50          push  eax
0000c e8 00 00 00 00 call  _TestSquare
```



```

00011      83 c4 04      add    esp, 4
00014      8b f0          mov    esi, eax
; Line 25
00016      8b 44 24 0c    mov    eax, DWORD PTR_b$(esp)
5  0001a      50            push   eax
0001b      e8 00 00 00 00 call   _TestSquare
00020      83 c4 04      add    esp, 4
00023      03 c6          add    eax, esi
; Line 28
10 00025      5e            pop    esi
00026      c3            ret    0
_TestSum ENDP
_TEXT      ENDS
;      COMDAT_TestSquare
15 _TEXT      SEGMENT
_x$ = 8
_TestSquare PROC NEAR                                ; COMDAT
; Line 30
00000      56            push   esi
20 ; Line 32
00001      8b 74 24 08    mov    esi, DWORD PTR_x$(esp)
00005      0f af f6      imul  esi, esi
; Line 34
00008      56            push   esi
25 00009      ff 15 00 00 00
          00            call   DWORD PTR_lpCallBackFunc2
0000f 83 c4 04      add    esp, 4
00012      8b c6          mov    eax, esi
; Line 36
30 00014      5e            pop    esi
00015      c3            ret    0
_TestSquare ENDP

```

Such information also is available from .map output files from the linker where the "f" between the address (i.e., 0a406100) and the object file (i.e. Algorithms.obj) means it is a "flat" address (i.e., hard coded by the linker) and the lack of an "f" means that it is an address pointer to be supplied at run time (load time) where the address that is contained in that address location is used and not the actual address location (i.e., the address that is contained at address location 5 0a406100 and not 0a406100 itself):

```

0001:00000000   _TestSum           0a401000 f Algorithms.obj
0001:00000030   _TestSquare        0a401030 f Algorithms.obj
10
0003:00001100   _lpCallbackFunc2   0a406100 Algorithms.obj
0003:00001104   _lpCallbackFunc1   0a406104 Algorithms.obj

```

When the code inside the .dll makes a "call" to a dereferenced pointer, it would jump to the correct function in the outer code and return the expected return value (if any). For example: 15

```

void CallbackFunction1(){
// This is the first function that exists in the Stub executable
// whose address has been placed at the appropriate location inside the "dll" code
20 // that has now been decrypted in a block of memory. The code inside the "dll"
// makes a function call to this function. In its encrypted state, the "dll" does not contain
// this address, but merely has a placeholder for the address. The "dll" has enough space
allocated to hold an
// address of this size. After the "dll" has been decrypted at run time, its address is
25 // placed in that location so the code inside the "dll" that references (or more
// appropriately dereferences) that address can jump (which is function call) to this
// address.
AfxMessageBox(
_T("This is the FIRST Stub.exe call back function being called from the dll.));
30     return;
}

```

- 25 -

```
void CallBackFunction2(DWORD nNumber){
// See comment for CallBackFunction1 except this function receives a parameter off
// of the stack. It could also return a value as well.
    CString
5    cString(
T("This is the SECOND Stub.exe call back function being called from the dll"));

    har buffer[20];
    ltoa(nNumber, buffer, 10);
10
    cString += _T(" with a parameter of ");
    cString += buffer;
    cString += _T(".");
    AfxMessageBox(cString.GetBuffer(0));
15    return;
}
```

The outer stub.exe would make the same kinds of jumps or function calls into the now protected decrypted code block as follows:

```
20    DWORD c;

// This command declares a function pointer. This command is different for different function
// calls. Here the called function takes two integer parameters and
25 // passes back a DWORD.
    DWORD (*lpFunc)(DWORD,DWORD);

// The function pointer is then pointed to the starting address of the function in the
// block of memory that now holds the decrypted DLL.
30 lpFunc = (DWORD (*)(DWORD,DWORD)) UnwrapFixUpAndRelocateDll();

// Now call that "function" which is really like all function calls, i.e., a jump to
```

// the address where that function exists. In this case, two
 // variables are passed to that function and returning a value from that function. This function
 illustrates that the function call
 // can be more complicated than merely a simple jump
 5 // to an address. Inline assembler code may be used to push the variables onto
 // the stack frame and return the variable from the eax register, but this function enables
 // the C++ compiler to do the same function.
 c = (DWORD) (*lpFunc)(a, b);

10 This mechanism requires the unwrap procedure and the now decrypted code to have intimate
 knowledge about procedural interfaces of each other but no knowledge about each other's
 implementation. This is the way most executable .exe files and .dll files behave but with the
 addition of a series of "wrapper" functions on either side for communication. This method works
 under Windows 95 and Windows NT 4.0 operating systems and should work under Windows NT
 15 3.51 and other operating systems.

Another modified version of this mechanism that works under the Windows NT 4.0
 operating system because of functions specific to Windows NT 4.0 would be to have another
 hidden and/or encrypted executable within the digital information product. This executable
 would be copied to a physical disk in an unencrypted form, launched or loaded with the
 20 CreateProcess() command in its current form but called with a parameter to load the executable
 in suspended mode:

```

  BOOL success = CreateProcess(cFrontEndExe.GetBuffer(0), 0, 0, 0, TRUE,
    CREATE_NEW_CONSOLE | CREATE_SUSPENDED,
  25    0, 0, &startUpInfo, &processInfo);

```

Then the first process would copy the encrypted dll into its own process and decrypt it, allocate
 enough memory using VirtualAllocEx() in its current form in the second process that has just
 loaded the expendable front end executable in a suspended state as follows:

```

  30 LPVOID lpvBlockEx = VirtualAllocEx(processInfo.hProcess,

```

- 27 -

(LPVOID) nPreferredLoadAddress, nPhantomFileSize,
MEM_RESERVE | MEM_COMMIT,
PAGE_READWRITE);

- 5 The decrypted code is copied from the first process to the second suspended process using WriteProcessMemory() in its current form:

```

10  BOOL bWriteProcessMemory = WriteProcessMemory((HANDLE) processInfo.hProcess,
        (LPVOID) lpvBlockEx, (LPVOID) nPreferredAddress,
        (DWORD) nPhantomFileSize, (LPDWORD) &nBytesWritten);

```

The primary thread of the previously launched second process is then resumed:

```

15  DWORD nResumed = ResumeThread(processInfo.hThread);

```

Any necessary function pointers are then placed in the correct locations by the second process, the area of memory is locked to prevent any writes to a swap file, and the memory protection is changed to execute only as follows:

```

20  WriteAddress((char*) 0x0a406104, (DWORD) &CallBackFunction1);
    WriteAddress((char*) 0x0a406100, (DWORD) &CallBackFunction2);

    BOOL bVLock = VirtualLock((LPVOID) nPreferredLoadAddress, nPhantomFileSize);
    DWORD lpflOldProtect; // variable to get old protection
25  BOOL bVProtect = VirtualProtect((LPVOID) nPreferredLoadAddress,
        nPhantomFileSize, PAGE_EXECUTE, &lpflOldProtect);

```

- 30 The program can continue running by making and receiving calls to and from the decrypted dynamic link library that now resides in the protected memory of its process using commands such as the following:

```

    DWORD c;

```

```
DWORD (*lpFunc)(DWORD,DWORD);  
lpFunc = (DWORD (*)(DWORD,DWORD)) ExpendableGetEntryAddress();  
c = (DWORD) (*lpFunc)(a, b);
```

- 5 The first process can either close down or launch another instance of that same process.

In either of these implementations using the same process or launching into a second process, the hidden encrypted code never passes through a file system driver or memory resident program in decrypted form. Code can be split up among different dynamic link libraries so that no two would reside in memory at the same time in order to protect code further. Both of these systems can be implemented using the Win32 function calls. If additional functions, similar to a
10 CreateProcess() command or a LoadLibrary() command but that take a process identifier and address location in memory to load in an executable instead of a physical file, are provided in an operating system then the entire executable and dynamic link library can be hidden, encrypted, and protected on the physical disk and then decrypted within protected memory and use the
15 operating system loader to load it directly to the operating system from memory without residing in decrypted form on any physical medium.

Having described the operation and use of the computer program product in accordance with the invention, embodiments of which are described above in connection with Figs. 3-8, and the operation of the unwrap procedure and device driver it contains, the process of constructing
20 such a computer program product will now be described in more detail. Referring now to Fig. 9, an embodiment of this process for creating a computer program product is shown. This process can be applied to any digital information including an arbitrary executable computer program, dynamic link libraries and related files of data. All digital information is treated as mere data by this process. Each separate data file is combined into a single file by this process, with an
25 executable program for performing the unwrap procedure, and optionally executable program code for a virtual device driver, into the computer program product. Each file of hidden information has a unique location and is identified by its own begin and end markers as shown in Fig. 3. The first step of this process is opening a new data file for the computer program using a name that will be used to indicate an executable file (step 200). For example, an executable
30 word processing program may be named "word_processor.exe" in the Windows95 operating system.

The three portions of the computer program product are then inserted into the open data file. First, the unwrap procedure is inserted at the beginning of the file in an executable format in step 202. The begin tag for the optional device driver is then inserted in step 204. The executable device driver program code is then inserted in step 206, followed by its corresponding end tag in step 208. For each hidden file to be inserted into this computer program product, steps 210 to 216 are performed. First, the begin tag is inserted in step 210. The begin tag also may include an indication of a unique name of the file which will be used as its name in the phantom directory created by the unwrap procedure. The hidden file is then encrypted and/or compressed in step 212 and inserted into the data file in step 214. The end tag for the hidden file is then inserted in step 216. The device driver and all of the tags may be encrypted also if the unwrap procedure has suitable decryption procedures. The computer program file is closed when the last hidden file is processed.

Using the present invention digital information, such as executable program code or various kinds of data, is loaded and unloaded as needed, and thus does not take up any more memory than is necessary. At no time does unencrypted digital information, such as computer program code, exist on disk in accessible and complete decrypted form. Because the original digital information is available as a read only file in one embodiment of the invention accessible only to the device driver, the digital information may be accessed over networks, from a CD-ROM or from a DVD, and can be made to have a limited number of uses. This mechanism is particularly useful for controlling distribution of computer programs, digitized movies or other information while reducing the cost of such distribution and control. For example, software may be distributed over a network on a single use basis, and charges may be levied on a per use basis. The ability to reverse engineer an application program also may be reduced.

One benefit with this system over some other systems for preventing unauthorized access to digital information is that the content provider maintains control of the encryption applied to the information how it may be decrypted. Any need for either a centralized facility or a predetermined decryption program is eliminated. An operating systems manufacturer or other platform vendor merely provides the capability for the information to be accessed and decrypted on the fly. Since the valuable information and any other tables of authorization codes, passwords, or hardware identifiers that the content provider may use to secure the information resides in one large encrypted file, it becomes difficult, if not impossible, for someone to determine just where any of this information exists.

A potential scenario with authorization procedure in which the present invention may be used is the following. A consumer purchases a DVD disk containing a movie. The user puts the disk into the player. This is the first time the disk is installed. The content provider's functions are loaded into the DVD chip, which looks in the encrypted table and sees that this is the first
5 time this disk is being played. The player then displays on a screen a numeric identifier and toll free phone number. The consumer calls the toll free phone number and inputs the numeric identifier that was displayed on the screen. The content provider provides a numeric password based on the numeric identifier that the user inputs into the DVD. The content provider may develop a database of information about its consumers that also may be used to detect pirating of
10 the digital information product. Now that this authorization has taken place, the software that the content provider wrote, and is now in the DVD chip, takes a hardware identifier from the DVD and encrypts it and puts it in the encrypted and buried table on the disk. Alternatively, the data may be decrypted in memory and re-encrypted back onto the disk using the hardware identifier as part of a key. Now that disk will run and show the movie and will only run on that DVD and
15 no other. The content provider could allow for a table of hardware id's so they could limit the number of DVD's that disk would run on or a limited number of times it can be shown. It should be understood that many other authorization procedures may be used.

In the foregoing scenario, the movie is encrypted on the same disk inside of the encrypted file that contains the table and functions the content provider distributed. The movie is decrypted
20 by the decryption functions contained in the file directly to the DVD chip. At no time does the movie reside anywhere in decrypted form. The content provider can protect the movie with any desired level of security (for both encryption and authorization).

In the present invention, the onus of protection of content does not reside with a hardware manufacturer or platform provider but in the hands of the content provider. The hardware
25 manufacturer only provides the mechanism to protect the digital information through the operating system. The technique and implementation of protection resides in the hands of the content provider. This mechanism allows the content providers to change the level of security as needed without any modifications to the hardware. The security of the content is provided by the encryption/decryption algorithms, public/private keys, and authorization methods which are
30 determined by the content provider. Even each individual product can have its own encryption/decryption algorithms and/or public/private keys. All of these can be changed and enhanced as the market demands.

The present invention also could be used for on-line or live use of digital information. For example, a movie could be retrieved on demand and recorded by a consumer. A set top box could receive the digital information, decrypt it, and then re-encrypt and store the information using, for example, a hardware identifier of the set top box. Since home movies digitally recorded would be encrypted using the hardware identifier of the device used in recording, that home movie could not be played on another or only on a limited number of other devices and/or for only a specified number of times depending on the wishes of the content provider. Since the algorithms are downloaded at the time of recording from a service provider, e.g., the cable company, the content provider (movie company) would provide the encrypted data to the service provider to present to their customers. The service provider need not be concerned with the encryption/decryption and authorization functions used by the content provider. Similar uses are possible with other data transmission systems including, but not limited to, telephone, cellular communications, audio transmission including communication and the like.

In another embodiment, the stub executable program is a first process that is implemented similar to a debugging tool such as the SoftIce debugger from NuMega Technologies or the WinDebug debugger from Microsoft Corporation for Ring 0 kernel level debugging for an Intel processor based architecture, or the CodeView debugger for ring 3 application level debugging. Such a debugger controls execution of a program to be debugged as a second process and steps through each program statement or opcode of the debugged program. The debugging tool could be modified to monitor each opcode that indicates a jump to a program fragment, such as each instruction or a block code. If the program fragment to be executed is not decrypted, the modified debugger decrypts the program fragment before the jump command is allowed to execute. Each program fragment may be re-encrypted after execution. Clearly, unnecessary debugging commands may be omitted from the modified debugger.

Having now described a few embodiments of the invention, it should be apparent to those skilled in the art that the foregoing is merely illustrative and not limiting, having been presented by way of example only. Numerous modifications and other embodiments are within the scope of one of ordinary skill in the art and are contemplated as falling within the scope of the invention as defined by the appended claims and equivalent thereto.

CLAIMS

1. A computer-implemented process for executing encrypted computer program logic while maintaining protection against copying of corresponding decrypted executable computer program logic, wherein the encrypted computer program logic is stored in association with first
5 executable computer program logic, the process comprising the steps of:
- through an operating system of a computer, reading, loading and executing the first executable computer program logic as a first process having a protected memory area defined by the operating system;
 - the first process decrypting the encrypted computer program logic into second executable
10 computer program logic and storing the second executable computer program logic in the protected memory area; and
 - the first process causing loading and execution of the decrypted second computer program logic in the protected memory area.
- 15 2. The process of claim 1, wherein the encrypted computer program logic and the first executable computer program logic are stored in a single data file accessible through the operating system.
3. The process of claim 1, wherein the execution of the decrypted second computer program
20 logic is performed as a second process having a second protected memory area defined by the operating system.
4. A digital information product including a computer readable medium having digital information stored thereon, the digital information including computer program logic defining
25 first executable computer program logic, wherein the first executable computer program logic when executed performs the following steps:
- storing the encrypted computer program logic in a data file accessible through an operating system of a computer, wherein the data file also includes first executable computer program logic;
 - 30 through the operating system, reading, loading and executing the first executable computer program logic from the data file as a first process having a protected memory area;

the first process decrypting the encrypted computer program logic into second executable computer program logic and storing the second executable computer program logic in the protected memory area; and

the first process causing loading and execution of the decrypted second computer
5 program logic in the protected memory area.

5. A computer system comprising:

a processor for executing computer program logic;

a main memory operatively connected to the processor for storing digital information
10 including executable computer program logic at memory locations addressed by the processor;
and

an operating system defined by executable computer program logic stored in the memory
and executed by the processor and having a command which when executed by the processor
defines means for creating a process in response to a request specifying a process identifier and a
15 memory location in the main memory, wherein the process identifier indicates the process
making the request and the memory location stores executable computer program logic which
when executed defines the process.

6. A computer system having an operating system, for decrypting digital information,
20 comprising:

means for storing the encrypted computer program logic in a data file accessible through
the operating system, wherein the data file also includes first executable computer program logic;

means, invokable through the operating system, for reading, loading and executing the
first executable computer program logic from the data file as a first process having a protected
25 memory area;

the first process defining means for decrypting the encrypted computer program logic
into second executable computer program logic and storing the second executable computer
program logic in the protected memory area; and

the first process defining means for causing loading and execution of the decrypted
30 second computer program logic in the protected memory area.

7. The computer system of claim 6, wherein the encrypted computer program logic and the first executable computer program logic are stored in a single data file accessible through the operating system.
- 5 8. The computer system of claim 6, wherein the execution of the decrypted second computer program logic is performed as a second process having a second protected memory area defined by the operating system.
9. A digital information product, including a computer readable medium with computer readable
10 information stored thereon, wherein the computer readable information comprises:
a first portion of executable computer program logic; and
a second portion of encrypted digital information; and
wherein the first portion of executable program logic, when executed, defines means,
operative in response to requests for digital information, for accessing the second portion of
15 encrypted digital information, for decrypting the encrypted digital information, and for
outputting the decrypted digital information.
10. The digital information product of claim 9, wherein the encrypted digital information is
encrypted executable computer program logic.
- 20 11. A computer program product including a self-decrypting encrypted executable computer
program, comprising:
a computer readable medium having computer program logic stored thereon, wherein the
computer program logic defines:
25 a first module,
a second module,
wherein the first module, when executed by a computer, defines means for loading the
second module into memory of the computer, and
a third module defining the encrypted executable computer program,
30 wherein the second module, when executed by a computer, defines means for
communicating with an operating system of the computer to receive requests for program code
from the encrypted executable computer program from the third module, and for processing the

requests to access and decrypt the encrypted executable computer program and for providing the decrypted executable code from the third module to the operating system.

12. A process for executing encrypted executable computer programs on a computer system
5 having a processor, memory and operating system, comprising the steps of:
- receiving computer program logic having a first module defining a start up routine, a
second module, and a third module containing the encrypted executable computer program;
 - executing the first module of the received computer program logic using the processor,
wherein the step of executing causes the second module to be loaded into the memory of
10 the computer system, and
 - generating requests from the operating system for data from the encrypted executable
computer program which are received by the second module, and
 - accessing and decrypting the encrypted executable computer program and returning the
decrypted executable computer program to the operating system.

15

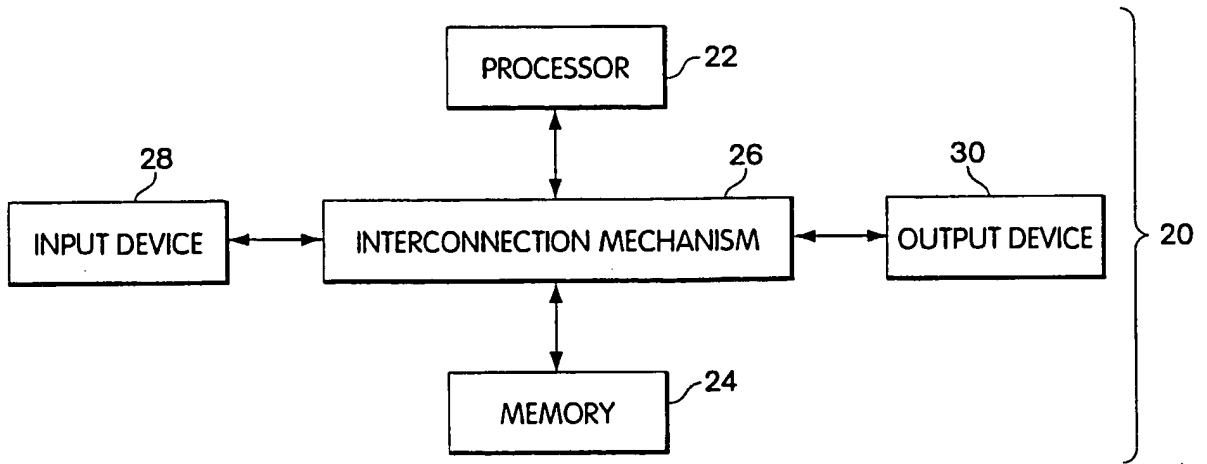


Fig. 1

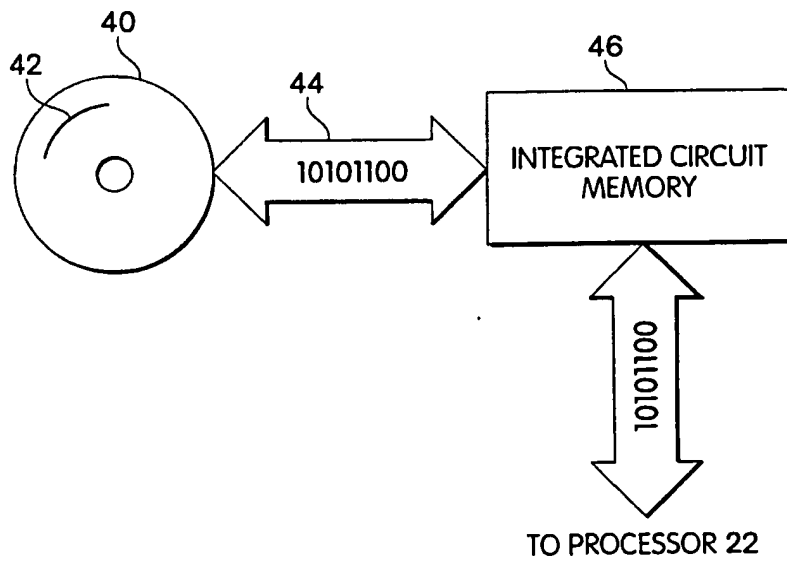


Fig. 2

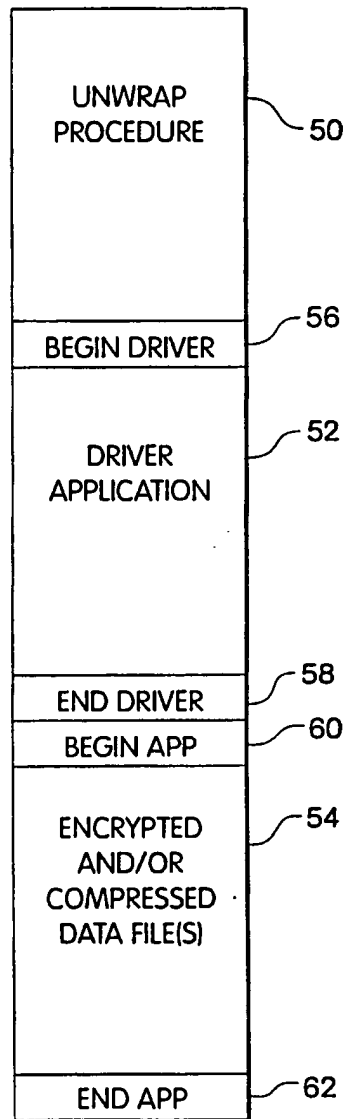


Fig. 3

3/8

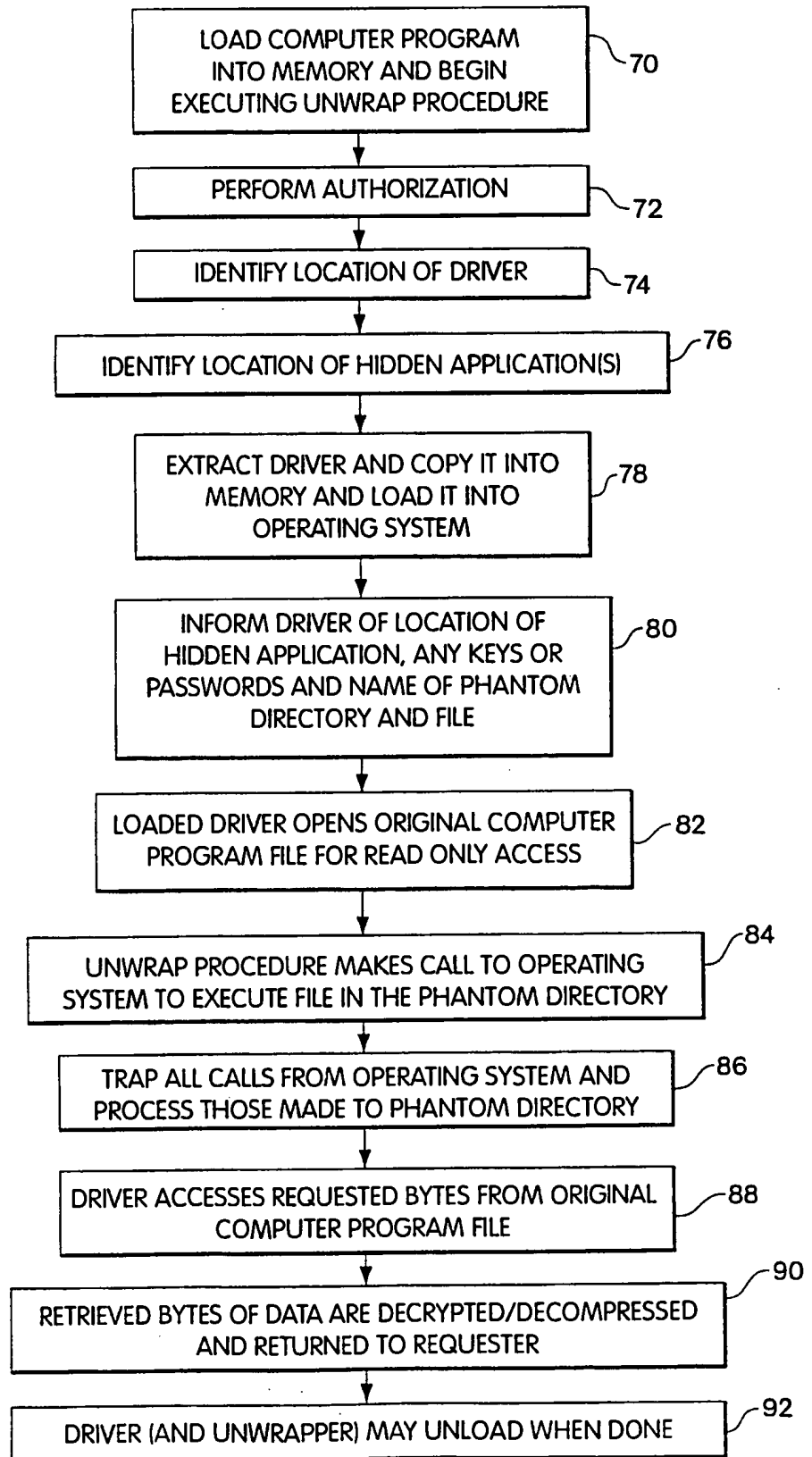


Fig.4

SUBSTITUTE SHEET (RULE 26)

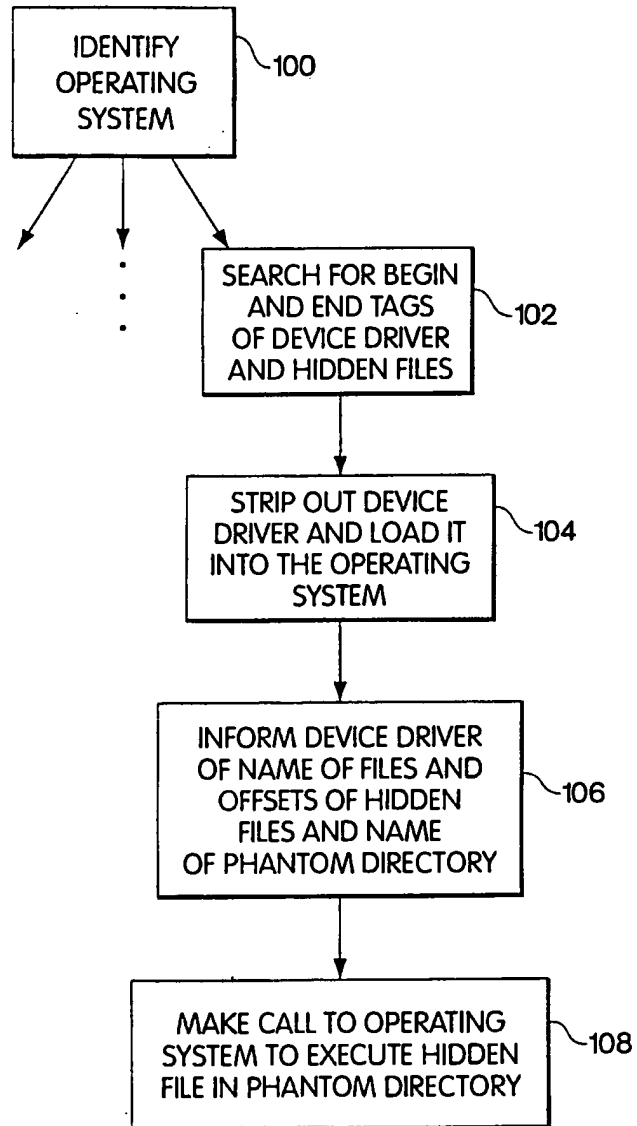


Fig. 5

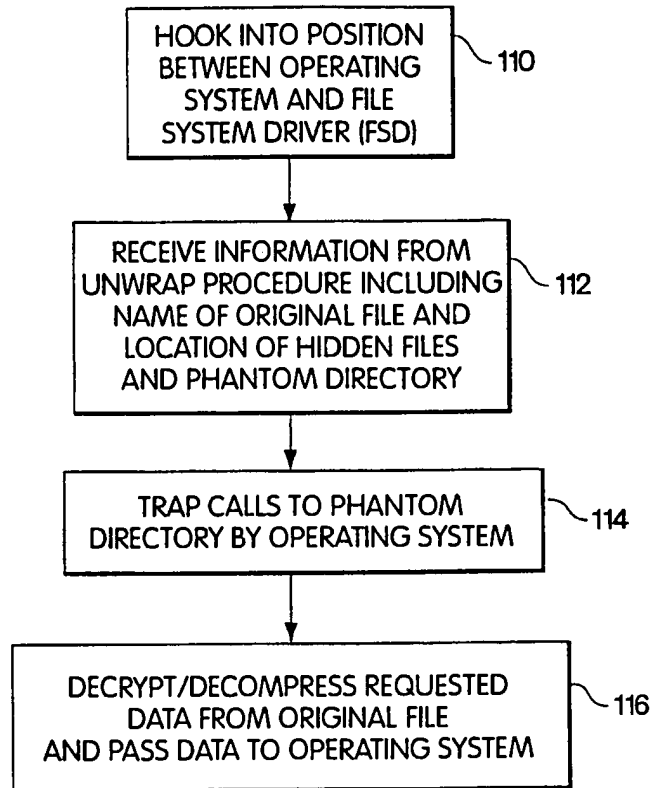


Fig. 6

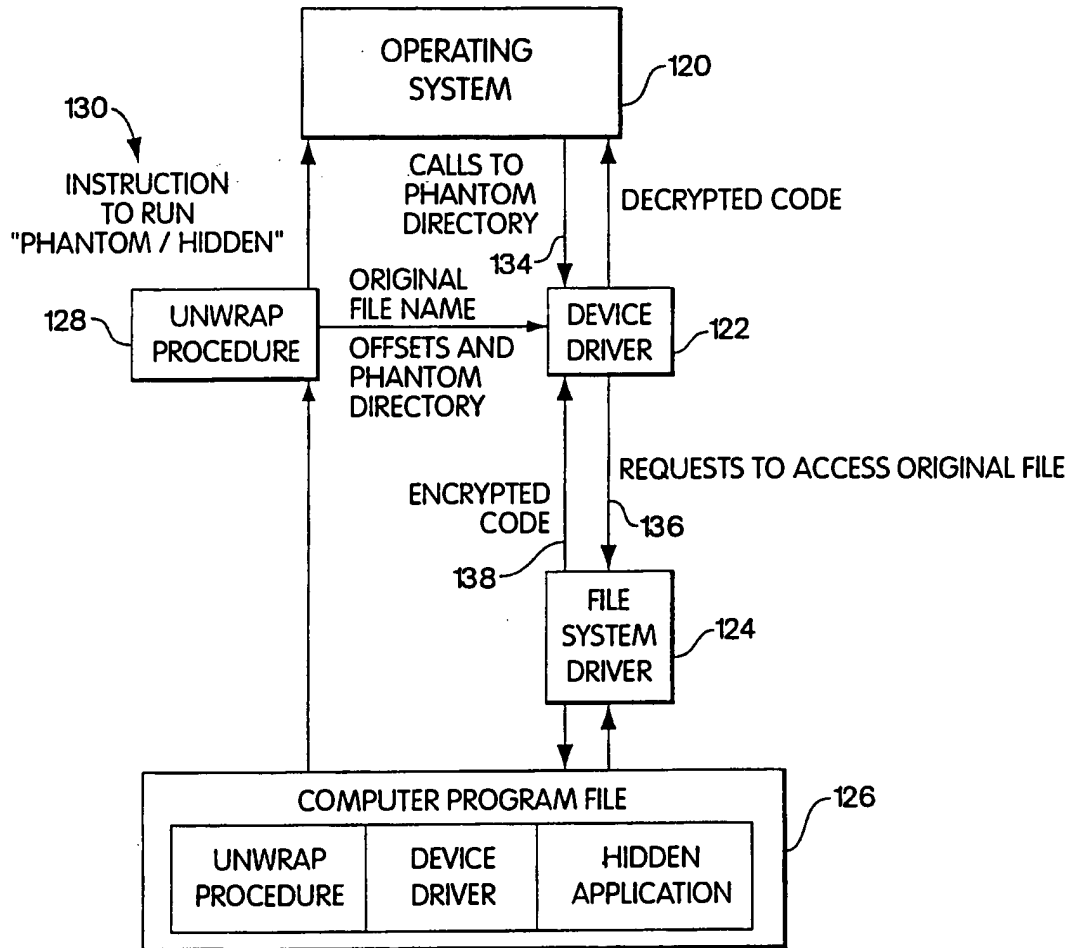


Fig. 7

7/8

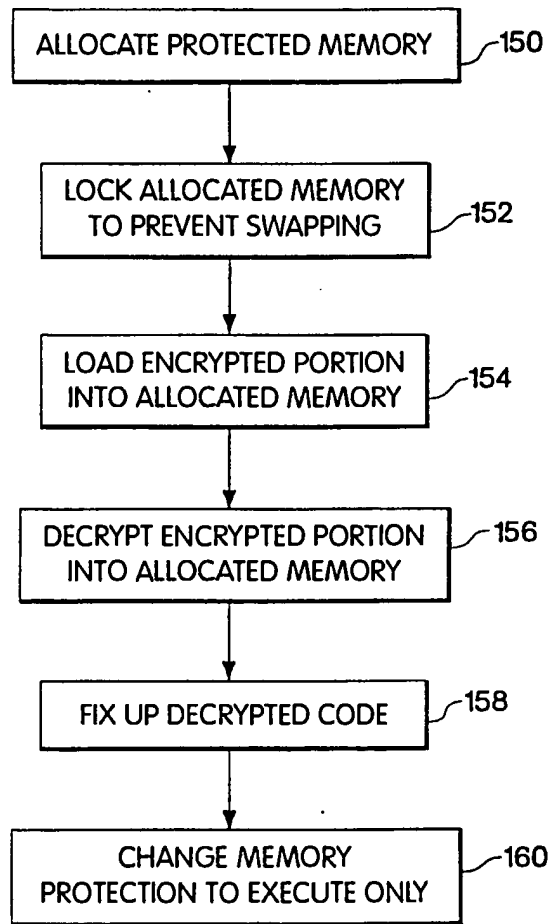


Fig. 8

8/8

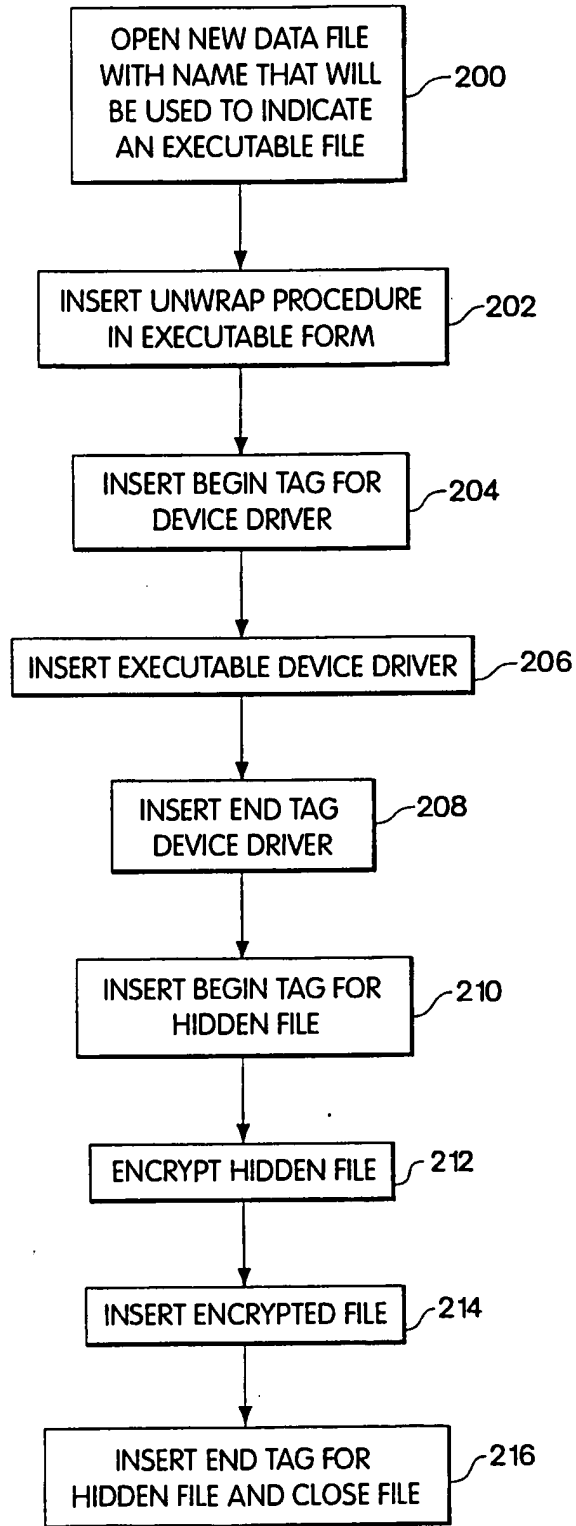


Fig. 9

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US97/16223

A. CLASSIFICATION OF SUBJECT MATTER
 IPC(6) :H04L 9/00
 US CL : 380/4
 According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED
 Minimum documentation searched (classification system followed by classification symbols)
 U.S. : 380/4,9,23,25,49,50,59

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 4,937,861 A (CUMMINS) 26 June 1990, see Abstract.	1-12
A	US 5,007,082 A (CUMMINS) 09 April 1991, see Abstract.	1-12
A	US 5,144,659 A (JONES) 01 September 1992, see Abstract.	1-12
A	US 5,155,827 A (GHERING) 13 October 1992, see Abstract.	1-12
A	US 5,396,609 A (SCHMIDT et al) 07 March 1995, see Abstract.	1-12

Further documents are listed in the continuation of Box C. See patent family annex.

* Special categories of cited documents:	*T later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
*A document defining the general state of the art which is not considered to be of particular relevance	*X document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
*B earlier document published on or after the international filing date	*Y document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
*L document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	*A document member of the same patent family
*O document referring to an oral disclosure, use, exhibition or other means	
*P document published prior to the international filing date but later than the priority date claimed	

Date of the actual completion of the international search 20 JANUARY 1998	Date of mailing of the international search report 18 FEB 1998
--	---

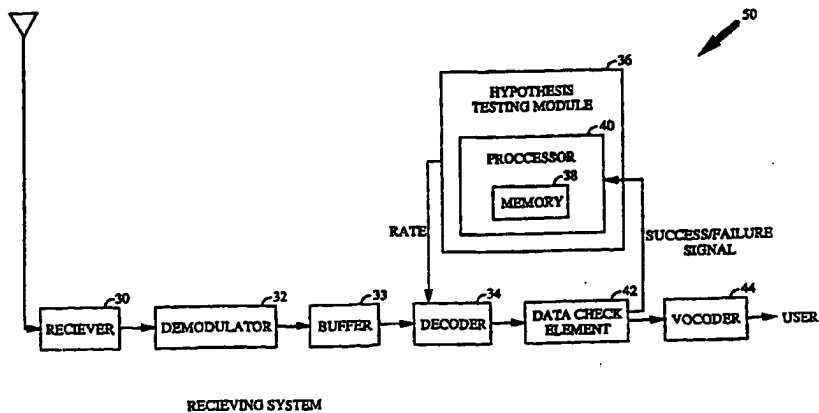
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. (703) 305-3230	Authorized officer <i>Diane Goodenough</i> BERNARR EARL GREGORY Telephone No. (703) 306-4153
---	--



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

<p>(51) International Patent Classification ⁶ : H04L 25/02</p>	<p>A1</p>	<p>(11) International Publication Number: WO 98/19431 (43) International Publication Date: 7 May 1998 (07.05.98)</p>
<p>(21) International Application Number: PCT/US97/19676 (22) International Filing Date: 27 October 1997 (27.10.97) (30) Priority Data: 08/741,273 30 October 1996 (30.10.96) US (71) Applicant: QUALCOMM INCORPORATED [US/US]; 6455 Lusk Boulevard, San Diego, CA 92121 (US). (72) Inventors: TIEDEMANN, Edward, G., Jr.; 4350 Bromfield Avenue, San Diego, CA 92122 (US). LIN, Yu-Chuan; 585 W. 63rd Avenue, Vancouver, British Columbia V6P 2G7 (CA). (74) Agents: OGROD, Gregory, D. et al.; Qualcomm Incorporated, 6455 Lusk Boulevard, San Diego, CA 92121 (US).</p>		<p>(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, UZ, VN, YU, ZW, ARIPO patent (GH, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).</p> <p>Published <i>With international search report. Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.</i></p>

(54) Title: METHOD AND APPARATUS FOR DECODING VARIABLE RATE DATA



(57) Abstract

A system and method for determining the data rate of a frame of data at a receiver (50) of a variable rate communications system. A vocoder at a transmitter encodes a frame of data at one of the rates of a predetermined set of rates. The data rate is dependent on the speech activity during the time frame of the data. The data frame is also formatted with overhead bits, including bits for error detection and detection. At the receiver (50), the data rate for the frame is determined based on hypothesis testing. Because the data rate is based on speech activity, a hypothesis test may be designed based on the statistics of speech activity. The received data frame is first decoded by a decoder (34) into information bits at the most probable rate as provided by the hypothesis testing module (36). Data check element (42) generates error metrics for the decoded information bits. If the error metrics indicate that the information bits are of good quality, then the information bits are presented to a vocoder (44) at the receiver to be processed for interface with the user. If the error metrics indicate that the information bits have not been properly decoded, then decoder (34) decodes the received data frame at the other rates of the set of rates until the actual data rate is determined.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

METHOD AND APPARATUS FOR DECODING VARIABLE RATE DATA

BACKGROUND OF THE INVENTION

5

I. Field of the Invention

The present invention relates to digital communications. More particularly, the present invention relates to a novel and improved system and method for determining, at a receiver of a variable rate communication system, the rate at which data has been encoded for transmission.

II. Description of the Related Art

15 The use of code division multiple access (CDMA) modulation techniques is one of several techniques for facilitating communications in which a large number of system users are present. Although other techniques such as time division multiple access (TDMA), frequency division multiple access (FDMA), and AM modulation schemes such as
20 amplitude companded single sideband (ACSSB) are known, CDMA has significant advantages over these other techniques. The use of CDMA techniques in a multiple access communication system is disclosed in U.S. Pat. No. 4,901,307, entitled "SPREAD SPECTRUM MULTIPLE ACCESS COMMUNICATION SYSTEM USING SATELLITE OR TERRESTRIAL
25 REPEATERS," assigned to the assignee of the present invention and incorporated by reference herein.

CDMA systems often employ a variable rate vocoder to encode data so that the data rate can be varied from one data frame to another. An exemplary embodiment of a variable rate vocoder is described in U.S. Pat.
30 No. 5,414,796, entitled "VARIABLE RATE VOCODER," assigned to the assignee of the present invention and incorporated by reference herein. The use of a variable rate communications channel reduces mutual interference by eliminating unnecessary transmissions when there is no useful speech to be transmitted. Algorithms are utilized within the vocoder for generating a
35 varying number of information bits in each frame in accordance with variations in speech activity. For example, a vocoder with a set of four rates may produce 20 millisecond data frames containing 16, 40, 80, or 171 information bits, depending on the activity of the speaker. It is desired to transmit each data frame in a fixed amount of time by varying the
40 transmission rate of communications.

SUBSTITUTE SHEET (RULE 26)

Additional details on the formatting of the vocoder data into data frames are described in U.S. Pat. No. 5,511,073, entitled "METHOD AND APPARATUS FOR THE FORMATTING OF DATA FOR TRANSMISSION," assigned to the assignee of the present invention and herein incorporated by
5 reference. The data frames may be further processed, spread spectrum modulated, and transmitted as described in U.S. Pat. No. 5,103,459, entitled "SYSTEM AND METHOD FOR GENERATING WAVEFORMS IN A CDMA CELLULAR TELEPHONE SYSTEM," assigned to the assignee of the present invention and incorporated by reference herein.

10 Variable rate systems can be developed which include explicit rate information. If the rate is included as part of a variable rate frame, then the rate is not recoverable until after the frame has already been properly decoded, at which point the rate has already been determined. Rather than including the rate in a variable rate frame, the rate could instead be sent in a
15 non-variable rate portion of the frame. However, only a few bits are typically needed to represent the rate, and these bits cannot be efficiently encoded and interleaved in order to provide error protection for fading communications channels. Furthermore, the rate information is only available after some decoding delay and are subject to error.

20 Alternatively, variable rate systems can be developed which do not include explicit rate information. One technique for the receiver to determine the rate of a received data frame where the rate information is not explicitly included in the frame is described in copending U.S. Patent Application Serial No. 08/233,570, entitled "METHOD AND APPARATUS
25 FOR DETERMINING DATA RATE OF TRANSMITTED VARIABLE RATE DATA IN A COMMUNICATIONS RECEIVER," filed April 26, 1994, assigned to the assignee of the present invention, and incorporated by reference. Another technique is described in copending U.S. Patent Application Serial No. 08/126,477, entitled "MULTIRATE SERIAL VITERBI
30 DECODER FOR CODE DIVISION MULTIPLE ACCESS SYSTEM APPLICATIONS," filed Sept. 24, 1993, assigned to the assignee of the present invention, and incorporated by reference. According to these techniques, each received data frame is decoded at each of the possible rates. Error metrics, describing the quality of the decoded symbols for each frame
35 decoded at each rate, are provided to a processor. The error metrics may include Cyclic Redundancy Check (CRC) results, Yamamoto Quality Metrics, and Symbol Error Rates. These error metrics are well-known in communications systems. The processor analyzes the error metrics and

determines the most probable rate at which the incoming symbols were transmitted.

Decoding each received data frame at each possible data rate will eventually generate the desired decoded data. However, the search through
5 all possible rates is not the most efficient use of processing resources in a receiver. Also, as higher transmission rates are used, power consumption for determining the transmission rate also increases. This is because there are more bits per frame to be processed. Furthermore, as technology evolves, variable rate systems may utilize larger sets of data rates for
10 communicating information. The use of larger sets of rates will make the exhaustive decoding at all possible rates infeasible. In addition, the decoding delay will not be tolerable for some system applications. Consequently, a more efficient rate determination system is needed in a variable rate communications environment. These problems and deficiencies are clearly
15 felt in the art and are solved by the present invention in the manner described below.

SUMMARY OF THE INVENTION

20 The present invention is a novel and improved system and method for determining the transmission rate of communications in a variable rate communications system. In a variable rate system, the data rate at which a data frame is encoded may be based on the speech activity during the time frame. Because the characteristics of speech are known, probability
25 functions may be defined for the data rates which are dependent on the characteristics of speech. The probability functions may in addition be dependent on the measured statistics of the received data frames. Furthermore, hypothesis tests can be designed based on the probability functions to determine the most likely data rate of a received frame of data.
30 These probability functions may be dependent on the selected service option. For example, the probability functions for data services will be different than for voice services.

At the receiver of the present invention, a processor causes a decoder to decode the received frame of data into information bits at the most
35 probable rate as determined by the hypothesis test. The most probable rate may, for example, be the rate of the previous frame of data. The decoder also generates error metrics for the decoded information bits. The decoded bits and the error metrics are provided to a data check element which checks the decoded bits for correctness. If the error metrics indicate that the

decoded information bits are of good quality, then the information bits are provided to a vocoder which further processes the data and provides speech to the user. Otherwise, a failure signal is presented to the processor. The processor then causes the decoder to decode the received frame of data at
5 other data rates until the correct data rate is found.

BRIEF DESCRIPTION OF THE DRAWINGS

The features, objects, and advantages of the present invention will
10 become more apparent from the detailed description set forth below when taken in conjunction with the drawings in which like reference characters identify correspondingly throughout and wherein:

FIG. 1 is a schematic overview of an exemplary CDMA cellular telephone system;

15 FIG. 2 is a block diagram of a variable rate receiving system with particular reference to the rate determination features of the present invention;

FIGS. 3 and 4 are flow charts illustrating two embodiments of the processing steps involved in rate determination wherein the hypothesis test
20 designates the rate of the previous frame of data as the most probable rate for the current frame of data;

FIGS. 5 and 6 are flow charts illustrating two embodiments of the processing steps involved in rate determination wherein the hypothesis test is based on the a priori probability distribution of the data rates; and

25 FIGS. 7 and 8 are flow charts illustrating two embodiments of the processing steps involved in rate determination wherein the hypothesis test is based on the conditional probability distribution of the data rates.

DETAILED DESCRIPTION OF THE PREFERRED 30 EMBODIMENTS

An exemplary cellular mobile telephone system in which the present invention is embodied is illustrated in FIG. 1. For purposes of example this system is described herein within the context of a CDMA cellular
35 communications system. However, it should be understood that the invention is applicable to other types of communication systems such as personal communication systems (PCS), wireless local loop, private branch exchange (PBX) or other known systems. Furthermore systems utilizing other well known transmission modulation schemes such as TDMA and

FDMA as well as other spread spectrum systems may employ the present invention.

An exemplary cellular system in which the rate determination system of the present invention may be implemented is illustrated in FIG. 1. In FIG. 1, system controller and switch 10 typically include appropriate interface and processing hardware for providing system control information to the cell-sites. Controller 10 controls the routing of telephone calls from the public switched telephone network (PSTN) to the appropriate cell-site for transmission to the appropriate mobile unit. Controller 10 also controls the routing of calls from the mobile units via at least one cell-site to the PSTN. Controller 10 may direct calls between mobile users via the appropriate cell-site stations since such mobile units do not typically communicate directly with one another.

Controller 10 may be coupled to the cell-sites by various means such as dedicated telephone lines, optical fiber links or by radio frequency communications. In FIG. 1, two exemplary cell-sites, 12 and 14, along with two exemplary mobile units, 16 and 18, which include cellular telephones, are illustrated. Arrows 20a-20b and 22a-22b respectively define the possible communication links between cell-site 12 and mobile units 16 and 18. Similarly, arrows 24a-24b and arrows 26a-26b respectively define the possible communication links between cell-site 14 and mobile units 18 and 16.

The cellular system illustrated in FIG. 1 may employ a variable rate data channel for communications between cell-sites 12, 14 and mobile units 16, 18. By example, a vocoder (not shown) may encode sampled voice information into symbols at four different rates according to the IS-95-A standard. The IS-95-A Mobile Station-Base Station Compatibility Standard for Dual Mode Wideband Spread Spectrum Cellular System has been provided by the telecommunications industry association (TIA) for CDMA communications. According to IS-95-A, speech is encoded at approximately 8,550 bits per second (bps), 4,000 bps, 2,000 bps, and 800 bps based on voice activity during a 20 millisecond (ms) frame of data. Each frame of vocoder data is then formatted with overhead bits as 9,600 bps, 4,800 bps, 2,400 bps, and 1,200 bps data frames for transmission. The 9,600 bps frame is referred to as a full rate frame; the 4,800 bps data frame is referred to as a half rate frame; a 2,400 bps data frame is referred to as a quarter rate frame; and a 1,200 bps data frame is referred to as an eighth rate frame. Although this example describes a set of four data rates of the IS-95-A standard, it should be recognized that the present invention is equally applicable in systems

utilizing different transmission rates and/or a different number of variable rates.

By encoding each frame of data based on speech activity, data compression is achievable without impacting the quality of the reconstructed speech. Since speech inherently contains periods of silence, i.e. pauses, the amount of data used to represent these periods can be reduced. Variable rate vocoding most effectively exploits this fact by reducing the data rate for these periods of silence. In a system with a set of four rates as described above, periods of active speech will generally be encoded at full rate, while periods of silence will generally be encoded at eighth rate. Most frames (about 80-90%) are encoded at full or eighth rate. Transitions between active speech and periods of silence will typically be encoded at half or quarter rate. An exemplary encoding technique which compresses data based on speech activity is described in U.S. Pat. No. 5,511,073 mentioned above.

The data frames are also formatted with overhead bits, which generally will include additional bits for error correction and detection, such as Cyclic Redundancy Check (CRC) bits. The CRC bits can be used by the decoder to determine whether or not a frame of data has been received correctly. CRC codes are produced by dividing the data block by a predetermined binary polynomial as is described in detail in IS-95-A.

In a preferred embodiment, each frame of symbol data is interleaved by an interleaver, preferably on a bit level basis, to increase time diversity for purposes of error detection. The formatted data frames undergo further processing, which include modulation, frequency upconversion to the radio frequency (RF) and amplification of the signals of data frames, before transmission.

When signals of the variable rate data frames are received by a receiver, the receiver must determine the rate of transmission in order to properly decode the signals. However, the rate of the received frame is not known by the mobile station a priori. Therefore, some other method of ascertaining the rate is necessary.

The present invention accomplishes rate determination through the use of hypothesis testing. Hypothesis tests are designed based on the probability distribution of the data rates of the frames of speech. Although the data rate of each received frame is not known a priori, the probability of receiving a frame at a given rate can be determined. As mentioned above, a variable rate vocoder encodes each frame of speech at one of a set of predetermined rates based on the speech activity during the time frame.

Since the characteristics of speech activity can be modeled, probabilistic functions of the data rates which depend on speech activity can be derived from the model. Hypothesis tests can then be designed based on the probabilistic functions of data rates to determine the most likely data rate for
5 each received frame of data.

The use of hypothesis testing for rate determination in a variable rate receiving system may be better appreciated by referring to FIG. 2. In a CDMA environment, for example, the receiving system 50 of FIG. 2 may be implemented in either a mobile unit or a cell site in order to determine the
10 data rate of received signals. The present invention offers particular advantages because it avoids the exhaustive decoding at all rates. By choosing a hypothesis and checking the hypothesis for correctness, the average amount of processing for each received frame is reduced. This is especially important in the mobile unit because reduced processing, and
15 thereby power consumption, in the decoding process can extend battery life in the receiver.

The variable rate receiving system 50 illustrated in FIG. 2 includes receiver 30 for collecting transmitted signals, including the data signal of interest. Receiver 30 amplifies and frequency downconverts the received
20 signals from the RF frequency band to the intermediate frequency (IF) band.

The IF signals are presented to demodulator 32. The design and implementation of demodulator 32 are described in detail in U.S. Pat. No. 5,490,165, entitled "DEMODULATION ELEMENT ASSIGNMENT IN A SYSTEM CAPABLE OF RECEIVING MULTIPLE SIGNALS," issued Feb. 6,
25 1996, and assigned to the assignee of the present invention, the disclosure of which is incorporated by reference herein. Demodulator 32 demodulates the IF signal to produce a data signal consisting of the symbols of one frame of data. Demodulator 32 generates the data signal by despreading and correlating the IF signal addressed to the receiver. The demodulated data
30 signal is then fed to buffer 33. Buffer 33 stores the demodulated data signal, or the received symbols, until it is properly decoded. Buffer 33 may also be the deinterleaver if the data frame had been interleaved for transmission. Buffer 33 provides the demodulated symbol data to decoder 34.

Hypothesis testing module 36 implements the hypothesis test for
35 determining the data rate of a received frame of data. Hypothesis testing module 36 comprises processor 40, which includes memory 38. The information needed in hypothesis testing such as the decoded rates from the previous frames and the probabilities are stored in memory 38. For each data frame received, processor 40 determines the most probable rate based

on the information stored in memory 38. Processor 40 then presents the most probable data rate to decoder 34 which decodes the data signal at this most probable rate to produce decoded bits.

5 In the exemplary embodiment, decoder 34 is a trellis decoder capable of decoding data of varying rates, such as a Viterbi decoder. The design and implementation of a multirate Viterbi decoder which exhaustively decodes a received signal at all rates of a set of rates is described in the
10 aforementioned U.S. Patent Applications 08/233,570 and 08/126,477. It will be understood by one skilled in the art that the multirate Viterbi decoder may be modified to decode at a selected rate. This may be accomplished by having the Viterbi decoder receive a rate indicator input, in response to which the decoder decodes the data signal according to the rate indicator. Thus, the modified Viterbi decoder may decode a received data frame based on a rate indicator supplied by processor 40 of hypothesis testing module 36.

15 Decoder 34 generates information data bits and error metrics characterizing the information bits. The error metrics include the previously described CRC bits, which were added into the data frames as overhead bits. Decoder 34 may also generate other error metrics, such as the Yamamoto Quality Metric and the Symbol Error Rate (SER). The
20 Yamamoto metric is determined by comparing the differences in the metrics of remerging paths in each step of the Viterbi decoding with a threshold and labeling a path as unreliable if the metric difference is less than a quality threshold. If the final path selected by the Viterbi decoder has been labeled as unreliable at any step, the decoder output is labeled as unreliable.
25 Otherwise, it is labeled as reliable. The Symbol Error Rate is determined by taking the decoded bits, re-encoding these bits to provided re-encoded symbols, and comparing these re-encoded symbols against the received symbols which are stored in buffer 33. The SER is a measure of the mismatching between the re-encoded symbols and the received symbols.
30 The decoded information bits and the error metrics are provided to data check element 42, which determines if the information bits have been correctly decoded.

In a preferred embodiment, data check element 42 first checks the CRC bits. If the CRC check fails, then data check element 42 provides a
35 signal indicative of the failure to processor 40. If the CRC check passes, then data check element 42 determines if the re-encoded SER is below a certain threshold. If the SER is above the threshold, then a signal indicative of failure is provided to processor 40. Otherwise, the data rate provided by hypothesis testing module 36 is determined to be correct, and a success

signal is provided to processor 40, whereupon no further decoding is performed on the data frame. The properly decoded data signal is presented to variable rate vocoder 44.

When processor 40 receives a failure signal indicating that data symbols have not been properly decoded into information bits, processor 40 will determine at least one other data rate from the set of data rates at which to decode the data symbols. Processor 40 provides the rate information to decoder 34, which decodes the data symbols at the rate provided. For each data rate at which the data signal is decoded, data check element 42 will determine the quality of the decoded information bits. Upon determination by data check element 42 that the correct data rate has been found, a signal of decoded information bits is provided to variable rate vocoder 44. Vocoder 44 will then process the information bits for interface with the user.

Hypothesis testing module 36 may implement any of a number of hypothesis tests for determining the data rate of a received frame of data. For example the hypothesis test may be based on known statistics of speech activity. It is known that for a set of four rates using 20 ms frames, a full rate frame will usually be followed by another full rate frame, while an eighth-rate frame will usually be followed by another eighth rate frame. Further, it is also known that most frames will either be full or eighth rate rather than half or quarter rate, because the periods of speech and silence do not occur in 20 ms bursts. Based on these characteristics, the hypothesis test may designate the rate of the previous frame of data as the most probable rate for the currently received frame of data.

In an exemplary implementation, the rate of the previous frame of data is stored in memory 38 of hypothesis testing module 36. When a data frame is received, processor 40 of hypothesis testing module 36 obtains the rate of the previous frame from memory 38 and presents it to decoder 34. Decoder 34 decodes the received data frame at the rate of the previous frame to produce information bits. Decoder 34 also generates error metrics which are then presented to data check element 42 along with the information bits. If data check element 42 determines from the error metrics that the decoded bits are of good quality, then the information bits are presented to vocoder 44. Otherwise, a failure indication is sent from data check element 42 to processor 40. Processor 40 may then have decoder 34 exhaustively decode the data frame at all other rates before determining the data rate. A flow chart illustrating some of the steps involved in rate determination as described in the embodiment above is shown in FIG. 3.

Alternatively, processor 40 may have decoder 34 sequentially decode the data frame according to a ranking from the next most likely rate to the least likely rate. The ranking may be determined in a number of ways, such as according to the probability distributions described below. For each
5 decoding, error metrics are generated by decoder 34 and checked by data check element 42 for correctness. When correctly decoded, the decoded frame is passed on to vocoder 44. A flow chart illustrating some of the processing steps of this embodiment is shown in FIG. 4.

Another implementation of hypothesis testing module 36 is based
10 upon the a priori probability distribution of data rates. For a set of four rates, the a priori probability distribution (P) of the data rates may be defined as:

$$P = \text{Prob}\{R_t\}, \quad (3)$$

15 where R_t refers to the full, half, quarter, or eighth rate at time t . The likelihood of receiving a frame at each of the different data rates of a set of rates are maintained in memory 38 of processor 40. Generally, the probability distribution of the data rates are determined based on the theoretical statistics or the empirical statistics of speech activity. The
20 likelihood of receiving a frame at the different rates are then permanently stored in memory 38 for determining the rate of every received frame of data. In a more sophisticated embodiment, the likelihood of the rates stored in memory 38 may be updated based on the actual statistics of the received frames of data.

25 For each new frame of data received, processor 40 obtains the most probable rate from memory 38 and presents the most probable rate to decoder 34. Decoder 34 decodes the data signal at this most probable data rate and presents the decoded data to data check element 42. Error metrics, including the CRC, are also generated by decoder 34 and presented to data
30 check element 42. Other error metrics may also be generated for checking by data check element 42. If the error metrics indicate that the decoded bits are of good quality, then the information bits are presented to vocoder 44. Otherwise, a failure indication is sent from data check element 42 to processor 40. Then, processor 40 obtains the second most likely data rate
35 from memory 38 and presents it to decoder 34, and the process of decoding and error checking is continued until the correct data rate is found. A flow chart of the processing steps of this embodiment is illustrated in FIG. 5. Alternatively, upon receipt of a failure signal by processor 40, processor 40 may cause decoder 34 to exhaustively decode the data frame at each of the

other data rates of the set of rates, and error metrics are checked for each decoding in order to determine the actual rate of transmission. A flow chart of the processing steps of this embodiment is illustrated in FIG. 6.

5 Instead of designing the hypothesis test based on the simple probability distribution of the data rates, conditional probabilities may be used to improve on the accuracy of the rate determination. For example, the probability of receiving a data frame at a given rate may be defined to be conditioned on the actual rates of the previous frames of data. Conditional probabilities based on the previous rates work well because transition
10 characteristics of the data signals are well known. For example, if the rate two frames ago was eighth rate and the rate for the previous frame was half rate, then the most likely rate for the current frame is full rate, because the transition to half rate indicates the onset of active speech. Conversely, if the rate two frames ago was full rate and the rate for the previous frame was
15 quarter rate, then the most likely rate for the present frame might be eighth rate, because the rate transition indicates the onset of silence.

The probability distribution of the data rates conditioned on the rates of the previous n frames of data may be defined as:

$$20 \quad P = \text{Prob}\{ R_t \mid R_{t-1}, R_{t-2}, \dots, R_{t-n} \} \quad (4)$$

where R_t again refers to the rate at time t, and $R_{t-1}, R_{t-2}, \dots, R_{t-n}$ refers to rate(s) of the previous n frame(s) of data, for $n \geq 1$. The likelihood of receiving a frame at each of the different data rates of a set of rates
25 conditioned on the previous n actual rates are stored in memory 38 of processor 40. In addition, the actual data rates of the previous n frames of data are maintained by processor 40, and may be stored in memory 38 as the rates are determined.

For each received frame of data, processor 40 will determine the most
30 probable data rate conditioned on the previous n actual data rates and present it to decoder 34. Decoder 34 will decode the frame at this most probable data rate and present the decoded bits to data check element 42. In addition, error metrics are generated by decoder 34 and presented to data check element 42. If the error metrics indicate that the decoded bits are of
35 good quality, then the information bits are presented to vocoder 44. Also, processor 40 is informed of the rate decision so that it can maintain the history of chosen rates. That is, processor 40 is supplied R_t so that it can be used in determining $\text{Prob}\{ R_t \mid R_{t-1}, R_{t-2}, \dots, R_{t-n} \}$ for the next frame. If error metrics indicate an unsuccessful decoding, then a failure indication signal is

sent from data check element 42 to processor 40, and processor 40 determines the second most probable data rate conditioned on the previous n actual data rates to decode the data frame. As in the simple probabilities case, the process of decoding and error checking is continued until the correct data rate is found. Some of the processing steps of this embodiment are illustrated in a flow chart in FIG. 7. Also as in the simple probabilities case, after a failed decoding at the most likely rate, decoder 34 may exhaustively decode the data frame at all of the other data rates and have error metrics checked for all decoding in order to determine the data rate. Some of the processing steps of this embodiment are illustrated in a flow chart in FIG. 8.

It should be understood that the conditional probability distribution of the data rates may depend on statistics other than the actual rates of the previous frames of data. For example, the probability distribution may be conditioned on one or more frame quality measurements. The probability distribution is then defined to be:

$$P = \text{Prob}\{R_t \mid X_1, X_2, \dots, X_k\}, \quad (5)$$

where R_t is the rate at time t , and X_1, X_2, \dots, X_k are one or more frame quality measurements. The k frame quality measurements may be measurements performed on the current frame of data, or measurements performed on previous frame(s) of data, or a combination of both. An example of a frame quality measurement is the SER error metric mentioned above. Thus, the probability of receiving a frame at a given rate is conditioned on the SER obtained from the previous decoding if a previous decoding had been performed.

The conditional probability distribution may also depend on a combination of the actual rates of the previous frames of data and the frame quality measurements. In this case, the probability distribution of the data rates is defined as:

$$P_t = \text{Prob}\{R_t \mid R_{t-1}, R_{t-2}, \dots, R_{t-n}, X_1, X_2, \dots, X_k\}, \quad (6)$$

where R_t is the rate at time t , $R_{t-1}, R_{t-2}, \dots, R_{t-n}$ are the rates of the previous frames of data and X_1, X_2, \dots, X_k are the frame quality measurements.

In the cases where the probability distribution is based on frame quality measurements, the frame quality measurements should be maintained in processor 40 of hypothesis testing module 36. As can be seen from the above description, the hypothesized frame rate may be conditioned

on a number of different statistics, and the rates of the previous frames and the frame quality measurements are examples of these statistics. For each data frame received, processor 40 uses the statistics to determine the rate at which to decode the frame.

5 A further refinement to the determination of the rate at which to decode a received frame of data considers the processing costs of decoding the frame at the various rates in conjunction with hypothesis testing. In this embodiment, an optimum test sequence of the rates is established based on both the probability distribution of the data rates and the cost of decoding
10 at each of the data rates. The optimum test sequence is maintained by processor 40, which causes decoder 34 to sequentially decode a received frame of data according to the optimum sequence until the correct rate is found. The optimum test sequence is established to minimize the total expected cost of the rate search. Denoting P_i to be the probability that the rate search will stop at test T_i , and C_i to be the cost for conducting test T_i , the total expected cost of the rate search using test sequence T_1, T_2, \dots, T_M , where M is the number of possible rates in the system and $1 \leq i \leq M$, can be modeled as:

$$20 \quad C_{\text{total}} = C_1 * P_1 + (C_1 + C_2) * P_2 + \dots + (C_1 + C_2 + \dots + C_M) * P_M. \quad (7)$$

The optimum test sequence is found by minimizing the total expected cost C_{total} .

25 In Equation (7), the cost C_i for conducting test T_i will generally be the processing power required for decoding a frame at the rate specified by test T_i . The cost may be assigned to be proportional to the frame rate specified by the test T_i because the computational complexity of decoder 34 is in general approximately proportional to the number of bits per frame. The
30 probabilities P_i may be assigned by the unconditioned a priori probability distribution of data rates as defined by Equation (3), or any of the conditional probability distributions defined by Equations (4), (5), or (6) above.

In a variable rate communications system where data frames are transmitted at 9,600 bps, 4,800 bps, 2,400 bps, and 1,200 bps, the following
35 example illustrates the formulation of the optimum test sequence for rate determination of a received frame. The costs to decode the 9,600 bps, 4,800 bps, 2,400 bps, and 1,200 bps frames are assumed to be 9.6, 4.8, 2.4, and 1.2, respectively. Further, the probability of receiving a frame at each of the four

rates is assumed to be the unconditioned a priori probabilities having the following values:

$$\text{Prob}(9,600 \text{ bps}) = 0.291, \quad (8)$$

$$5 \quad \text{Prob}(4,800 \text{ bps}) = 0.039, \quad (9)$$

$$\text{Prob}(2,400 \text{ bps}) = 0.072, \text{ and} \quad (10)$$

$$\text{Prob}(1,200 \text{ bps}) = 0.598. \quad (11)$$

10 The probabilities given in Equations (8)-(11) are derived from steady state empirical data.

15 A listing of all possible test sequences for rate determination in the system transmitting frames at 9,600, 4,800, 2,400, and 1,200 bps is shown in Table I below. In Table I, column 1 lists all possible test sequences T_1, T_2, T_3, T_4 , where $T_i = 1$ refers to the test of decoding at 9,600 bps, $T_i = 1/2$ refers to the test of decoding at 4,800 bps, $T_i = 1/4$ refers to the test of decoding at 2,400 bps, and $T_i = 1/8$ refers to the test of decoding at 1,200 bps. Columns 2 and 3 list the probability P_1 and the cost C_1 of performing the test T_1 , columns 4 and 5 list the probability P_2 and the cost C_2 of performing the test T_2 , columns 6 and 7 list the probability P_3 and the cost C_3 of performing the test T_3 , and columns 8 and 9 list the probability P_4 and the cost C_4 of performing the test T_4 . The total cost C_{total} of performing the test sequence T_1, T_2, T_3, T_4 is listed in column 10.

T ₁ , T ₂ , T ₃ , T ₄	P ₁	C ₁	P ₂	C ₂	P ₃	C ₃	P ₄	C ₄	C _{total}
1, 1/2, 1/4, 1/8	0.291	9.6	0.039	4.8	0.072	2.4	0.598	1.2	15.33
1, 1/2, 1/8, 1/4	0.291	9.6	0.039	4.8	0.598	1.2	0.072	2.4	13.98
1, 1/4, 1/2, 1/8	0.291	9.6	0.072	2.4	0.039	4.8	0.598	1.2	15.08
1, 1/4, 1/8, 1/2	0.291	9.6	0.072	2.4	0.598	1.2	0.039	4.8	12.25
1, 1/8, 1/2, 1/4	0.291	9.6	0.598	1.2	0.039	4.8	0.072	2.4	11.16
1, 1/8, 1/4, 1/2	0.291	9.6	0.598	1.2	0.072	2.4	0.039	4.8	10.90
1/2, 1, 1/4, 1/8	0.039	4.8	0.291	9.6	0.072	2.4	0.598	1.2	16.35
1/2, 1, 1/8, 1/4	0.039	4.8	0.291	9.6	0.598	1.2	0.072	2.4	15.00
1/2, 1/4, 1, 1/8	0.039	4.8	0.072	2.4	0.291	9.6	0.598	1.2	16.36
1/2, 1/4, 1/8, 1	0.039	4.8	0.072	2.4	0.598	1.2	0.291	9.6	10.97
1/2, 1/8, 1, 1/4	0.039	4.8	0.598	1.2	0.291	9.6	0.072	2.4	9.61
1/2, 1/8, 1/4, 1	0.039	4.8	0.598	1.2	0.072	2.4	0.291	9.6	9.62
1/4, 1, 1/2, 1/8	0.072	2.4	0.291	9.6	0.039	4.8	0.598	1.2	15.08
1/4, 1, 1/8, 1/2	0.072	2.4	0.291	9.6	0.598	1.2	0.039	4.8	12.26
1/4, 1/2, 1, 1/8	0.072	2.4	0.039	4.8	0.291	9.6	0.598	1.2	16.11
1/4, 1/2, 1/8, 1	0.072	2.4	0.039	4.8	0.598	1.2	0.291	9.6	10.71
1/4, 1/8, 1/2, 1	0.072	2.4	0.598	1.2	0.039	4.8	0.291	9.6	7.89
1/4, 1/8, 1, 1/2	0.072	2.4	0.598	1.2	0.291	9.6	0.039	4.8	6.87
1/8, 1, 1/4, 1/2	0.598	1.2	0.291	9.6	0.072	2.4	0.039	4.8	5.51
1/8, 1, 1/2, 1/4	0.598	1.2	0.291	9.6	0.039	4.8	0.072	2.4	5.76
1/8, 1/2, 1, 1/4	0.598	1.2	0.039	4.8	0.291	9.6	0.072	2.4	6.79
1/8, 1/2, 1/4, 1	0.598	1.2	0.039	4.8	0.072	2.4	0.291	9.6	6.79
1/8, 1/4, 1/2, 1	0.598	1.2	0.072	2.4	0.039	4.8	0.291	9.6	6.54
1/8, 1/4, 1, 1/2	0.598	1.2	0.072	2.4	0.291	9.6	0.039	4.8	5.52

Table I

5 As shown in Table I, the optimum test sequence is the sequence 1/8,
 1, 1/4, 1/2 shown in the 19th row. This test sequence offers the lowest total
 expected cost of processing. Therefore, the rate determination system would
 decode a received frame of data at 1,200 bps first. If the decoding at 1,200 bps
 is not successful, then the frame would be decoded sequentially at 9,600 bps,
 10 2,400 bps, and 4,800 bps until the correct rate is found. In a preferred
 embodiment, the optimum test sequence is maintained by processor 40 of
 hypothesis testing module 36. For each frame of data received, processor 40
 causes decoder 34 to decode the frame sequentially according to the
 optimum test sequence, with each decoding checked by data check element
 15 42, until the correct data rate is found. Processing resources are efficiently
 utilized in this rate determination system because the decoding is performed
 sequentially according to an optimum search sequence.

Based on the embodiments described above, it will be understood by
 one skilled in the art that the present invention is applicable to all systems

in which data has been encoded according to a variable rate scheme and the data must be decoded in order to determine the rate. Even more generally, the invention is applicable to all systems in which the encoded data E is a function of the data D and some key k, and there exists some information in D or E which permits the verification of the correct D by the receiver. The sequence k may be time varying. The encoded data is represented as:

$$E = f(D,k), \quad (1)$$

where k is from a small set K of keys and where some probability function exists on the set of keys. The inverse of the encoding, or the decoding, can be represented as:

$$D = f^{-1}(E,k), \quad (2)$$

where k is chosen so that D is correct.

As an example, assume that D is data composed of fixed-length sequence D1 and fixed length sequence D2 so that $D = D1, D2$. Sequence D2 is the Cyclic Redundancy Code (CRC) of D1, so that $D2 = f_{crc}(D1)$. Assume also that the encoding function, $f(D,k)$, is an exclusive-OR of a fixed-length D with the fixed length sequence k. Then, the decoding, $f^{-1}(E,k)$, would be the exclusive-OR of E with the correct k. The correct k is verified by checking whether $D2 = f_{crc}(D1)$. The correct k can be found by decoding all possible k's in K and then determining whether the CRC check passes. Alternatively, it can be done by sequentially decoding using one k at a time, with no further decoding once the "correct" k is found. According to the present invention, the order of sequential decoding is to be determined by hypothesis testing. A number of hypothesis tests, including the tests described above, may be utilized. The order of sequential decoding may in addition depend on the cost of processing, as described above. The use of hypothesis testing and/or cost functions in formulating a test sequence for rate determination reduces the average amount of processing as fewer k's will have to be tried.

The previous description of the preferred embodiments is provided to enable any person skilled in the art to make or use the present invention. The various modifications to these embodiments will be readily apparent to those skilled in the art, and the generic principles defined herein may be applied to other embodiments without the use of the inventive faculty. Thus, the present invention is not intended to be limited to the

embodiments shown herein but is to be accorded the widest scope consistent with the principles and novel features disclosed herein.

CLAIMS

1. In a variable rate communications system, a sub-system for
2 determining, at a receiver, the data rate of a received data frame, comprising:
a processor for generating a signal indicating the most likely rate of
4 said received data frame in accordance with a predetermined hypothesis test;
and
6 a decoder for receiving said most likely rate signal and for decoding
said received data frame into a decoded frame of bits at said most likely rate.

2. The rate determination sub-system of claim 1 wherein said
2 most likely rate is the rate of the previous data frame.

3. The rate determination sub-system of claim 1 wherein said
2 hypothesis test is based on an a priori probability distribution of data rates.

4. The rate determination sub-system of claim 1 wherein said
2 hypothesis test is based on a conditional probability distribution of data rates
conditioned on the rate of at least one previous data frame.

5. The rate determination sub-system of claim 1 wherein said
2 hypothesis test is based on a conditional probability distribution of data rates
conditioned on at least one frame quality measurement.

6. The rate determination sub-system of claim 1 further
2 comprising a data check element for receiving said decoded bits, generating
error metrics characterizing said decoded bits, and generating a quality
4 indication based on said error metrics for said decoded bits.

7. The rate determination sub-system of claim 6,
2 further comprising a vocoder for receiving said decoded bits and
processing said decoded bits to provide speech to an user upon generation of
4 a positive indication of said quality; and
wherein upon generation of a negative indication of said quality, said
6 processor further causes said decoder to perform additional decoding of said
received data frame in accordance with at least one rate other than said most
8 likely rate.

8. The rate determination sub-system of claim 7,
2 wherein said additional decoding is performed sequentially in
accordance with a predetermined test sequence of data rates;
4 wherein said data check element generates error metrics for each said
additional decoding and generates a quality indication based on said error
6 metrics for each said additional decoding; and
wherein said additional decoding terminates upon generation of a
8 positive indication of said quality.

9. The rate determination sub-system of claim 7,
2 wherein said additional decoding comprises exhaustive decoding of
said received data frame at all rates of a rate set except said most likely rate;
4 and
wherein said data check element generates error metrics for each said
6 additional decoding and determines the rate of said received data frame in
accordance with said error metrics.

10. The rate determination sub-system of claim 6 wherein said
2 error metrics include a Cyclic Redundancy Check result.

11. The rate determination sub-system of claim 6 wherein said
2 error metrics include a Symbol Error Rate metric.

12. The rate determination sub-system of claim 6 wherein said
2 error metrics include a Yamamoto quality metric.

13. The rate determination sub-system of claim 1 wherein said
2 processor comprises a memory for storing said most likely rate.

14. The rate determination sub-system of claim 1 wherein said
2 decoder is a Viterbi decoder.

15. In a variable rate communications system, a sub-system for
2 determining, at a receiver, the data rate of a received data frame, comprising:
a processor for generating a test sequence of data rates for determining
4 the rate of a received data frame, said test sequence being generated in
accordance with a predetermined hypothesis test;

6 a decoder for decoding said received data frame sequentially according
to said test sequence and generating a decoded frame of bits for each rate at
8 which said received data frame is decoded;
a data check element for generating error metrics characterizing said
10 decoded bits and for generating a quality indication based on said error
metrics for each rate at which said received data frame is decoded; and
12 wherein no further decoding is performed upon generation of a
positive indication of said quality.

16. The rate determination sub-system of claim 15 wherein said
2 hypothesis test is based on an a priori probability distribution of data rates.

17. The rate determination sub-system of claim 15 wherein said
2 hypothesis test is based on a conditional probability distribution of data rates
conditioned on the rate of at least one previous data frame.

18. The rate determination sub-system of claim 15 wherein said
2 hypothesis test is based on a conditional probability distribution of data rates
conditioned on at least one frame quality measurement.

19. The rate determination sub-system of claim 16 wherein said
2 test sequence is generated further in accordance with the cost of decoding
said received data frame at each of said data rates.

20. The rate determination sub-system of claim 17 wherein said
2 test sequence is generated further in accordance with the cost of decoding
said received data frame at each of said data rates.

21. The rate determination sub-system of claim 18 wherein said
2 test sequence is generated further in accordance with the cost of decoding
said received data frame at each of said data rates.

22. The rate determination sub-system of claim 15 further
2 comprising a vocoder for receiving said decoded bits and processing said
decoded bits to provide speech to an user upon generation of a positive
4 indication of said quality.

23. The rate determination sub-system of claim 15 wherein said
2 error metrics include a Cyclic Redundancy Check result.

24. The rate determination sub-system of claim 15 wherein said
2 error metrics include a Symbol Error Rate metric.

25. The rate determination sub-system of claim 15 wherein said
2 error metrics include a Yamamoto quality metric.

26. The rate determination sub-system of claim 15 wherein said
2 processor comprises a memory for storing said test sequence of data rates.

27. The rate determination sub-system of claim 15 wherein said
2 decoder is a Viterbi decoder.

28. A method for determining the rate of a received data frame in
2 a variable rate communications system, comprising the steps of:
receiving a wide-band signal;
4 demodulating said wide-band signal to produce a data signal, wherein
said data signal has been transmitted at one of a set of possible transmission
6 rates;
generating a test sequence of data rates for determining the rate of said
8 data signal, said test sequence being generated in accordance with a
predetermined hypothesis test;
10 decoding said data signal sequentially according to said test sequence
to generate a decoded frame of bits for each rate at which said data signal is
12 decoded;
generating error metrics characterizing said decoded frame of bits for
14 each rate at which said data signal is decoded;
generating a quality indication based on said error metrics for each
16 rate at which said data signal is decoded; and
upon generation of a positive indication of said quality, providing
18 said decoded frame of bits to a vocoder which processes said decoded bits to
provide speech to an user.

29. The method of claim 28 wherein said hypothesis test is based
2 on an a priori probability distribution of data rates.

30. The method of claim 28 wherein said hypothesis test is based
2 on a conditional probability distribution of data rates conditioned on the rate
of at least one previous data frame.

31. The method of claim 28 wherein said hypothesis test is based
2 on a conditional probability distribution of data rates conditioned on at least
one frame quality measurement.

32. The method of claim 29 wherein said test sequence is generated
2 further in accordance with the cost of decoding said received data frame at
each of said data rates.

33. The method of claim 30 wherein said test sequence is generated
2 further in accordance with the cost of decoding said received data frame at
each of said data rates.

34. The method of claim 31 wherein said test sequence is generated
2 further in accordance with the cost of decoding said received data frame at
each of said data rates.

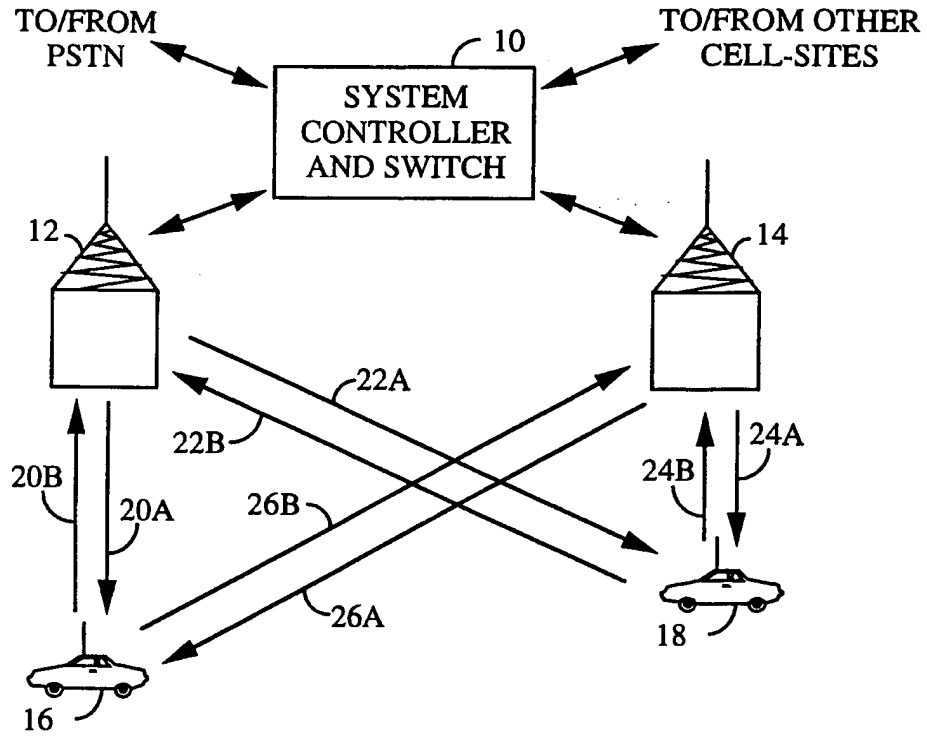


FIG. 1

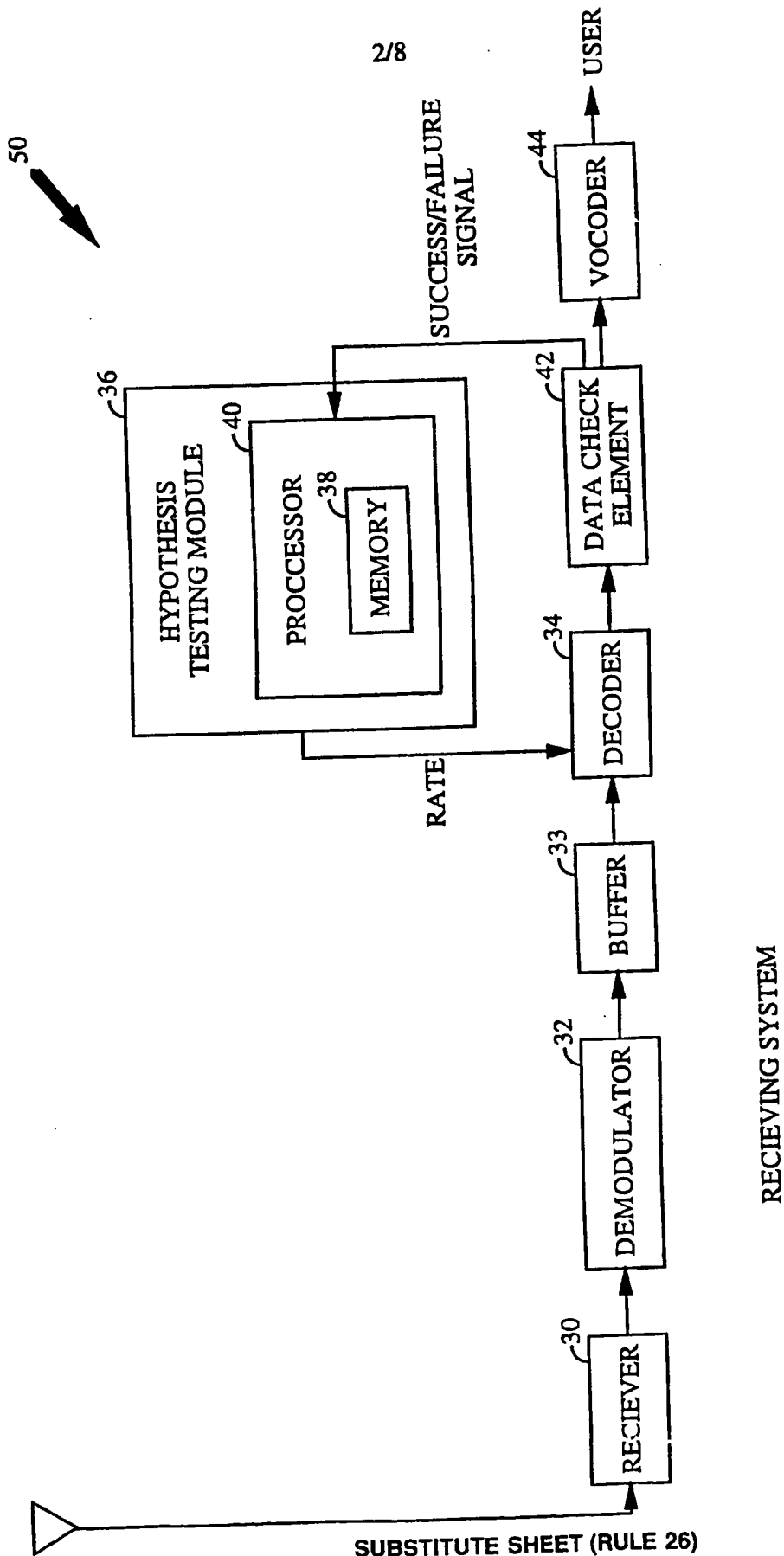


FIG. 2

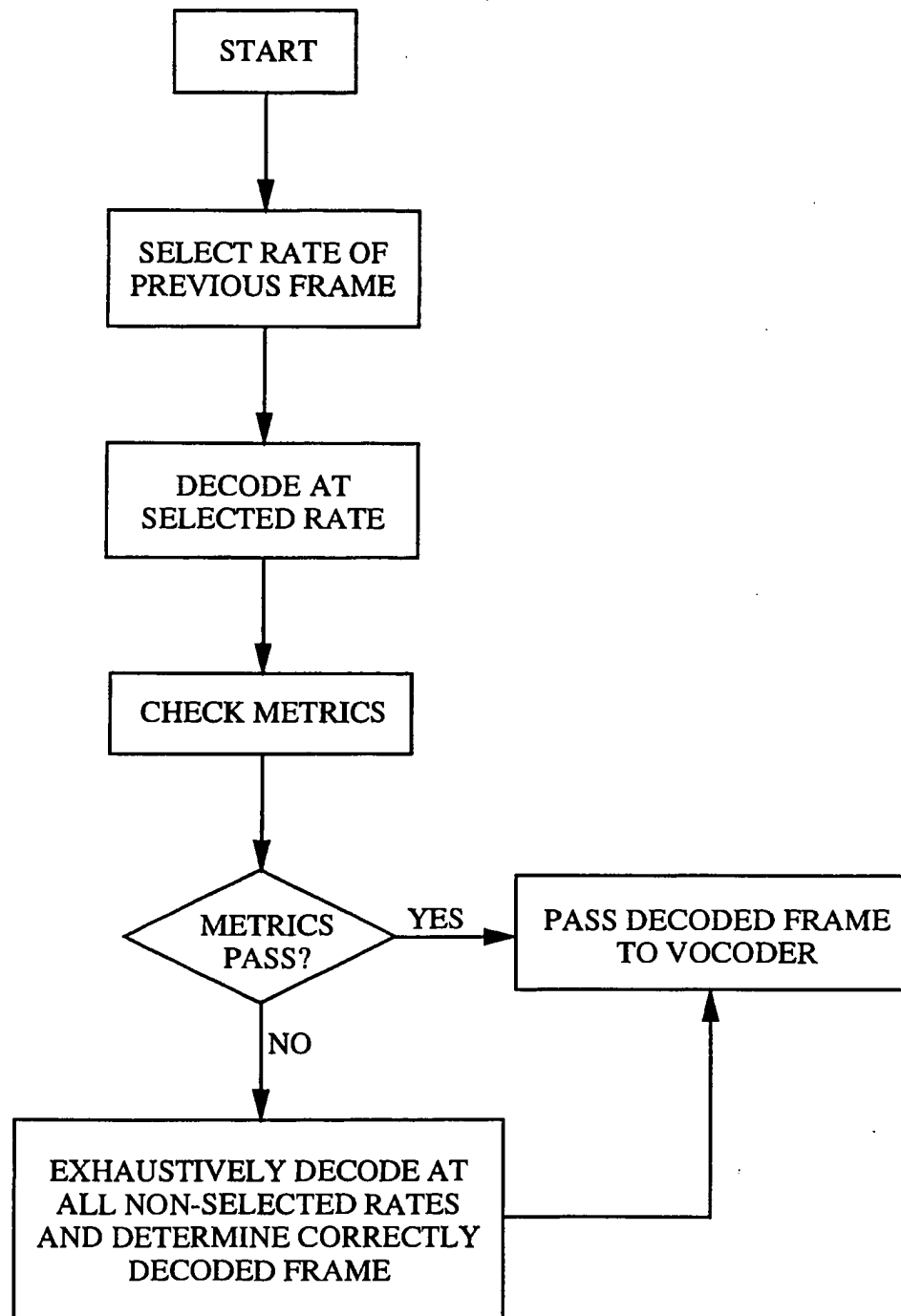


FIG. 3

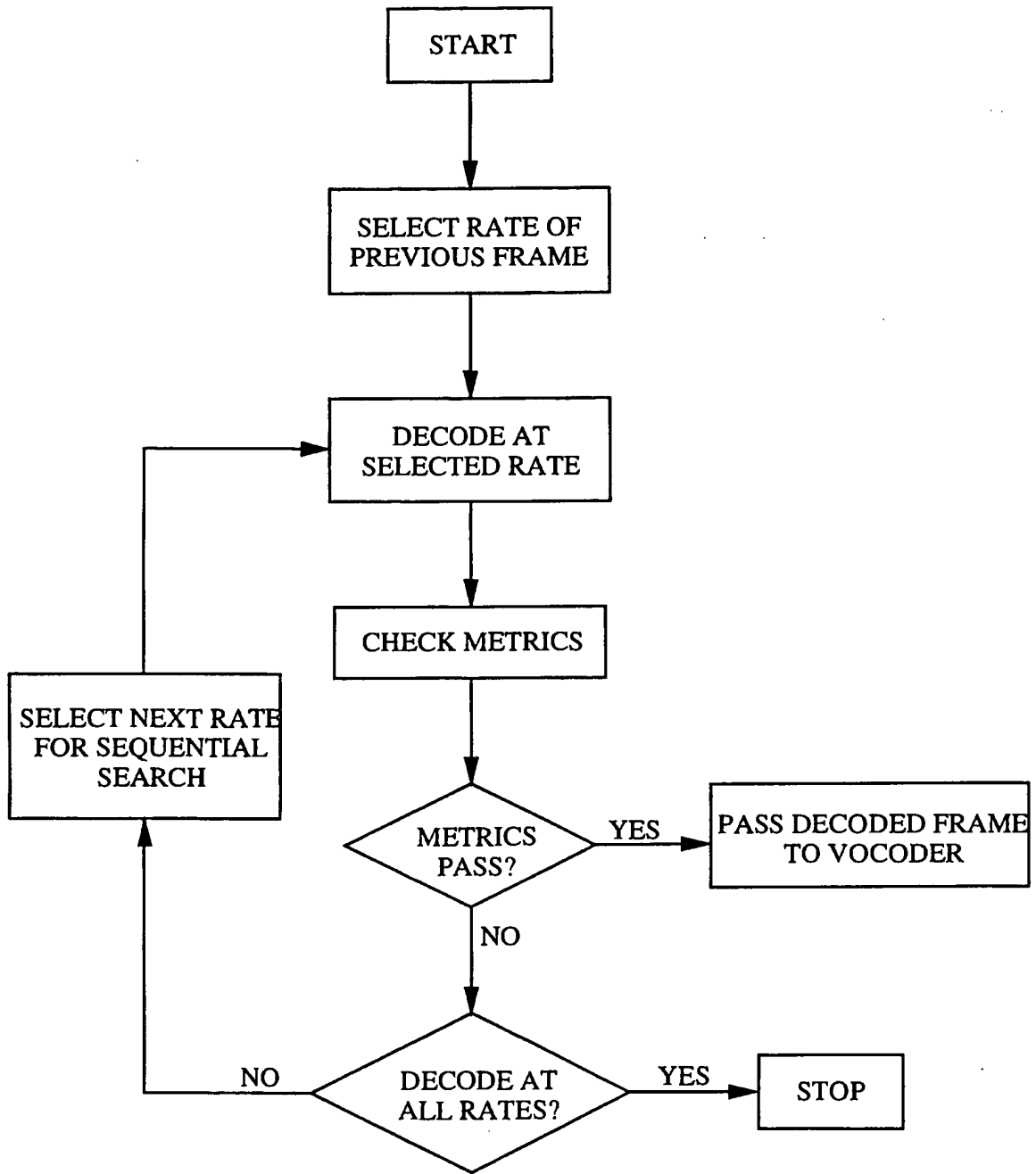


FIG. 4

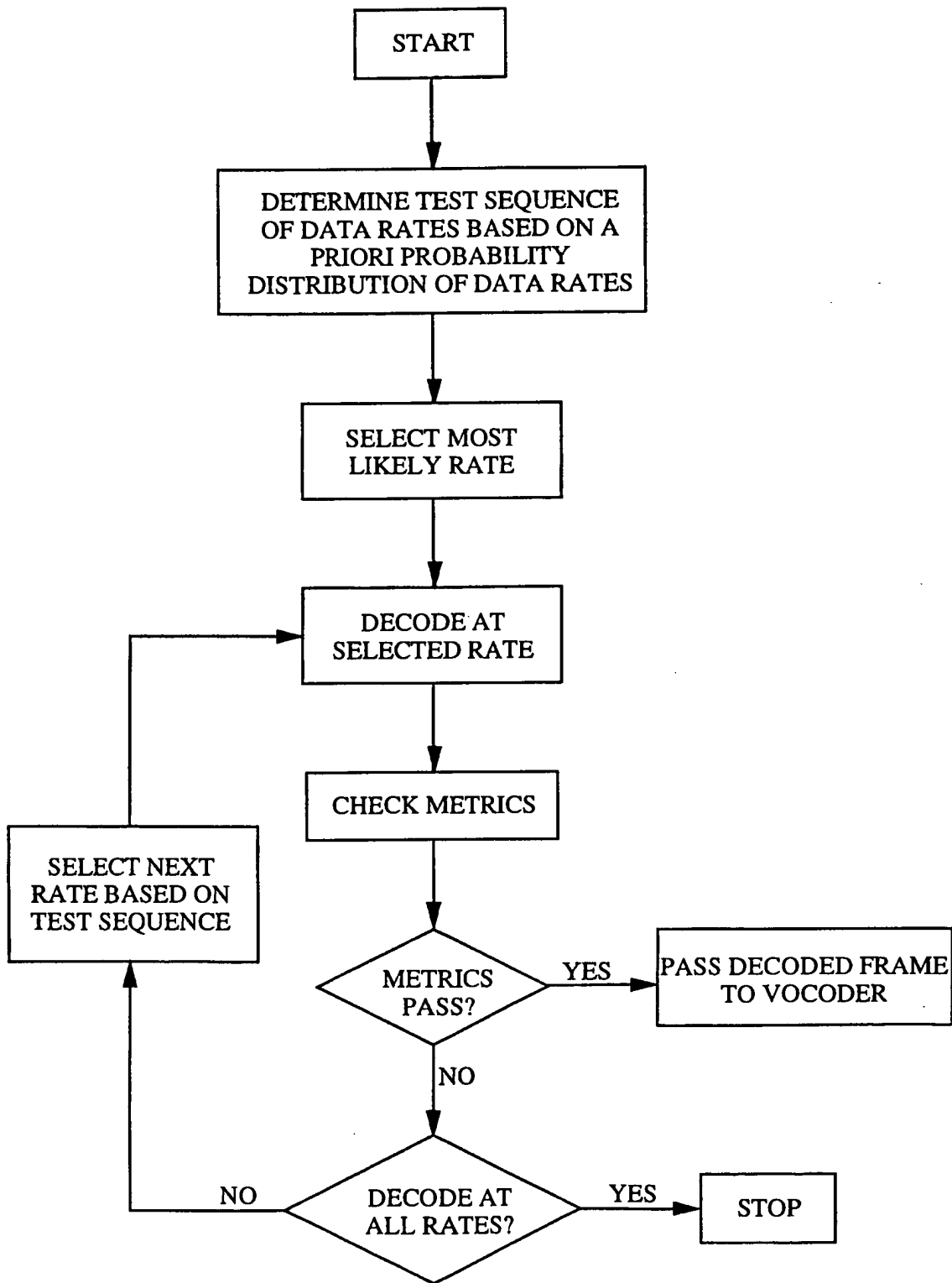


FIG. 5

6/8

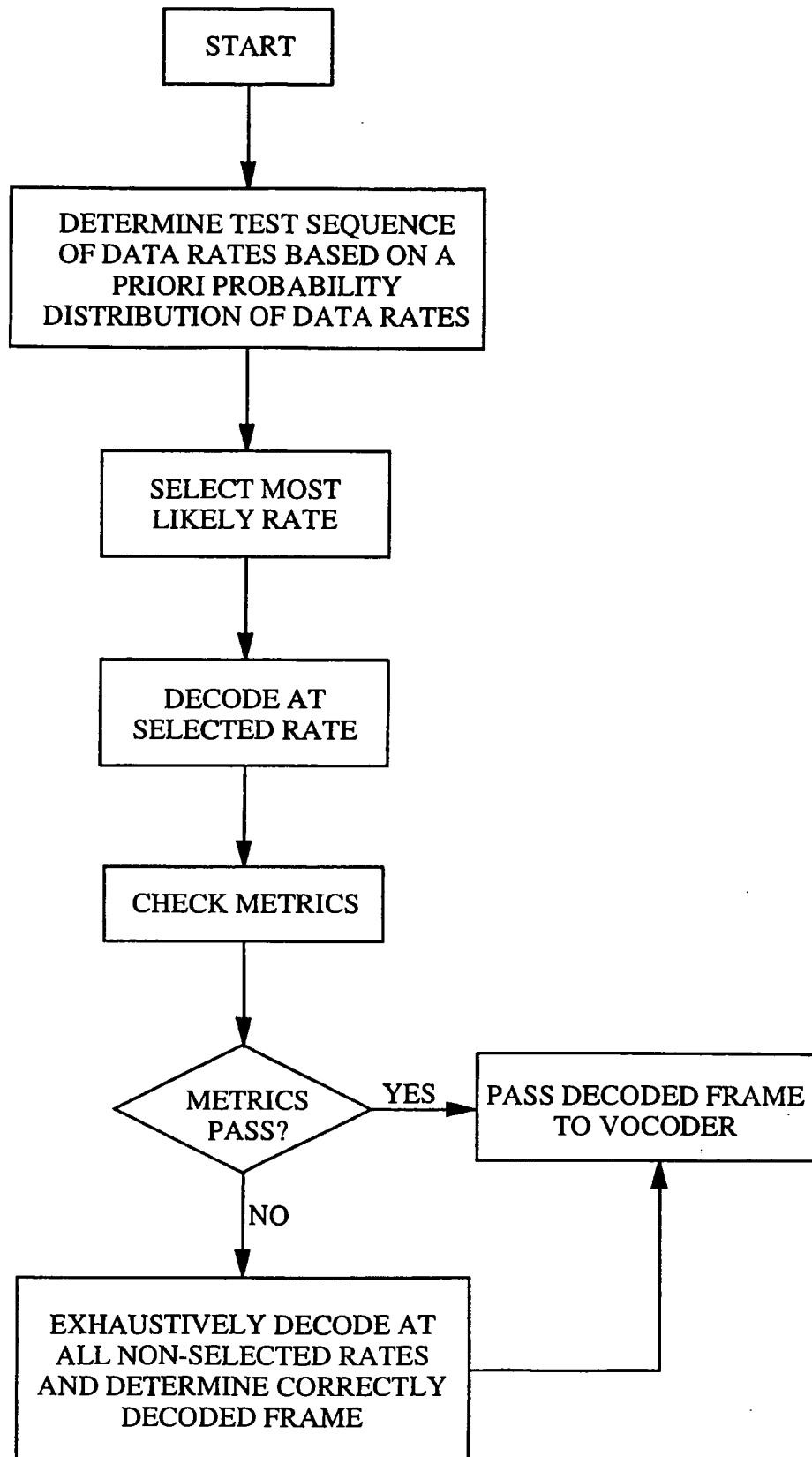


FIG. 6

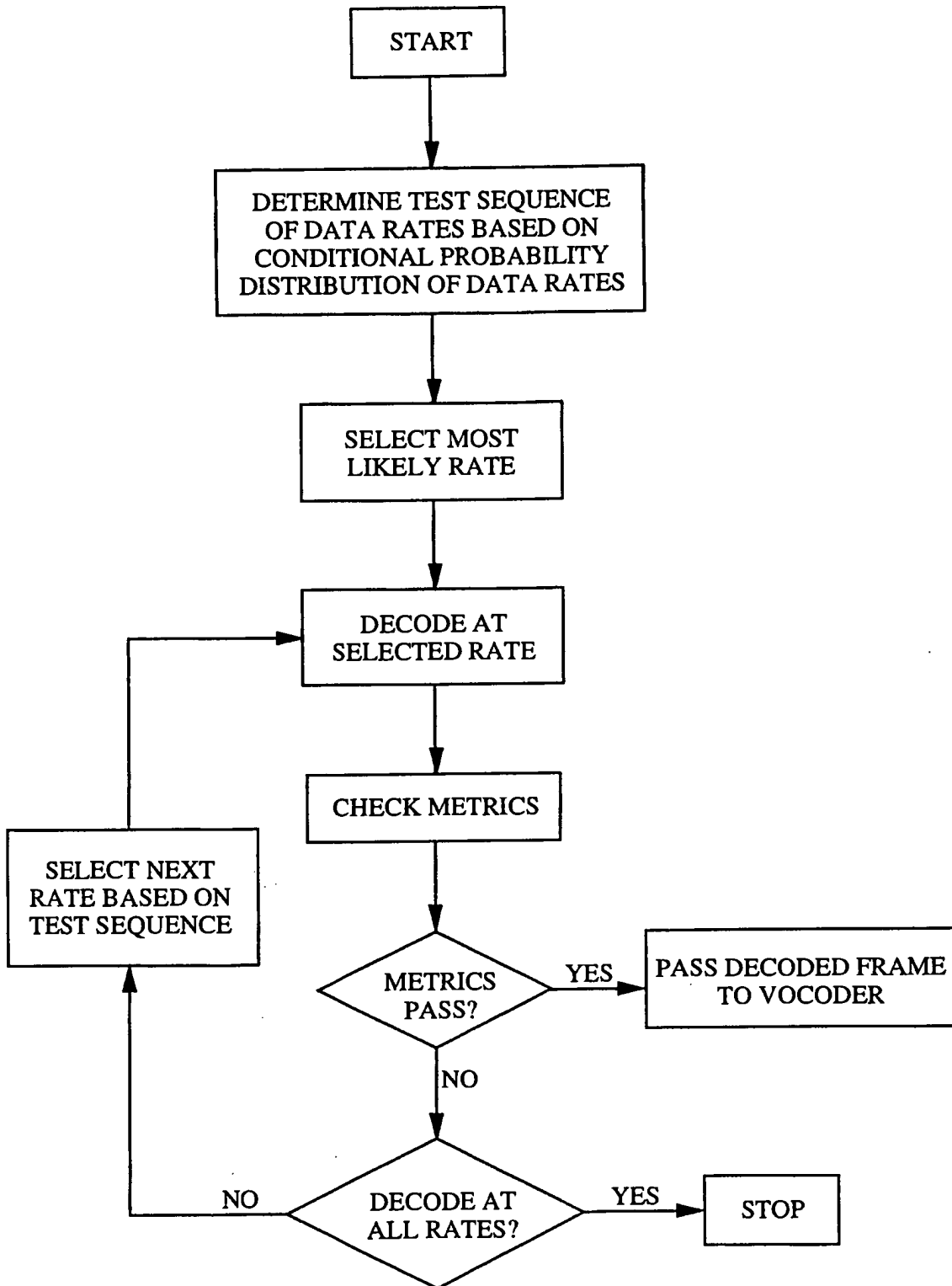


FIG. 7

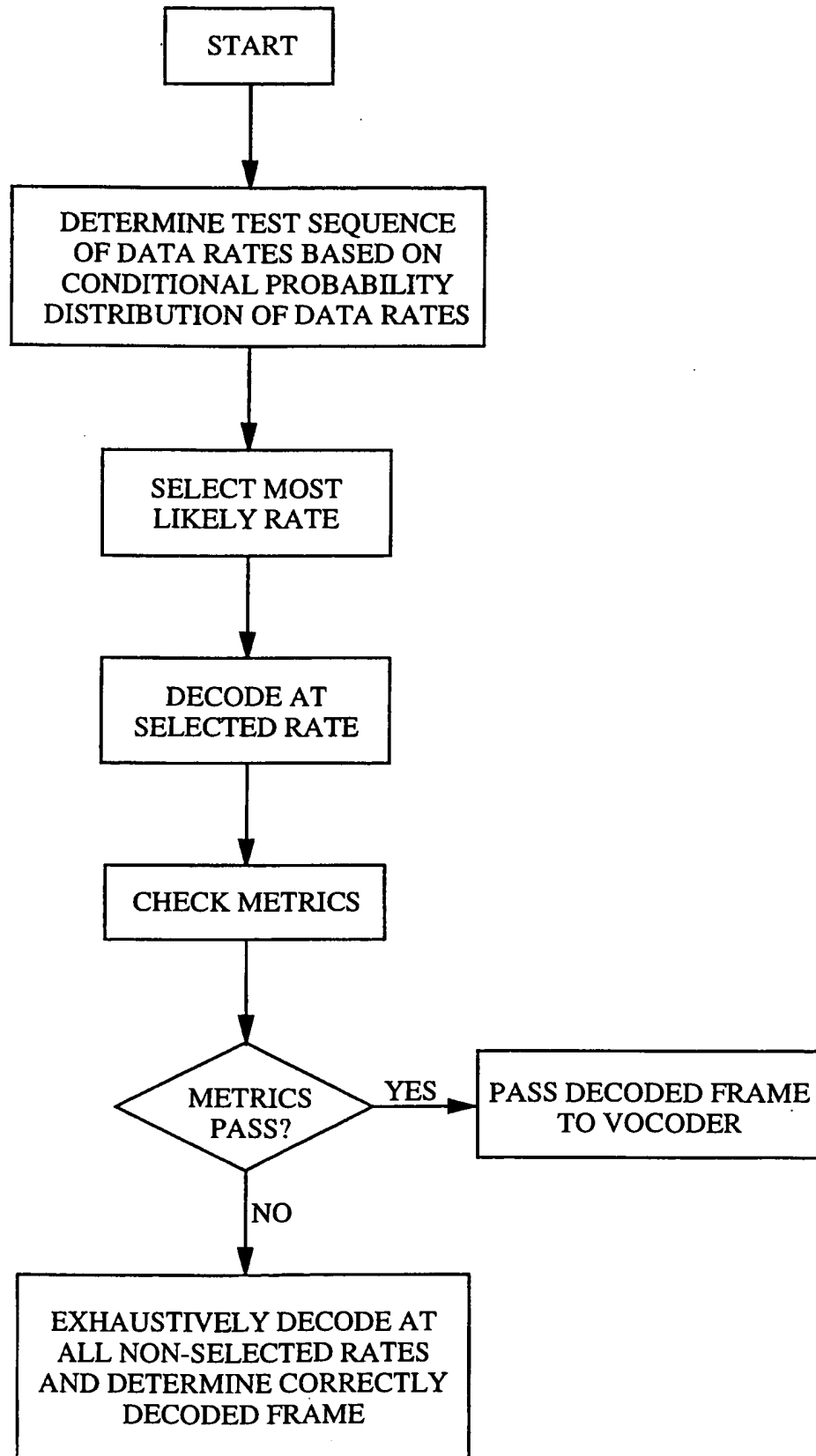


FIG. 8

INTERNATIONAL SEARCH REPORT

International Application No

PCT/US 97/19676

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6 H04L25/02

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 6 H04L

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	EP 0 711 056 A (NIPPON ELECTRIC CO) 8 May 1996 see abstract see page 2, column 2, line 46 - page 3, column 3, line 26 see page 4, column 5, line 13 - column 6, line 23; figure 2 ---	1, 14, 15, 22, 27, 28
A	EP 0 713 305 A (NIPPON ELECTRIC CO) 22 May 1996 see abstract see page 2, column 2, line 45 - page 3, column 3, line 12 see page 3, column 3, line 33 - column 4, line 2; figure 1 see page 3, column 4, line 44 - page 4, column 5, line 13 --- -/--	1, 14, 15, 22, 27, 28

Further documents are listed in the continuation of box C.

Patent family members are listed in annex.

* Special categories of cited documents :

- "A" document defining the general state of the art which is not considered to be of particular relevance
- "E" earlier document but published on or after the international filing date
- "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- "O" document referring to an oral disclosure, use, exhibition or other means
- "P" document published prior to the international filing date but later than the priority date claimed

- "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
- "&" document member of the same patent family

Date of the actual completion of the international search <p style="text-align: center;">27 February 1998</p>	Date of mailing of the international search report <p style="text-align: center;">09/03/1998</p>
Name and mailing address of the ISA European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Authorized officer <p style="text-align: center;">Bossen, M</p>

INTERNATIONAL SEARCH REPORT

International Application No
PCT/US 97/19676

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 5 566 206 A (BUTLER BRIAN K ET AL) 15 October 1996 cited in the application see column 2, line 37 - line 47 see column 5, line 57 - column 6, line 32; figure 2 see column 7, line 35 - line 52; figure 4 -----	7,9-12, 14, 22-25,27

1

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No

PCT/US 97/19676

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
EP 0711056 A	08-05-96	JP 8130535 A	21-05-96
		AU 3662595 A	09-05-96
		CA 2163134 A	03-05-96

EP 0713305 A	22-05-96	JP 2596392 B	02-04-97
		JP 8149567 A	07-06-96
		AU 686026 B	29-01-98
		AU 3787595 A	23-05-96
		CA 2162417 A	17-05-96
		FI 955398 A	17-05-96

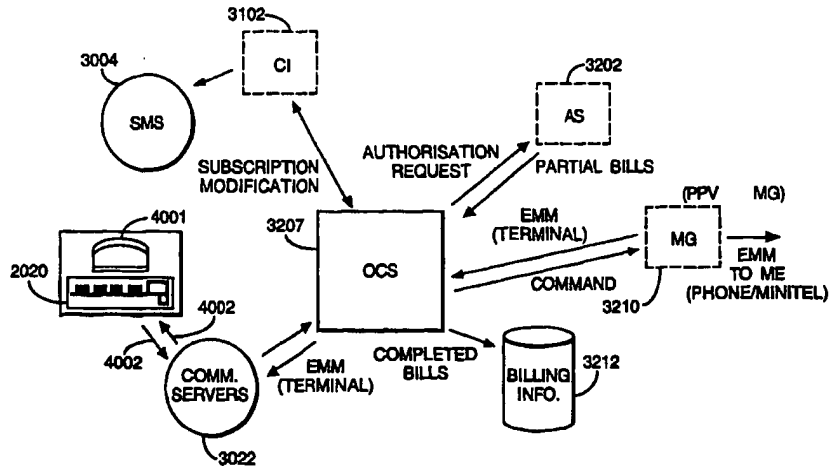
US 5566206 A	15-10-96	AT 158910 T	15-10-97
		AU 683479 B	13-11-97
		AU 7113694 A	17-01-95
		BR 9406891 A	26-03-96
		DE 69405997 D	06-11-97
		EP 0705512 A	10-04-96
		FI 956091 A	16-02-96
		JP 9501548 T	10-02-97
		WO 9501032 A	05-01-95
		CN 1108834 A	20-09-95
		IL 109842 A	30-09-97
		MX 9404610 A	31-01-95
		ZA 9404032 A	09-03-95



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

<p>(51) International Patent Classification ⁶ : H04N 7/16, 7/167</p>	<p>A1</p>	<p>(11) International Publication Number: WO 98/43426 (43) International Publication Date: 1 October 1998 (01.10.98)</p>
<p>(21) International Application Number: PCT/EP97/02108 (22) International Filing Date: 25 April 1997 (25.04.97) (30) Priority Data: 97400650.4 21 March 1997 (21.03.97) EP (34) Countries for which the regional or international application was filed: FR et al. (71) Applicant (for all designated States except US): CANAL+ SOCIETE ANONYME [FR/FR]; 85/89, quai André Citroën, F-75711 Paris Cedex 15 (FR). (72) Inventors; and (75) Inventors/Applicants (for US only): BAYASSI, Mulham [FR/FR]; 30, rue de Chambéry, F-75015 Paris (FR). DE LA TULLAYE, Pierre [FR/FR]; 7, allée Marcel Jouhandeau, F-92500 Rueil Malmaison (FR). JEZEQUEL, Jean-François [FR/FR]; 35, rue du Commandant Kieffer, F-95240 Corneille en Parisis (FR). (74) Agent: COZENS, Paul, Dennis; Mathys & Squire, 100 Grays Inn Road, London WC1X 8AL (GB).</p>		<p>(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, HU, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ARIPO patent (GH, KE, LS, MW, SD, SZ, UG), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG). Published With international search report.</p>

(54) Title: BROADCAST AND RECEPTION SYSTEM, AND CONDITIONAL ACCESS SYSTEM THEREFOR



(57) Abstract

A digital satellite television system has a plurality of set-top-boxes associated with a plurality of end users' television receivers, a modem and a decoder housed in each STB, a Subscriber Authorization System (SAS) incorporating or having associated therewith a plurality of communication servers, means included in the SAS for generating Electronic Managements Messages (EMM), a back channel interconnecting each of the STBs individually with the SAS, means included in the SAS and each STB so that the necessary information required to inject a relevant EMM into the system is supplied directly to the relevant communication server included in or associated with the SAS to authorise the release of the said EMM and/or means to connect the modem to the back channel and means whereby an EMM is transmissible to the decoder directly from a relevant communication server included in or associated with the SAS. Further important features are also disclosed.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Licchtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

- 1 -

**BROADCAST AND RECEPTION SYSTEM, AND CONDITIONAL ACCESS
SYSTEM THEREFOR**

The present invention relates to a broadcast and reception system, in particular to a mass-market digital interactive satellite television system, and to a conditional access
5 system therefor.

In particular, but not exclusively, the invention relates to a mass-market broadcast system having some or all of the following preferred features:-

- It is an information broadcast system, preferably a radio and/or television broadcast system
- 10 ● It is a satellite system (although it could be applicable to cable or terrestrial transmission)
- It is a digital system, preferably using the MPEG, more preferably the MPEG-2, compression system for data/signal transmission
- It affords the possibility of interactivity.

15 More particularly the present invention relates to so-called pay television (or radio) where a user/viewer selects a programme/film/game to be viewed for which payment is to be made, this being referred to as a pay-per-view (PPV) or in the case of data to be downloaded a so-called pay-per-file (PPF).

20 With such known PPV or PPF systems a significant amount of time is required to be spent by the user/viewer in order to carry out the actions necessary to actually access the product being selected.

For example, in one known system the sequence of steps which have to be carried out are as follows:-

25 I) The user telephones a so-called Subscriber Management System (SMS) which in this known system includes a number of human operators which answer the subscriber's call and to whom the subscriber communicates the necessary information concerning the selected product and concerning the financial status of the subscriber

- 2 -

to a so-called Subscriber Authorization System (SAS) which has included in it or associated with it a plurality of communications servers.

- ii) The operator at the SMS then has to check the financial status of the user before authorising the connection from the communications servers to the user's television set so that the product can be delivered and viewed by the user.

In another known system the human operator is replaced by an automatic voice server so that when the user telephones the SMS he/she hears a voice activated recording to which the user conveys the same information as I) above.

- This second arrangement reduces the delay inherent in the first described arrangement which can be more easily overloaded when large numbers of users are wishing to order a product at the same time.

However, even with this second arrangement the user is involved in inputting significant information in the form of lengthy serial numbers which operation provides plenty of scope for error as well as being time consuming.

- The third known arrangement involves the user making use of existing screen based systems such as MINITEL in France and PRESTEL in the United Kingdom, which systems replace the voice activated server referred to above in connection with the second arrangement. The MINITEL and PRESTEL systems themselves incorporate a modem at the consumer end.

- In all these known arrangements the user is involved in the expenditure of significant time and effort in inputting all the information necessary to enable the system to in effect authorize the transmission of the chosen product to the user's television set.

In the case of a satellite television system there is a further delay involved in the user actually receiving the product selected.

- 3 -

In PPV and PPF systems the key element in controlling the user's access to products are so-called Entitlement Management Messages (EMM) which have to be injected into the system in order to give the user product access. More particularly the EMMs are the mechanism by which the encrypted data representative of a product is
5 decrypted for a particular individual user.

In known satellite television systems the EMMs are transmitted to the user's televisions via the satellite link at regular intervals in the MPEG-2 data stream. Thus in the case of a particular user's EMM there can be a significant delay of perhaps several minutes before the user's next EMM transmission arrives at that user's
10 television set.

This transmission delay is in addition to the delay referred to earlier which is inherent in the user having to manually input certain data into the system. The cumulative effect of these two delays is that it may take perhaps typically five minutes for a user to be able to gain access to the selected product.

15 The present invention is concerned with overcoming this problem.

In a first aspect, the present invention provides a conditional access system comprising:

means for generating a plurality of (preferably conditional access) messages;
and

20 means for receiving the messages, said receiving means being adapted to communicate with said generating means via a communications server connected directly to said generating means.

Preferably, the message is an entitlement message for transmission (for example by broadcast) to the receiving means, said generating means being adapted to generate
25 entitlement messages in response to data received from said receiving means.

The generating means may be arranged to transmit a message as a packet of digital

- 4 -

data to said receiving means either via said communications server or via a satellite transponder.

The receiving means may be connectable to said communications server via a modem and telephone link.

- 5 In a related aspect, the present invention provides a conditional access system for affording conditional access to subscribers, comprising:
- a subscriber management system;
 - a subscriber authorization system coupled to the subscriber management system; and
 - 10 a communications server; said server being connected directly to the subscriber authorization system.

The system may further comprise a receiver/decoder for the subscriber, the receiver/decoder being connectable to said communications server, and hence to said subscriber authorization system, via a modem and telephone link.

- 15 In a second aspect, the present invention provides a broadcast and reception system including a conditional access system as described above.

In a third aspect, the present invention provides a broadcast and reception system comprising:

- 20 means for generating a plurality of entitlement messages relating to broadcast programs;
- means for receiving said messages from said generating means; and
- means for connecting the receiving means to the generating means to receive said messages, said connecting means being capable of effecting a dedicated connection between the receiving means and the generating means.

- 25 The dedicated connection would typically be a hard-wired connection and/or a modemmed connection, with the possibility of the connection been made via a cellular

- 5 -

telephone system. In other words, the dedicated connection is capable of forming a channel of communication (from point to point). This is in contrast to broadcasting of information through the air or ambient medium. The connecting means would typically be a modem at the receiving means.

5 Hence, in a closely related aspect, the present invention provides a broadcast and reception system comprising:

means for generating a plurality of entitlement messages relating to broadcast programs;

means for receiving said messages from said generating means via a modem;

10 and

means for connecting said modem to said generating means and said receiving means.

The above features can afford the advantage of providing the user the necessary viewing authorization (via the EMM) more quickly than has hitherto been possible, partly because, since the SAS typically uses a smaller amount of computer code than
15 the SMS, the SAS can operate more efficiently (and in real time), partly because the SAS can itself, directly, generate the requisite EMM, and partly because the EMM can be passed to the user or subscriber via a dedicated (typically modemmed) link.

Preferably, the generating means is connected to said modem via a communications
20 server which is preferably included in or associated with said generating means.

The receiving means may be further adapted to receive said entitlement messages via a satellite transponder.

The receiving means may be a receiver/decoder comprising means for receiving a compressed MPEG-type signal, means for decoding the received signal to provide a
25 television signal and means for supplying the television signal to a television.

Preferably, the receiving means is adapted to communicate with said generating means

- 6 -

via said modem and connecting means. The receiving means may comprise means for reading a smartcard insertable thereinto by an end user, the smartcard having stored therein data to initiate automatically the transmission of a message from said receiving means to said generating means upon insertion of the smartcard by the end user.

In addition, the system may further comprise a voice link to enable the end user of the broadcast and reception system to communicate with the generating means.

It will be understood from the above that the present invention provides two arrangements by which the time it takes for an end user to access a desired product is reduced. Preferably both arrangements are employed to achieve the maximum time saving but either arrangement can be used individually.

According to a further aspect of the present invention, there is provided a broadcast and reception system, comprising, at the broadcast end:

a broadcast system including means for broadcasting a callback request;
and at the reception end:
a receiver including means for calling back the broadcast system in response to the callback request.

By providing that the broadcast system can request the receiver to call it back, the possibility is afforded of the broadcast system obtaining information from the receiver about the state of the receiver.

Preferably, the means for calling back the broadcast system includes a modem connectable to a telephone system. By using a modemed back channel, a simple way of putting the invention into effect can be provided.

Preferably also, the means for calling back the broadcast system is arranged to transfer to the broadcast system information concerning the receiver. This information might include the number of remaining tokens, the number of pre-booked sessions, and so

- 7 -

on.

Preferably, the broadcast system includes means for storing the information, so that it can be processed at a later time, as desired.

5 Preferably, the broadcast means is arranged to broadcast a callback request which includes a command that the callback be made at a given time, and the means for calling back the broadcast system is arranged to respond to said command. By arranging for the callback to be later than the actual request, greater flexibility can be imparted to the system.

10 The broadcasting means may be arranged to broadcast as the callback request one or more Entitlement Messages for broadcast.

15 Preferably, the broadcast system includes means for generating a check message (such as a random number) and passing this to the receiver, the receiver includes means for encrypting the check message and passing this to the broadcast system, and the broadcast system further includes means for decrypting the check message received from the receiver and comparing this with the original check message. In this way it can be checked whether the receiver is genuine.

Any of the above features may be combined together in any appropriate combination. They may also be provided, as appropriate, in method aspects.

20 Preferred features of the present invention will now be described, purely by way of example, with reference to the accompanying drawings, in which:-

Figure 1 shows the overall architecture of a digital television system according to the preferred embodiment of the present invention;

Figure 2 shows the architecture of a conditional access system of the digital television system;

- 8 -

Figure 3 shows the structure of an Entitlement Management Message used in the conditional access system;

Figure 4 is a schematic diagram of the hardware of a Subscriber Authorisation System (SAS) according to a preferred embodiment of the present invention;

5 Figure 5 is a schematic diagram of the architecture of the SAS;

Figure 6 is a schematic diagram of a Subscriber Technical Management server forming part of the SAS;

Figure 7 is a flow diagram of the procedure for automatic renewal of subscriptions as implemented by the SAS;

10 Figure 8 is a schematic diagram of a group subscription bitmap used in the automatic renewal procedure;

Figure 9 shows the structure of an EMM used in the automatic renewal procedure;

Figure 10 shows in detail the structure of the EMM;

15 Figure 11 is a schematic diagram of an order centralized server when used to receive commands directly through communications servers;

Figure 12 illustrates diagrammatically a part of Figure 2 showing one embodiment of the present invention;

Figure 13 is a schematic diagram of the order centralized server when used to receive commands from the subscriber authorization system to request a callback;

20 Figure 14 is a schematic diagram of the communications servers;

- 9 -

Figure 15 shows the manner in which EMM emission cycle rate is varied according to the timing of a PPV event;

Figure 16 is a schematic diagram of a Message Emitter used to emit EMMs;

Figure 17 is a schematic diagram showing the manner of storage of EMMs within the
5 Message Emitter;

Figure 18 is a schematic diagram of a smartcard;

Figure 19 is a schematic diagram of an arrangement of zones in the memory of the smartcard; and

Figure 20 is a schematic diagram of a PPV event description.

10 An overview of a digital television broadcast and reception system 1000 according to the present invention is shown in Figure 1. The invention includes a mostly conventional digital television system 2000 which uses the known MPEG-2 compression system to transmit compressed digital signals. In more detail, MPEG-2 compressor 2002 in a broadcast centre receives a digital signal stream (typically a
15 stream of video signals). The compressor 2002 is connected to a multiplexer and scrambler 2004 by linkage 2006. The multiplexer 2004 receives a plurality of further input signals, assembles one or more transport streams and transmits compressed digital signals to a transmitter 2008 of the broadcast centre via linkage 2010, which can of course take a wide variety of forms including telecom links. The transmitter
20 2008 transmits electromagnetic signals via uplink 2012 towards a satellite transponder 2014, where they are electronically processed and broadcast via notional downlink 2016 to earth receiver 2018, conventionally in the form of a dish owned or rented by the end user. The signals received by receiver 2018 are transmitted to an integrated receiver/decoder 2020 owned or rented by the end user and connected to the end user's
25 television set 2022. The receiver/decoder 2020 decodes the compressed MPEG-2 signal into a television signal for the television set 2022.

- 10 -

A conditional access system 3000 is connected to the multiplexer 2004 and the receiver/decoder 2020, and is located partly in the broadcast centre and partly in the decoder. It enables the end user to access digital television broadcasts from one or more broadcast suppliers. A smartcard, capable of decrypting messages relating to commercial offers (that is, one or several television programmes sold by the broadcast supplier), can be inserted into the receiver/decoder 2020. Using the decoder 2020 and smartcard, the end user may purchase events in either a subscription mode or a pay-per-view mode.

An interactive system 4000, also connected to the multiplexer 2004 and the receiver/decoder 2020 and again located partly in the broadcast centre and partly in the decoder, enables the end user to interact with various applications via a modemmed back channel 4002.

The conditional access system 3000 is now described in more detail.

With reference to Figure 2, in overview the conditional access system 3000 includes a Subscriber Authorization System (SAS) 3002. The SAS 3002 is connected to one or more Subscriber Management Systems (SMS) 3004, one SMS for each broadcast supplier, by a respective TCP-IP linkage 3006 (although other types of linkage could alternatively be used). Alternatively, one SMS could be shared between two broadcast suppliers, or one supplier could use two SMSs, and so on.

First encrypting units in the form of ciphering units 3008 utilising "mother" smartcards 3010 are connected to the SAS by linkage 3012. Second encrypting units again in the form of ciphering units 3014 utilising mother smartcards 3016 are connected to the multiplexer 2004 by linkage 3018. The receiver/decoder 2020 receives a "daughter" smartcard 3020. It is connected directly to the SAS 3002 by Communications Servers 3022 via the modemmed back channel 4002. The SAS sends amongst other things subscription rights to the daughter smartcard on request.

The smartcards contain the secrets of one or more commercial operators. The

- 11 -

"mother" smartcard encrypts different kinds of messages and the "daughter" smartcards decrypt the messages, if they have the rights to do so.

The first and second ciphering units 3008 and 3014 comprise a rack, an electronic VME card with software stored on an EEPROM, up to 20 electronic cards and one
5 smartcard 3010 and 3016 respectively, for each electronic card, one (card 3016) for encrypting the ECMs and one (card 3010) for encrypting the EMMs.

The operation of the conditional access system 3000 of the digital television system will now be described in more detail with reference to the various components of the television system 2000 and the conditional access system 3000.

10 Multiplexer and Scrambler

With reference to Figures 1 and 2, in the broadcast centre, the digital video signal is first compressed (or bit rate reduced), using the MPEG-2 compressor 2002. This compressed signal is then transmitted to the multiplexer and scrambler 2004 via the linkage 2006 in order to be multiplexed with other data, such as other compressed
15 data.

The scrambler generates a control word used in the scrambling process and included in the MPEG-2 stream in the multiplexer 2004. The control word is generated internally and enables the end user's integrated receiver/decoder 2020 to descramble the programme.

20 Access criteria, indicating how the programme is commercialised, are also added to the MPEG-2 stream. The programme may be commercialised in either one of a number of "subscription" modes and/or one of a number of "Pay Per View" (PPV) modes or events. In the subscription mode, the end user subscribes to one or more commercial offers, or "bouquets", thus getting the rights to watch every channel inside
25 those bouquets. In the preferred embodiment, up to 960 commercial offers may be selected from a bouquet of channels. In the Pay Per View mode, the end user is provided with the capability to purchase events as he wishes. This can be achieved

- 12 -

by either pre-booking the event in advance ("pre-book mode"), or by purchasing the event as soon as it is broadcast ("impulse mode"). In the preferred embodiment, all users are subscribers, whether or not they watch in subscription or PPV mode, but of course PPV viewers need not necessarily be subscribers.

5 Both the control word and the access criteria are used to build an Entitlement Control Message (ECM); this is a message sent in relation with one scrambled program; the message contains a control word (which allows for the descrambling of the program) and the access criteria of the broadcast program. The access criteria and control word are transmitted to the second encrypting unit 3014 via the linkage 3018. In this unit,
10 an ECM is generated, encrypted and transmitted on to the multiplexer and scrambler 2004.

Each service broadcast by a broadcast supplier in a data stream comprises a number of distinct components; for example a television programme includes a video component, an audio component, a sub-title component and so on. Each of these
15 components of a service is individually scrambled and encrypted for subsequent broadcast to the transponder 2014. In respect of each scrambled component of the service, a separate ECM is required.

Programme Transmission

The multiplexer 2004 receives electrical signals comprising encrypted EMMs from the
20 SAS 3002, encrypted ECMs from the second encrypting unit 3014 and compressed programmes from the compressor 2002. The multiplexer 2004 scrambles the programmes and transmits the scrambled programmes, the encrypted EMMs and the encrypted ECMs as electric signals to a transmitter 2008 of the broadcast centre via linkage 2010. The transmitter 2008 transmits electromagnetic signals towards the
25 satellite transponder 2014 via uplink 2012.

Programme Reception

The satellite transponder 2014 receives and processes the electromagnetic signals transmitted by the transmitter 2008 and transmits the signals on to the earth receiver

- 13 -

2018, conventionally in the form of a dish owned or rented by the end user, via
downlink 2016. The signals received by receiver 2018 are transmitted to the
integrated receiver/decoder 2020 owned or rented by the end user and connected to
the end user's television set 2022. The receiver/decoder 2020 demultiplexes the
5 signals to obtain scrambled programmes with encrypted EMMs and encrypted ECMs.

If the programme is not scrambled, that is, no ECM has been transmitted with the
MPEG-2 stream, the receiver/decoder 2020 decompresses the data and transforms the
signal into a video signal for transmission to television set 2022.

If the programme is scrambled, the receiver/decoder 2020 extracts the corresponding
10 ECM from the MPEG-2 stream and passes the ECM to the "daughter" smartcard 3020
of the end user. This slots into a housing in the receiver/decoder 2020. The daughter
smartcard 3020 controls whether the end user has the right to decrypt the ECM and
to access the programme. If not, a negative status is passed to the receiver/decoder
2020 to indicate that the programme cannot be descrambled. If the end user does
15 have the rights, the ECM is decrypted and the control word extracted. The decoder
2020 can then descramble the programme using this control word. The MPEG-2
stream is decompressed and translated into a video signal for onward transmission to
television set 2022.

Subscriber Management System (SMS)

20 A Subscriber Management System (SMS) 3004 includes a database 3024 which
manages, amongst others, all of the end user files, commercial offers (such as tariffs
and promotions), subscriptions, PPV details, and data regarding end user consumption
and authorization. The SMS may be physically remote from the SAS.

Each SMS 3004 transmits messages to the SAS 3002 via respective linkage 3006
25 which imply modifications to or creations of Entitlement Management Messages
(EMMs) to be transmitted to end users.

The SMS 3004 also transmits messages to the SAS 3002 which imply no

- 14 -

modifications or creations of EMMs but imply only a change in an end user's state (relating to the authorization granted to the end user when ordering products or to the amount that the end user will be charged).

As described later, the SAS 3002 sends messages (typically requesting information such as call-back information or billing information) to the SMS 3004, so that it will be apparent that communication between the two is two-way.

Entitlement Management Messages (EMMs)

The EMM is a message dedicated to an individual end user (subscriber), or a group of end users, only (in contrast with an ECM, which is dedicated to one scrambled programme only or a set of scrambled programmes if part of the same commercial offer). Each group may contain a given number of end users. This organisation as a group aims at optimising the bandwidth; that is, access to one group can permit the reaching of a great number of end users.

Various specific types of EMM are used in putting the present invention into practice. Individual EMMs are dedicated to individual subscribers, and are typically used in the provision of Pay Per View services; these contain the group identifier and the position of the subscriber in that group. So-called "Group" subscription EMMs are dedicated to groups of, say, 256 individual users, and are typically used in the administration of some subscription services. This EMM has a group identifier and a subscribers' group bitmap. Audience EMMs are dedicated to entire audiences, and might for example be used by a particular operator to provide certain free services. An "audience" is the totality of subscribers having smartcards which bear the same Operator Identifier (OPI). Finally, a "unique" EMM is addressed to the unique identifier of the smartcard.

The structure of a typical EMM is now described with reference to Figure 3. Basically, the EMM, which is implemented as a series of digital data bits, comprises a header 3060, the EMM proper 3062, and a signature 3064. The header 3060 in turn comprises a type identifier 3066 to identify whether the type is individual, group,

- 15 -

audience or some other type, a length identifier 3068 which gives the length of the EMM, an optional address 3070 for the EMM, an operator identifier 3072 and a key identifier 3074. The EMM proper 3062 of course varies greatly according to its type. Finally, the signature 3064, which is typically of 8 bytes long, provides a number of
5 checks against corruption of the remaining data in the EMM.

Subscriber Authorization System (SAS)

The messages generated by the SMS 3004 are passed via linkage 3006 to the Subscriber Authorization System (SAS) 3002, which in turn generates messages acknowledging receipt of the messages generated by the SMS 3004 and passes these
10 acknowledgements to the SMS 3004.

As shown in Figure 4, at the hardware level the SAS comprises in known fashion a mainframe computer 3050 (in the preferred embodiment a DEC machine) connected to one or more keyboards 3052 for data and command input, one or more Visual Display Units (VDUs) 3054 for display of output information and data storage means
15 3056. Some redundancy in hardware may be provided.

At the software level the SAS runs, in the preferred embodiment on a standard open VMS operating system, a suite of software whose architecture is now described in overview with reference to Figure 5; it will be understood that the software could alternatively be implemented in hardware.

20 In overview the SAS comprises a Subscription Chain area 3100 to give rights for subscription mode and to renew the rights automatically each month, a Pay Per View Chain area 3200 to give rights for PPV events, and an EMM Injector 3300 for passing EMMs created by the Subscription and PPV chain areas to the multiplexer and scrambler 2004, and hence to feed the MPEG stream with EMMs. If other rights are
25 to be granted, such as Pay Per File (PPF) rights in the case of downloading computer software to a user's Personal Computer, other similar areas are also provided.

One function of the SAS 3002 is to manage the access rights to television

- 16 -

programmes, available as commercial offers in subscription mode or sold as PPV events according to different modes of commercialisation (pre-book mode, impulse mode). The SAS 3002, according to those rights and to information received from the SMS 3004, generates EMMs for the subscriber.

- 5 The Subscription Chain area 3100 comprises a Command Interface (CI) 3102, a Subscriber Technical Management (STM) server 3104, a Message Generator (MG) 3106, and the Cipherring Unit 3008.

The PPV Chain area 3200 comprises an Authorisation Server (AS) 3202, a relational database 3204 for storing relevant details of the end users, a local blacklist database
10 3205, Database Servers 3206 for the database, an Order Centralized Server (OCS) 3207, a Server for Programme Broadcaster (SPB) 3208, a Message Generator (MG) 3210 whose function is basically the same as that for the Subscription Chain area and is hence not described further in any detail, and the Cipherring Unit 3008.

The EMM Injector 3300 comprises a plurality of Message Emitters (MEs) 3302, 3304,
15 3306 and 3308 and Software Multiplexers (SMUXs) 3310 and 3312. In the preferred embodiment, there are two MEs, 3302 and 3304 for the Message Generator 3106, with the other two MEs 3306 and 3308 for the Message Generator 3210. MEs 3302 and 3306 are connected to the SMUX 3310 whilst MEs 3304 and 3308 are connected to the SMUX 3312.

- 20 Each of the three main components of the SAS (the Subscription Chain area, the PPV Chain area and the EMM Injector) are now considered in more detail.

Subscription Chain Area

Considering first the Subscription Chain area 3100, the Command Interface 3102 is primarily for despatching messages from the SMS 3004 to the STM server 3104, as
25 well as to the OCS 3206, and from the OCS to the SMS. The Command Interface takes as input from the SMS either direct commands or batch files containing commands. It performs syntactic analysis on the messages coming from the STM

- 17 -

server, and is able to emit accurate messages when an error occurs in a message (parameter out of range, missing parameter, and so on). It traces incoming commands in textual form in a trace file 3110 and also in binary form in a replay file 3112 in order to be able to replay a series of commands. Traces can be disabled and the size
5 of files limited.

Detailed discussion of the STM server 3104 is now provided with particular reference to Figure 6. The STM server is effectively the main engine of the Subscription Chain area, and has the purpose of managing free rights, the creation of new subscribers and the renewal of existing subscribers. As shown in the figure, commands are passed on
10 to the Message Generator 3106, albeit in a different format from that in which the commands are passed to the STM server. For each command, the STM server is arranged to send an acknowledgement message to the CI only when the relevant command has been successfully processed and sent to the MG.

The STM server includes a subscriber database 3120, in which all the relevant
15 parameters of the subscribers are stored (smartcard number, commercial offers, state, group and position in the group, and so on). The database performs semantic checks of the commands sent by the CI 3102 against the content of the database, and updates the database when the commands are valid.

The STM server further manages a First In First Out (FIFO) buffer 3122 between the
20 STM server and the MG, as well as a backup disk FIFO 3124. The purpose of the FIFOs is to average the flow of commands from the CI if the MG is not able to respond for a while for any reason. They can also ensure that in the case of a crash of the STM server or MG no command will be lost, since the STM server is arranged to empty (that is, send to the MG) its FIFOs when restarted. The FIFOs are
25 implemented as files.

The STM server includes at its core an automatic renewal server 3126 which automatically generates renewals, and, if required by the operators, free rights. In this context, the generation of renewals may be thought of as including the generation of

- 18 -

rights for the first time, although it will be understood that the generation of new rights is initiated at the SMS. As will become apparent, the two can be treated by roughly the same commands and EMMs.

5 Having the STM separate from the SAS, and the automatic renewal server within the SAS rather than (in known systems) in the SMS 3004, is a particularly important feature, since it can significantly reduce the number of commands which need to be passed from the SMS to the SAS (bearing in mind that the SMS and SAS may be in different locations and operated by different operators). In fact, the two main commands required from the SMS are merely commands that a new subscription
10 should be started and that an existing subscription should be stopped (for example in the case of non-payment). By minimising command exchange between the SMS and SAS, the possibility of failure of command transfer in the linkage 3006 between the two is reduced; also, the design of the SMS does not need to take into account the features of the conditional access system 3000 generally.

15 Automatic renewal proceeds in the fashion indicated in the flow diagram of Figure 7. In order to reduce bandwidth, and given that a very high percentage of all renewals are standard, renewal proceeds in groups of subscribers; in the preferred embodiments there are 256 individual subscribers per group. The flow diagram begins with the start step 3130, and proceeds to step 3132 where a monthly activation of the renewal
20 function is made (although of course it will be appreciated that other frequencies are also possible). With a monthly frequency, rights are given to the end user for the current month and all of the following month, at which point they expire if not renewed.

In step 3134 the subscriber database 3120 is accessed in respect of each group and
25 each individual within that group to determine whether rights for the particular individual are to be renewed.

In step 3136, a group subscription bitmap is set up according to the contents of the subscriber database, as shown in Figure 8. The bitmap comprises a group identifier

- 19 -

(in this case Group 1 - "G1") 3138 and 256 individual subscriber zones 3140. The individual bits in the bitmap are set to 1 or zero according to whether or not the particular subscriber is to have his rights renewed. A typical set of binary data is shown in the figure.

- 5 In step 3142 the appropriate commands, including the group subscription bitmap, are passed to the Message Generator 3106. In step 3143 the Message Generator sets an obsolescence date to indicate to the smartcard the date beyond which the particular subscription EMM is not valid; typically this date is set as the end of the next month.

- 10 In step 3144 the Message Generator generates from the commands appropriate group subscription EMMs and asks the Ciphering Unit 3008 to cipher the EMMs, the ciphered EMMs being then passed to the EMM Injector 3300, which, in step 3146, injects the EMMs into the MPEG-2 data stream.

Step 3148 indicates that the above described procedure is repeated for each and every group. The process is finally brought to an end at stop step 3150.

- 15 The flow diagram described above with reference to Figure 7 relates in fact specifically to the renewal of subscriptions. The STM also manages in a similar way free audience rights and new subscribers.

- 20 In the case of free audience rights, available for specific television programmes or groups of such programmes, these are made available by the STM issuing a command to the Message Generator to generate appropriate audience EMMs (for a whole audience) with an obsolescence date a given number of days (or weeks) hence. The MG computes the precise obsolescence date based on the STM command.

- 25 In the case of new subscribers, these are dealt with in two stages. Firstly, on purchase the smartcard in the receiver/decoder 2020 (if desired by the operator) affords the subscriber free rights for a given period (typically a few days). This is achieved by generating a bitmap for the subscriber which includes the relevant obsolescence date.

- 20 -

The subscriber then passes his completed paperwork to the operator managing the subscriber (at the SMS). Once the paperwork has been processed, the SMS supplies to the SAS a start command for that particular subscriber. On receipt by the SAS of the start command, the STM commands the MG to assign a unique address to the new
5 subscriber (with a particular group number and position within the group) and to generate a special, so-called "commercial offer" subscription EMM (as opposed to the more usual "group" subscription EMM used for renewals) to provide the particular subscriber with rights until the end of the next month. From this point renewal of the subscriber can occur automatically as described above. By this two stage process it
10 is possible to grant new subscribers rights until the SMS issues a stop command.

It is to be noted that the commercial offer subscription EMM is used for new subscribers and for reactivation of existing subscribers. The group subscription EMM is used for renewal and suspension purposes.

With reference to Figure 9, a typical subscription EMM proper (that is, ignoring the
15 header and signature) generated by the above procedure comprises the following main portions, namely typically a 256 bit subscription (or subscribers' group) bitmap 3152, 128 bits of management ciphering keys 3154 for the ciphering of the EMM, 64 bits of each exploitation ciphering key 3156 to enable the smartcard 3020 to decipher a control word to provide access to broadcast programmes, and 16 bits of obsolescence
20 date 3158 to indicate the date beyond which the smartcard will ignore the EMM. In fact in the preferred embodiment three exploitation keys are provided, one set for the present month, one set for the next month, and one for resume purposes in the event of system failure.

In more detail, the group subscription EMM proper has all of the above components,
25 except the management ciphering keys 3154. The commercial offer subscription EMM proper (which is for an individual subscriber) includes instead of the full subscribers' group bitmap 3152 the group ID followed by the position in the group, and then management ciphering keys 3154 and three exploitation keys 3156, followed by the relevant obsolescence date 3158.

- 21 -

The Message Generator 3106 serves to transform commands issued by the STM server 3104 into EMMs for passing to the Message Emitter 3302. With reference to Figure 5, firstly, the MG produces the EMMs proper and passes them to the Ciphering Unit 3008 for ciphering with respect to the management and exploitation keys. The CU
5 completes the signature 3064 on the EMM (see Figure 3) and passes the EMM back to the MG, where the header 3060 is added. The EMMs which are passed to the Message Emitter are thus complete EMMs. The Message Generator also determines the broadcast start and stop time and the rate of emission of the EMMs, and passes these as appropriate directions along with the EMMs to the Message Emitter. The
10 MG only generates a given EMM once; it is the ME which performs its cyclic transmission.

Again with reference to Figure 5, the Message Generator includes its own EMM database 3160 which, for the lifetime of the relevant EMM, stores it. It is erased once its emission duration has expired. The database is used to ensure consistency between
15 the MG and ME, so that for example when an end user is suspended the ME will not continue to send renewals. In this regard the MG computes the relevant operations and sends them to the ME.

On generation of an EMM, the MG assigns a unique identifier to the EMM. When the MG passes the EMM to the ME, it also passes the EMM ID. This enables
20 identification of a particular EMM at both the MG and the ME.

Also concerning the Subscription Chain area, the Message Generator includes two FIFOs 3162 and 3164, one for each of the relevant Message Emitters 3302 and 3304 in the EMM Injector 3300, for storing the ciphered EMMs. Since the Subscription Chain area and EMM Injector may be a significant distance apart, the use of FIFOs
25 can allow full continuity in EMM transmission even if the links 3166 and 3168 between the two fail. Similar FIFO's are provided in the Pay Per View Chain area.

One particular feature of the Message Generator in particular and the conditional access system in general concerns the way that it reduces the length of the EMM

- 22 -

proper 3062 by mixing parameter length and identifier to save space. This is now described with reference to Figure 10 which illustrates an exemplary EMM (in fact a PPV EMM, which is the simplest EMM). The reduction in length occurs in the Pid (Packet or "Parameter" identifier) 3170. This comprises two portions, the actual ID
 5 3172, and the length parameter for the packet 3174 (necessary in order that the start of the next packet can be identified). The whole Pid is expressed in just one byte of information, 4 bits being reserved for the ID, and four for the length. Because 4 bits is not sufficient to define the length in true binary fashion, a different correspondence between the bits and the actual length is used, this correspondence being represented
 10 in a look-up table, stored in storage area 3178 in the Message Generator (see Figure 5). The correspondence is typically as follows:-

	0000 =	0
	0001 =	1
	0010 =	2
15	0011 =	3
	0100 =	4
	0101 =	5
	0110 =	6
	0111 =	7
20	1000 =	8
	1001 =	9
	1010 =	10
	1011 =	11
	1100 =	12
25	1101 =	16
	1110 =	24
	1111 =	32

It will be seen that the length parameter is not directly proportional to the actual length of the packet; the relationship is in part more quadratic rather than linear. This
 30 allows for a greater range of packet length.

- 23 -

Pay Per View Chain Area

Concerning the Pay Per View Chain area 3200, with reference to Figure 5 in more detail the Authorisation Server 3202 has as its client the Order Centralized Server 3207, which requests information about each subscriber which connects to the
5 Communications Servers 3022 to purchase a PPV product.

If the subscriber is known from the AS 3202, a set of transactions takes place. If the subscriber is authorized for the order, the AS creates a bill and sends it to the OCS. Otherwise, it signals to the OCS that the order is not authorized.

It is only at the end of this set of transactions that the AS updates the end users
10 database 3204 via the database servers (DBAS) 3206, if at least one transaction was authorized; this optimizes the number of database accesses.

The criteria according to which the AS authorizes purchase are stored in the database, accessed through DBAS processes. In one embodiment, the database is the same as the database accessed by the STM.

15 Depending on consumer profile, the authorization may be denied (PPV_Forbidden,Casino_Forbidden ...). These kind of criteria are updated by STM 3104, on behalf of the SMS 3004.

Other parameters are checked, such as limits allowed for purchase (either by credit card, automatic payment, or number of authorized token purchases per day).

20 In case of payment with a credit card, the number of the card is checked against a local blacklist stored in the local blacklist database 3205.

When all the verifications are successful, the AS:-

1. Generates a bill and sends it to the OCS, which completes this bill and stores it in a file, this file being later sent to the SMS for processing (customer actual
25 billing); and

- 24 -

2. Updates the database, mainly to set new purchase limits.

This check-and-generate-bill-if-OK mechanism applies for each command a subscriber may request during a single connection (it is possible to order e.g. 5 movies in a single session).

5 It is to be noted that the AS has a reduced amount of information concerning the subscriber, by comparison with that held by the SMS. For example, the AS does not hold the name or address of the subscriber. On the other hand, the AS does hold the smartcard number of the subscriber, the subscriber's consumer category (so that different offers can be made to different subscribers), and various flags which state
10 whether, for example, the subscriber may purchase on credit, or he is suspended or his smartcard has been stolen. Use of a reduced amount of information can help to reduce the amount of time taken to authorize a particular subscriber request.

The main purpose of the DBASs 3206 is to increase database performance seen from the AS, by paralleling the accesses (so actually it does not make much sense to define
15 a configuration with only one DBAS). An AS parameter determines how many DBASes should connect. A given DBAS may be connected to only one AS.

The OCS 2307 mainly deals with PPV commands. It operates in several modes.

Firstly, it operates to process commands issued by the SMS, such as product refreshment (for instance, if the bill is already stored by the SMS, no bill is generated
20 by the OCS), update of the wallet in the smartcard 3020, and session cancellation/update.

The various steps in the procedure are:-

1. Identifying the relevant subscriber (using the AS 3202);
2. If valid, generate adequate commands to the Message Generator, in order to
25 send an appropriate EMM. Commands may be:

Product commands,
Update of the wallet,

- 25 -

Session erasure.

Note that these operations do not imply creation of billing information, since billing is already known from the SMS. These operations are assimilated to "free products" purchase.

5 Secondly, the OCS deals with commands received from the subscribers through the Communications Servers 3022. These may be received either via a modem connected to the receiver/decoder 2020, or by voice activation via the telephone 4001, or by key activation via a MINITEL, PRESTEL or like system where available.

10 Thirdly, the OCS deals with callback requests issued by the SMS. These last two modes of operation are now discussed in more detail.

In the second type of mode described above it was stated that the OCS deals with commands received directly from the end user (subscriber) through the Communications Servers 3022. These include product orders (such as for a particular PPV event), a subscription modification requested by the subscriber, and a reset of a
15 parental code (a parental code being a code by which parents may restrict the right of access to certain programmes or classes of programmes).

The way in which these commands are dealt with is now described in more detail with reference to Figure 11.

Product orders by a subscriber involve the following steps:

- 20
1. Identifying through the AS the caller who is making a call through the CS 3022 ordering a particular product;
 2. Checking the caller's request validity, again using the AS (where the order is placed using the receiver/decoder 2020, this is achieved by verifying the smartcard 3020 details);
 - 25 3. Ascertain the price of the purchase;
 4. Check that the price does not exceed the caller's credit limit etc;
 5. Receiving a partial bill from the AS;

- 26 -

6. Filling additional fields in the bill to form a completed bill;
7. Adding the completed bill to a billing information storage file 3212 for later processing; and
8. Sending corresponding command(s) to the PPV Message Generator 3210 to generate the relevant EMM(s).

The EMM(s) is sent either on the modem line 4002 if the consumer placed the product order using the receiver/decoder 2020 (more details of this are described later), or else it is broadcast. The one exception to this is where there is some failure of the modem connection (in the case where the consumer places the order using the receiver/decoder); in this event the EMM is broadcast over the air.

A subscription modification requested by a subscriber involves:

1. Identifying the caller (using the AS);
2. Sending information to the Command Interface; the CI in turn forwards this information to the SMS; and
3. Via the CI, the OCS then receives an answer from the SMS (in terms of the cost of the modification, if the modification is possible).

If modification was requested using the receiver/decoder, the OCS generates a confirmation to the SMS. Otherwise, for example in the case of phone or Minitel, the subscriber is prompted for confirmation and this answer sent to the SMS via the OCS and the CI.

Reset of a parental code involves:

1. Identifying the caller (using AS); and
 2. Sending a command to the MG to generate an appropriate EMM bearing an appropriate reset password.
- In the case of reset of parental code, the command to reset the code is for security reasons not permitted to originate from the receiver/decoder. Only the SMS, telephone and MINITEL or like can originate such a command. Hence in this

- 27 -

particular case the EMM(s) are broadcast only on air, never on the telephone line.

It will be understood from the above examples of different modes of operation of the OCS that the user can have direct access to the SAS, and in particular the OCS and AS, in that the Communications Servers are directly connected to the SAS, and in particular the OCS. This important feature is concerned with reducing the time for
5 the user to communicate his command to the SAS.

This feature is illustrated further with reference to Figure 12, from which it can be seen that the end user's Set-Top-Box, and in particular its receiver/decoder 2020, has the capability of communicating directly with the Communications Servers 3022
10 associated with the SAS 3002. Instead of the connection from the end user to the Communications Servers 3022 of the SAS 3002 being through the SMS 3004 the connection is directly to the SAS 3002.

In fact, as directly mentioned two direct connections are provided.

The first direct connection is by a voice link via a telephone 4001 and appropriate
15 telephone line (and/or by MINITEL or like connection where available) where the end users still have to input a series of voice commands or code numbers but time is saved compared with the communication being via the SMS 3004.

The second direct connection is from the receiver/decoder 2020 and the input of data is achieved automatically by the end user inserting his own daughter smartcard 3020
20 thus relieving the end user of the job of having to input the relevant data which in turn reduces the time taken and the likelihood of errors in making that input.

A further important feature which arises out of the above discussion is concerned with reducing the time taken for the resulting EMM to be transmitted to the end user in order to initiate viewing by the end user of the selected product.

25 In broad terms, and with reference to Figure 12, the feature is again achieved by

- 28 -

providing the end user's receiver/decoder 2020 with the capability of communicating directly with the Communications Servers 3022 associated with the SAS 3002.

As described earlier the integrated receiver/decoder 2020 is connected directly to the Communications Servers 3022 by the modemmed back channel 4002 so that
5 commands from the decoder 2020 are processed by the SAS 3002, messages generated (including EMMs) and then sent back directly to the decoder 2020 through the back channel 4002. A protocol is used in the communication between the CS 3022 and the receiver/decoder 2020 (as described later), so that the CS receive acknowledgement of receipt of the relevant EMM, thereby adding certainty to the procedure.

10 Thus, for example, in the case of a pre-book mode the SAS 3002 receives messages from the end user via the smartcard and decoder 2020 via its modem and via the telephone line 4002, requesting access to a specific event/product, and returns a suitable EMM via the telephone line 4002 and modem to the decoder 2020, the modem and decoder being preferably located together in a Set-Top-Box (STB). This
15 is thus achieved without having to transmit the EMM in the MPEG-2 data stream 2002 via the multiplexer and scrambler 2004, the uplink 2012, satellite 2014 and datalink 2016 to enable the end user to view the event/product. This can save considerably on time and bandwidth. Virtual certainty is provided that as soon as the subscriber has paid for his purchase the EMM will arrive at the receiver/decoder 2020.

20 In the third type of mode of operation of the OCS 3207 described above, the OCS deals with callback requests issued by the SAS. This is illustrated with reference to Figure 13. Typical callback requests have the purpose of ensuring that the receiver/decoder 2020 calls back the SAS via the modemmed back channel 4002 with the information that the SAS requires of the receiver /decoder.

25 As instructed by the Command Interface 3102, the subscription chain Message Generator 3106 generates and sends to the receiver/decoder 202 a callback EMM. This EMM is ciphered by the Ciphering Unit 3008 for security reasons. The EMM may contain the time/date at which the receiver/decoder should wake up and perform

- 29 -

a callback on its own, without being explicitly solicited; the EMM may also typically contain the phone numbers which the terminal must dial, the number of further attempts after unsuccessful calls and the delay between two calls.

When receiving the EMM, or at the specified time-date, the receiver/decoder connects
5 to the Communications Servers 3022. The OCS 3207 first identifies the caller, using
the AS 3202, and verifies certain details, such as smartcard operator and subscriber
details. The OCS then asks the smartcard 3020 to send various ciphered information
(such as the relevant session numbers, when the session was watched, how many times
10 the subscriber is allowed to view the session again, the way in which the session was
viewed, the number of remaining tokens, the number of prebooked sessions, etc). This
information is deciphered by the PPV chain Message Generator 3210, again using the
Ciphering Unit 3008. The OCS adds this information to a callback information
storage file 3214 for later processing and passing to the SMS 3004. The information
15 is ciphered for security reasons. The whole procedure is repeated until there is
nothing more to be read from the smartcard.

One particular preferred feature of the callback facility is that before reading the
smartcard (so just after the identification of the caller using the AS 3202 as described
above) a check is made by the SAS 3002 that the receiver/decoder is indeed a genuine
one rather than a pirated version or computer simulation. Such a check is carried out
20 in the following manner. The SAS generates a random number, which is received by
the receiver/decoder, ciphered, and then returned to the SAS. The SAS decipheres this
number. If the deciphering is successful and the original random number is retrieved,
it is concluded that the receiver/decoder is genuine, and the procedure continues.
Otherwise, the procedure is discontinued.

25 Other functions which may occur during the callback are erasure of obsolete sessions
on the smartcard, or filling of the wallet (this latter also being described later under
the section entitled "Smartcard").

Also as regards the Pay Per View Chain area 3200, description is now made of the

- 30 -

Communications Servers 3022. At the hardware level, these comprise in the preferred embodiment a DEC Four parallel processor machine. At the software architecture level, with reference to Figure 14, in many respects the Communications Servers are conventional. One particular divergence from conventional designs arises from the fact that the Servers must serve both receiver/decoders 2020 and voice communication with conventional telephones 4001, as well possibly as MINITEL or like systems.

It will be noted in passing that two Order Centralized Servers 3207 are shown in Figure 14 (as "OCS1" and "OCS2"). Naturally any desired number may be provided.

The Communication Servers include two main servers ("CS1" and "CS2") as well as a number of frontal servers ("Frontal 1" and "Frontal 2"); whilst two frontal servers are shown in the figure, typically 10 or 12 may be provided per main server. Indeed, although two main servers CS1 and CS2 and two frontal servers, Frontal 1 and Frontal 2, have been shown, any number could be used. Some redundancy is usually desirable.

CS1 and CS2 are coupled to OCS1 and OCS2 via high level TCP/IP links 3230, whilst CS1 and CS2 are coupled to Frontal 1 and Frontal 2 via further TCP/IP links 3232.

As illustrated, CS1 and CS2 comprise servers for "SENDER" (transmission), "RECVR" (reception), "VTX" (MINITEL, PRESTEL or the like), "VOX" (voice communication), and "TRM" (communication with the receiver/decoder). These are coupled to the "BUS" for communication of signals to the Frontal servers.

CS1 and CS2 communicate directly with the receiver/decoders 2020 via their modemmed back channels 4002 using the X25 public network common protocol. The relatively low-level protocol between the Communications Servers 3022 and the receiver/decoders 3020 is in one preferred embodiment based upon the V42 standard international CCITT protocol, which provides reliability by having error detection and data re-transmission facilities, and uses a checksum routine to check the integrity of

- 31 -

the re-transmission. An escape mechanism is also provided in order to prevent the transmission of disallowed characters.

On the other hand, voice telephone communication is carried out via the Frontal Communications Servers, each capable of picking up, say, 30 simultaneous voice
5 connections from the connection 3234 to the local telephone network via the high speed "T2" (E1) standard telephony ISDN lines.

Three particular functions of the software portion of the Communications Servers (which could of course alternatively be implemented fully in hardware) are firstly to convert the relatively low level protocol information received from the
10 receiver/decoder into the relatively high level protocol information output to the OCS, secondly to attenuate or control the number of simultaneous connections being made, and thirdly to provide several simultaneous channels without any mixing. In this last regard, the Communications Servers play the role of a form of multiplexer, with the interactions in a particular channel being defined by a given Session ID (identifier),
15 which is in fact used throughout the communication chain.

Finally as regards the Pay Per View Chain area 3200, and with reference again to Figure 5, the Server for Programme Broadcast (SPB) 3208 is coupled to one or more Programme Broadcasters 3250 (which would typically be located remotely from the SAS) to receive programme information. The SPB filters out for further use
20 information corresponding to PPV events (sessions).

A particularly important feature is that the filtered programme event information is passed by the SPB to the MG which in turn sends a directive (control command) to the ME to change the rate of cyclic emission of the EMMs in given circumstances; this is done by the ME finding all EMMs with the relevant session identifier and
25 changing the cycle rate allocated to such EMMs. This feature might be thought of as a dynamic allocation of bandwidth for specific EMMs. Cyclic EMM emission is discussed in more detail in the section below concerned with the EMM Injector.

- 32 -

The circumstances in which the cycle rate is changed are now described with reference to Figure 15, which demonstrates how cycle rate 3252 is raised a short while (say 10 minutes) before a particular PPV programme event until the end of the event from a slow cycle rate of say once every 30 minutes to a fast cycle rate of say once every 30 seconds to 1 minute in order to meet the anticipated extra user demand for PPV events at those times. In this way bandwidth can be allocated dynamically according to the anticipated user demand. This can assist in reducing the overall bandwidth requirement.

The cycle rate of other EMMs may also be varied. For example the cycle rate of subscription EMMs may be varied by the Multiplexer and Scrambler 2004 sending the appropriate bitrate directive.

EMM Injector

Concerning the EMM Injector 3300, details of the Message Emitters 3302 to 3308, forming part of the EMM Injector and acting as output means for the Message Generator, are now described with reference to Figure 16. Their function is take the EMMs and to pass them cyclically (in the manner of a carousel) via respective links 3314 and 3316 to the Software Multiplexers 3310 and 3312 and thence to the hardware multiplexers and scramblers 2004. In return the software multiplexers and scramblers 2004 generate a global bitrate directive to control the overall cycling rate of the EMMs; to do so, the MEs take into account various parameters such as the cycle time, the size of EMM, and so on. In the figure, EMM_X and EMM_Y are group EMMs for operators X and Y, whilst EMM_Z are other EMMs for either operator X or operator Y.

Further description proceeds for an exemplary one of the Message Emitters; it will be appreciated that the remaining MEs operate in similar fashion. The ME operates under control of directives from the MG, most notably transmission start and stop time and emission rate, as well as session number if the EMM is a PPV EMM. In relation to the emission rate, in the preferred embodiment the relevant directive may take one of five values from Very fast to Very slow. The numeric values are not specified in

- 33 -

the directive, but rather the ME maps the directive to an actual numeric value which is supplied by the relevant part of the SAS. In the preferred embodiment, the 5 emission rates are as follows:-

1. Very fast - every 30 seconds
- 5 2. Fast - every minute
3. Medium - every 15 minutes
4. Slow - every 30 minutes
5. Very slow - every 30 minutes

The ME has first and second databases 3320 and 3322. The first database is for those
10 EMMs which have not yet achieved their broadcast date; these are stored in a series of chronological files in the database. The second database is for EMMs for immediate broadcast. In the event of a system crash, the ME is arranged to have the ability to re-read the relevant stored file and perform correct broadcast. All the files stored in the databases are updated upon request from the MG, when the MG wishes
15 to maintain consistency between incoming directives and EMMs already sent to the ME. The EMMs actually being broadcast are also stored in Random Access Memory 3324.

A combination of the FIFOs 3162 and 3164 in the Message Generator and the databases 3320 and 3322 in the Message Emitter means that the two can operate in
20 standalone mode if the link 3166 between them is temporarily broken; the ME can still broadcast EMMs.

The Software Multiplexers (SMUX) 3310 and 3312 provide an interface between the MEs and the hardware multiplexers 2004. In the preferred embodiment, they each receive EMMs from two of the MEs, although in general there is no restriction on the
25 number of MEs that can be connected with one SMUX. The SMUXs concentrate the EMMs and then pass them according to the type of EMM to the appropriate hardware multiplexer. This is necessary because the hardware multiplexers take the different types of EMMs and place them at different places in the MPEG-2 stream. The

- 34 -

SMUX's also forward global bitrate directives from the hardware multiplexers to the MEs.

One particularly important feature of the ME is that it emits EMMs in random order. The reason for this is as follows. The Message Emitter has no ability to sense or control what it emits to the multiplexer. Hence it is possible that it may transmit two
5 EMMs which are to be received and decoded by the receiver/decoder 2020 back to back. In such circumstances, further, it is possible that if the EMMs are insufficiently separated the receiver/decoder and smartcard will be unable to sense and decode properly the second of the EMMs. Cyclically emitting the EMMs in random order
10 can solve this problem.

The manner in which randomization is achieved is now described with reference to Figure 17; in the preferred embodiment the necessary software logic is implemented in the ADA computer language. A particularly important part of the randomization is the correct storage of the EMMs in the databases 3320 and 3322 (which are used
15 for backup purposes) and in the RAM 3324. For a particular cycle rate and operator, the EMMs are stored in a two-dimensional array, by rank 3330 (going say from A to Z) and number in the rank 3332 (going from 0 to N). A third dimension is added by cycle rate 3334, so that there are as many two-dimensional arrays as there are cycle rates. In the preferred embodiment there are 256 ranks and typically 200 or 300
20 EMMs in each rank; there are 5 cycle rates. A final dimension to the array is added by the presence of different operators; there are as many three-dimensional arrays as there are operators. Storage of the data in this fashion can permit rapid retrieval in the event that the MG wants to delete a particular EMM.

Storage of the EMMs takes place according to the "hash" algorithm (otherwise known
25 as the "one-way hash function". This operates on a modulo approach, so that successive ranks are filled before a higher number in the rank is used, and the number of EMMs in each rank remains roughly constant. The example is considered of there being 256 ranks. When the MG sends the ME an EMM with identifier (ID) 1, the rank "1" is assigned to this EMM, and it takes the first number 3332 in the rank 3330.

- 35 -

The EMM with ID 2 is assigned the rank "2", and so on, up to the rank 256. The EMM with ID 257 is assigned the rank "1" again (based on the modulo function), and takes the second number in the first rank, and so on.

5 Retrieval of a specific EMM, for example when deletion of a specific EMM is requested by the MG, is effected by means of the inverse of the above. The hash algorithm is applied to the EMM ID to obtain the rank, after which the number in the rank is found.

10 The actual randomization occurs when the EMMs are, on a cyclical basis, retrieved from RAM 3324 using the randomization means 3340 which is implemented in the hardware and/or software of the Message Emitter. The retrieval is random, and again based on the hash algorithm. Firstly, a random number (in the above example initially in the range 1 to 256) is chosen, to yield the particular rank of interest. Secondly, a further random number is chosen to yield the particular number in the rank. The further random number is selected according to the total number of EMMs in a given rank. Once a given EMM has been selected and broadcast, it is moved to a second identical storage area in the RAM 3324, again using the hash function. Hence the first area diminishes in size as the EMMs are broadcast, to the extent that, once a complete rank has been used, this is deleted. Once the first storage area is completely empty, it is replaced by the second storage area before a new round of EMM broadcast, and vice versa.

15

20

In the above fashion, after two or three cycles of the EMMs, statistically the chances of any two EMMs destined for the same end user being transmitted back to back is negligible.

25 At regular intervals whilst the EMMs are being stored the computer 3050 computes the number of bytes in storage and from this computes the bitrate of emission given the global bitrate directive from the multiplexer and software multiplexer.

Reference was made above to the backup databases 3320 and 3322. These are in fact

- 36 -

in the preferred embodiment sequential file stores, which hold a backup version of what is in the RAM 3324. In the event of failure of the Message Emitter and subsequent restart, or more generally when the ME is being restarted for whatever reason, a link is made between the RAM and the databases, over which the stored
5 EMMs are uploaded to RAM. In this way, the risk of losing EMMs in the event of failure can be removed.

Similar storage of PPV EMMs occurs to that described above in relation to subscription EMMs, with the rank typically corresponding to a given operator and the number in the rank corresponding to the session number.

10 Smartcard

A daughter, or "subscriber", smartcard 3020 is schematically shown in Figure 18 and comprises an 8 bit microprocessor 110, such as a Motorola 6805 microprocessor, having an input/output bus coupled to a standard array of contacts 120 which in use are connected to a corresponding array of contacts in the card reader of the
15 receiver/decoder 2020, the card reader being of conventional design. The microprocessor 110 is also provided with bus connections to preferably masked ROM 130, RAM 140 and EEPROM 150. The smartcard complies with the ISO 7816-1, 7816-2 and 7816-3 standard protocols which determine certain physical parameters of the smartcard, the positions of the contacts on the chip and certain communications
20 between the external system (and particularly the receiver/decoder 2020) and the smartcard respectively and which will therefore not be further described here. One function of the microprocessor 110 is to manage the memory in the smartcard, as now described.

The EEPROM 150 contains certain dynamically-created operator zones 154, 155, 156
25 and dynamically-created data zones which will now be described with reference to Figure 19.

Referring to Figure 19, EEPROM 150 comprises a permanent "card ID" (or manufacturer) zone 151 of 8 bytes which contains a permanent subscriber smartcard

- 37 -

identifier set by the manufacturer of the smartcard 3020.

When the smartcard is reset, the microprocessor 110 issues a signal to receiver/decoder 2020, the signal comprising an identifier of the conditional access system used by the smartcard and data generated from data stored in the smartcard, including the card ID. This signal is stored by the receiver/decoder 2020, which subsequently utilises the stored signal to check whether the smartcard is compatible with the conditional access system used by the receiver/decoder 2020.

The EEPROM 150 also contains a permanent "random number generator" zone 152 which contains a program for generating pseudo-random numbers. Such random numbers are used for diversifying transaction output signals generated by the smartcard 3020 and sent back to the broadcaster.

Below the random number generator zone 152 a permanent "management" zone 153 of 144 bytes is provided. The permanent management zone 153 is a specific operator zone utilised by a program in the ROM 130 in the dynamic creation (and removal) of zones 154, 155, 156... as described below. The permanent management zone 153 contains data relating to the rights of the smartcard to create or remove zones.

The program for dynamically creating and removing zones is responsive to specific zone creation (or removal) EMMs which are transmitted by the SAS 3002 and received by the receiver/decoder 2020 and passed to the subscriber smartcard 3020. In order to create the EMMs the operator requires specific keys dedicated to the management zone. This prevents one operator from deleting zones relating to another operator.

Below the management zone 153 is a series of "operator ID" zones 154, 155, 156 for operators 1, 2 N respectively. Normally at least one operator ID zone will be preloaded into the EEPROM of the subscriber smartcard 3020 so that the end user can decrypt programmes broadcast by that operator. However further operator ID zones can subsequently be dynamically created using the management zone 153 in response

- 38 -

to a transaction output signal generated via his smartcard 3020 by the end user (subscriber), as will subsequently be described.

Each operator zone 154, 155, 156 contains the identifier of the group to which the smartcard 3020 belongs, and the position of the smartcard within the group. This data
5 enables the smartcard (along with the other smartcards in its group) to be responsive to a broadcast "group" subscription EMM having that group's address (but not the smartcard's position in the group) as well as to an "individual" (or commercial offers subscription) EMM addressed only to that smartcard within the group. There can be 256 member smartcards of each such group and this feature therefore reduces
10 significantly the bandwidth required for broadcasting EMMs.

In order to reduce further the bandwidth required for broadcasting "group" subscription EMMs, the group data in each operator zone 154, 155, 156 and all similar zones in the EEPROM of smartcard 3020 and the other daughter smartcards is continually updated to enable a particular smartcard to change its position in each group to fill
15 any holes created by e.g. deletion of a member of the group. The holes are filled by the SAS 3002 as in the STM server 3104 there is a list of such holes.

In this manner fragmentation is reduced and each group's membership is maintained at or near the maximum of 256 members.

Each operator zone 154, 155, 156 is associated with one or more "operator data
20 objects" stored in the EEPROM 150. As shown in Figure 19, a series of dynamically created "operator data" objects 157-165 are located below the operator ID zones. Each of these objects is labelled with:

- a) an "identifier" 1, 2, 3 N corresponding to its associated operator 1, 2, 3 ... N as shown in its left hand section in Figure 19;
- 25 b) an "ID" indicating the type of object; and
- c) a "data" zone reserved for data, as shown in the right hand section of each relevant operator object in Figure 19. It should be understood that each operator is associated with a similar set of data objects so that the following description of the

- 39 -

types of data in the data objects of operator 1 is also applicable to the data objects of all the other operators. Also it will be noted that the data objects are located in contiguous physical regions of the EEPROM and that their order is immaterial.

Deletion of a data object creates a "hole" 166 in the smartcard, that is, the number of bytes that the deleted objects had previously occupied are not immediately occupied. The thus "freed" number of bytes, or "hole" are labelled with:

- a) an "identifier" 0; and
- b) an "ID" indicating that the bytes are free to receive an object.

The next data object created fills the hole, as identified by the identifier 0. In this manner the limited memory capacity (4 kilobytes) of the EEPROM 150 is efficiently utilised.

Turning now to the set of data objects associated with each operator, examples of the data objects are now described.

Data object 157 contains an EMM key used for decrypting encrypted EMM's received by the receiver/decoder 2020. This EMM key is permanently stored in the data object 157. This data object 157 may be created prior to distribution of the smartcard 3020, and/or may be created dynamically when creating a new operator zone (as described above).

Data object 159 contains ECM keys which are sent by the associated operator (in this case operator 1) to enable the end user to decrypt the particular "bouquet" of programs to which he has subscribed. New ECM keys are sent typically every month, along with a group subscription (renewal) EMM which renews the end user's overall right to view the broadcast from (in this case) operator 1. The use of separate EMM and ECM keys enables viewing rights to be purchased in different ways (in this embodiment by subscription and individually (Pay Per View)) and also increases security. The Pay Per View (PPV) mode will be described subsequently.

- 40 -

Since new ECM keys are sent periodically, it is essential to prevent a user from using old ECM keys, for example by switching off the receiver/decoder or re-setting a clock to prevent expiry of an old ECM key so that a timer in the receiver/decoder 2020 could be overridden. Accordingly operator zone 154 comprises an area (typically
5 having a size of 2 bytes) containing an obsolescence date of the ECM keys. The smartcard 3020 is arranged to compare this date with the current date which is contained in received ECMs and to prevent decryption if the current date is later than the obsolescence date. The obsolescence date is transmitted via EMMs, as described above.

10 Data object 161 contains a 64 bit subscription bitmap which is an exact representation of the broadcast operator's programs to which the subscriber has subscribed. Every bit represents a program and is set to "1" if it is subscribed to and "0" if it is not.

Data object 163 contains a quantity of tokens which can be used by the consumer in PPV mode to buy viewing rights to an imminent broadcast e.g. in response to a free
15 preview or other advertisement. Data object 163 also contains a limit value, which may be set to e.g. a negative value to allow credit to the consumer. Tokens can be purchased e.g. by credit and via the modemed back channel 4002, or by using a voice server in combination with a credit card, for example. A particular event can be charged as one token or a number of tokens.

20 Data object 165 contains a description of a PPV event, as shown with reference to table 167 of Figure 20.

The PPV event description 167 contains a "session ID" 168 identifying the viewing session (corresponding to the program and the time and date of broadcasting) a
25 "session mode" 169 indicating how the viewing right is being purchased (e.g. in pre-book mode), a "session index" 170 and a "session view" 171.

In respect of receiving a programme in PPV mode, the receiver decoder 2020 determines whether the programme is one sold in PPV mode. If so, the decoder 2020

- 41 -

checks, using the items stored in the PPV event description 167 whether the session ID for the programme is stored therein. If the session ID is stored therein, the control word is extracted from the ECM.

If the session ID is not stored therein, by means of a specific application the receiver/decoder 2020 displays a message to the end user indicating that he has the
5 right to view the session at a cost of, say, 25 tokens, as read from the ECM or to connect to the communications servers 3022 to purchase the event. Using the tokens, if the end user answers "yes" (by means of remote controller 2026 (see Figure 2)) the decoder 2020 sends the ECM to the smartcard, the smartcard decreases the wallet of
10 the smartcard 3020 by 25 tokens, writes the session ID 168, the session mode 169, the session index 170 and the session view 171 in the PPV event description 167 and extracts and deciphers the control word from the ECM.

In the "pre-book" mode, an EMM will be passed to the smartcard 3020 so that the smartcard will write the session ID 168, the session mode 169, the session index 170
15 and the session view 171 in the PPV event description 167 using the EMM.

The session index 170 can be set to differentiate one broadcast from the other. This feature permits authorization to be given for a subset of broadcasts, for example, 3 times out of 5 broadcasts. As soon as an ECM with a session index different from the current session index 170 stored in the PPV event description 167 is passed to the
20 smartcard, the number of the session view 171 is decreased by one. When the session view reaches zero, the smartcard will refuse to decipher an ECM with a different session index to the current session index.

The initial value of the session view depends only on the way in which the broadcast supplier wishes to define the event to which it relates; the session view for a
25 respective event may take any value.

The microprocessor 110 in the smartcard implements a counting and a comparison program to detect when the limit to the number of viewings of a particular program

- 42 -

has been reached.

All of the session ID 168, the session mode 169, the session index 170 and the session view 171 in the PPV event description 167 may be extracted from the smartcard using the "call-back" procedure as described previously.

5 Each receiver/decoder 2020 contains an identifier which may either identify uniquely that receiver/decoder or identify its manufacturer or may classify it in some other way in order to enable it to work only with a particular individual smartcard, a particular class of smartcards made by the same or a corresponding manufacturer or any other class of smartcards which are intended for use with that class of receiver/decoders
10 exclusively.

In this manner the receiver/decoders 2020 which have been supplied by one broadcast supplier to the consumer are protected against the use of non-authorized daughter smartcards 3020.

15 Additionally or alternatively to this first "handshake" between the smartcard and the receiver, the EEPROM of the smartcard 3020 could contain a field or bitmap describing the categories of receiver/decoders 2020 with which it can function. These could be specified either during the manufacture of the smartcard 3020 or by a specific EMM.

20 The bitmap stored in the smartcard 3020 typically comprises a list of up to 80 receiver/decoders, each identified with a corresponding receiver/decoder ID with which the smartcard may be used. Associated with each receiver/decoder is a level "1" or "0" indicating whether the smartcard may be used with the receiver/decoder or not, respectively. A program in the memory 2024 of the receiver/decoder searches for the identifier of the receiver/decoder in the bitmap stored in the smartcard. If the
25 identifier is found, and the value associated with the identifier is "1", then the smartcard is "enabled"; if not, then the smartcard will not function with that receiver/decoder.

- 43 -

In addition, if, typically because of an agreement between operators, it is desired to authorize the use of other smartcards in a particular receiver/decoder, specific EMMs will be sent to those smartcards to change their bitmap via the transponder 2014.

Each broadcast supplier may differentiate his subscribers according to certain predetermined criteria. For example, a number of subscribers may be classed as "VIPs". Accordingly, each broadcast supplier may divide his subscribers into a plurality of subsets, each subset comprising any number of subscribers.

The subset to which a particular subscriber belongs is set in the SMS 3004. In turn, the SAS 3002 transmits an EMM to the subscriber which writes information (typically of length 1 byte) concerning the subset to which the subscriber belongs into the relevant operator data zone, say 154, of the EEPROM of the smartcard. In turn, as events are broadcast by the broadcast supplier, an ECM, typically of 256 bits, is transmitted with the event and indicating which of the subsets of subscribers may view the event. If, according to the information stored in the operator zone, the subscriber does not have the right to view the event, as determined by the ECM, programme viewing is denied.

This facility may be used, for example, to switch off all of a given operator's smartcards in a particular geographical region during the transmission of a particular program, in particular a program relating to a sports fixture taking place in that geographical region. In this manner football clubs and other sport bodies can sell broadcasting rights outside their locality whilst preventing local supporters from viewing the fixture on television. In this manner the local supporters are encouraged to buy tickets and attend the fixture.

Each of the features associated with zones 151 to 172 is considered to be a separate invention independent of the dynamic creation of zones.

It will be understood that the present invention has been described above purely by way of example, and modifications of detail can be made within the scope of the

- 44 -

invention.

Each feature disclosed in the description, and (where appropriate) the claims and drawings may be provided independently or in any appropriate combination.

In the aforementioned preferred embodiments, certain features of the present invention
5 have been implemented using computer software. However, it will of course be clear to the skilled man that any of these features may be implemented using hardware. Furthermore, it will be readily understood that the functions performed by the hardware, the computer software, and such like are performed on or using electrical and like signals.

10 Cross reference is made to our co-pending applications, all bearing the same filing date, and entitled Signal Generation and Broadcasting (Attorney Reference no. PC/ASB/19707), Smartcard for use with a Receiver of Encrypted Broadcast Signals, and Receiver (Attorney Reference No. PC/ASB/19708), Broadcast and Reception
15 System and Conditional Access System therefor (Attorney Reference No. PC/ASB/19710), Downloading a Computer File from a Transmitter via a Receiver/Decoder to a Computer (Attorney Reference No. PC/ASB/19711), Transmission and Reception of Television Programmes and Other Data (Attorney Reference No. PC/ASB/19712), Downloading Data (Attorney Reference No. PC/ASB/19713), Computer Memory Organisation (Attorney Reference No.
20 PC/ASB/19714), Television or Radio Control System Development (Attorney Reference No. PC/ASB/19715), Extracting Data Sections from a Transmitted Data Stream (Attorney Reference No. PC/ASB/19716), Access Control System (Attorney Reference No. PC/ASB/19717), Data Processing System (Attorney Reference No. PC/ASB/19718), and Broadcast and Reception System, and Receiver/Decoder and
25 Remote Controller therefor (Attorney Reference No. PC/ASB/19720). The disclosures of these documents are incorporated herein by reference. The list of applications includes the present application.

- 45 -

CLAIMS

1. A conditional access system comprising:
means for generating a plurality of messages; and
means for receiving the messages, said receiving means being adapted to
5 communicate with said generating means via a communications server connected
directly to said generating means.
2. A conditional access system according to Claim 1, wherein said message is an
entitlement message for transmission to the receiving means, said generating means
being adapted to generate entitlement messages in response to data received from said
10 receiving means.
3. A conditional access system according to Claim 1 or 2, wherein said generating
means is arranged to transmit a message as a packet of digital data to said receiving
means either via said communications server or via a satellite transponder.
4. A conditional access system according to any preceding claim, wherein said
15 receiving means is connectable to said communications server via a modem and
telephone link.
5. A conditional access system for affording conditional access to subscribers,
comprising:
a subscriber management system;
20 a subscriber authorization system coupled to the subscriber management
system; and
a communications server; said server being connected directly to the subscriber
authorization system.
6. A conditional access system according to Claim 5, further comprising a
25 receiver/decoder for the subscriber, the receiver/decoder being connectable to said
communications server, and hence to said subscriber authorization system, via a

- 46 -

modem and telephone link.

7. A broadcast and reception system including a conditional access system according to any preceding claim.

8. A broadcast and reception system comprising:

5 means for generating a plurality of entitlement messages relating to broadcast programs;

means for receiving said messages from said generating means; and

10 means for connecting the receiving means to the generating means to receive said messages, said connecting means being capable of effecting a dedicated connection between the receiving means and the generating means.

9. A broadcast and reception system comprising:

means for generating a plurality of entitlement messages relating to broadcast programs;

means for receiving said messages from said generating means via a modem;

15 and

means for connecting said modem to said generating means and said receiving means.

10. A broadcast and reception system according to Claim 9, wherein said generating means is connected to said modem via a communications server.

20 11. A broadcast and reception system according to Claim 9 or 10, wherein said receiving means is adapted to communicate with said generating means via said modem and connecting means.

25 12. A broadcast and reception system according to Claim 11, wherein said receiving means comprises means for reading a smartcard insertable thereinto by an end user, the smartcard having stored therein data to initiate automatically the transmission of a message from said receiving means to said generating means upon

- 47 -

insertion of the smartcard by the end user.

13. A broadcast and reception system according to Claim 11 or 12, further comprising a voice link to enable the end user of the broadcast and reception system to communicate with the generating means.

5 14. A broadcast and reception system according to any of Claims 8 to 13, wherein said receiving means comprises a receiver/decoder comprising means for receiving a compressed MPEG-type signal, means for decoding the received signal to provide a television signal and means for supplying the television signal to a television.

10 15. A broadcast and reception system, comprising, at the broadcast end:
a broadcast system including means for broadcasting a callback request;
and at the reception end:
a receiver including means for calling back the broadcast system in response to the callback request.

15 16. A system according to Claim 15, wherein the means for calling back the broadcast system includes a modem connectable to a telephone system.

17. A system according to Claim 15 or 16, wherein the means for calling back the broadcast system is arranged to transfer to the broadcast system information concerning the receiver.

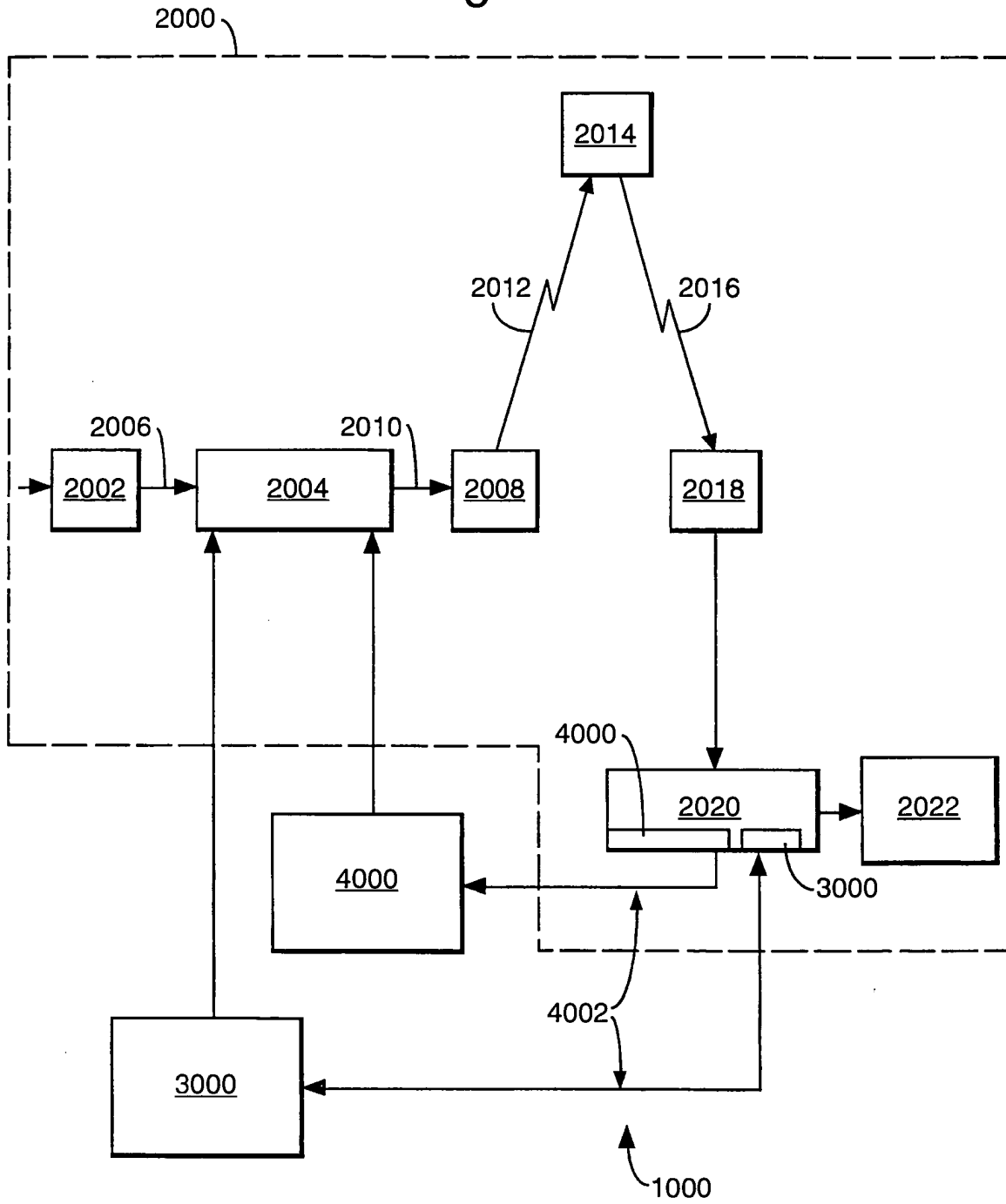
20 18. A system according to Claim 17, wherein the broadcast system includes means for storing the information.

19. A system according to any of Claims 15 to 18, wherein the broadcast means is arranged to broadcast a callback request which includes a command that the callback be made at a given time, and the means for calling back the broadcast system is arranged to respond to said command.

- 48 -

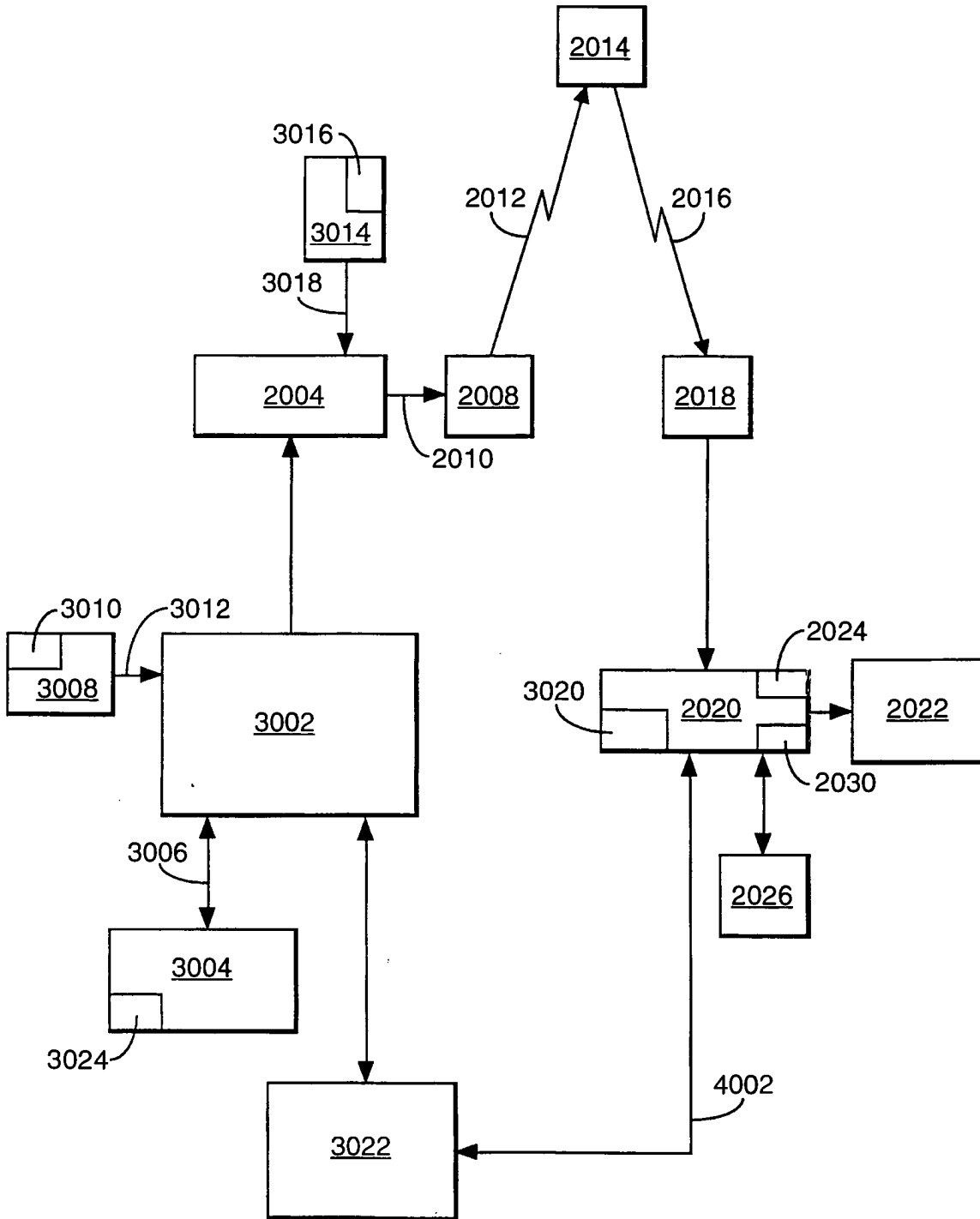
20. A system according to any of Claims 15 to 19, wherein the broadcasting means is arranged to broadcast as the callback request one or more entitlement messages for broadcast.
21. A system according to any of Claims 15 to 20, wherein the broadcast system includes means for generating a check message and passing this to the receiver, the receiver includes means for encrypting the check message and passing this to the broadcast system, and the broadcast system further includes means for decrypting the check message received from the receiver and comparing this with the original check message.
22. A conditional access system or a broadcast and reception system substantially as herein described with reference to and as illustrated in the accompanying drawings, and especially Figures 12, 13 or 14 thereof.

Fig. 1.



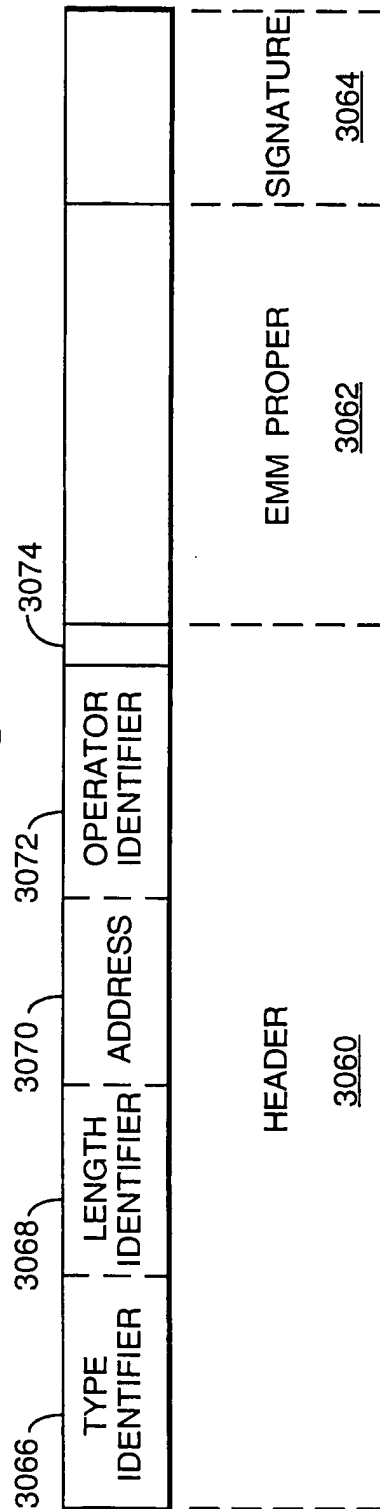
SUBSTITUTE SHEET (RULE 26)

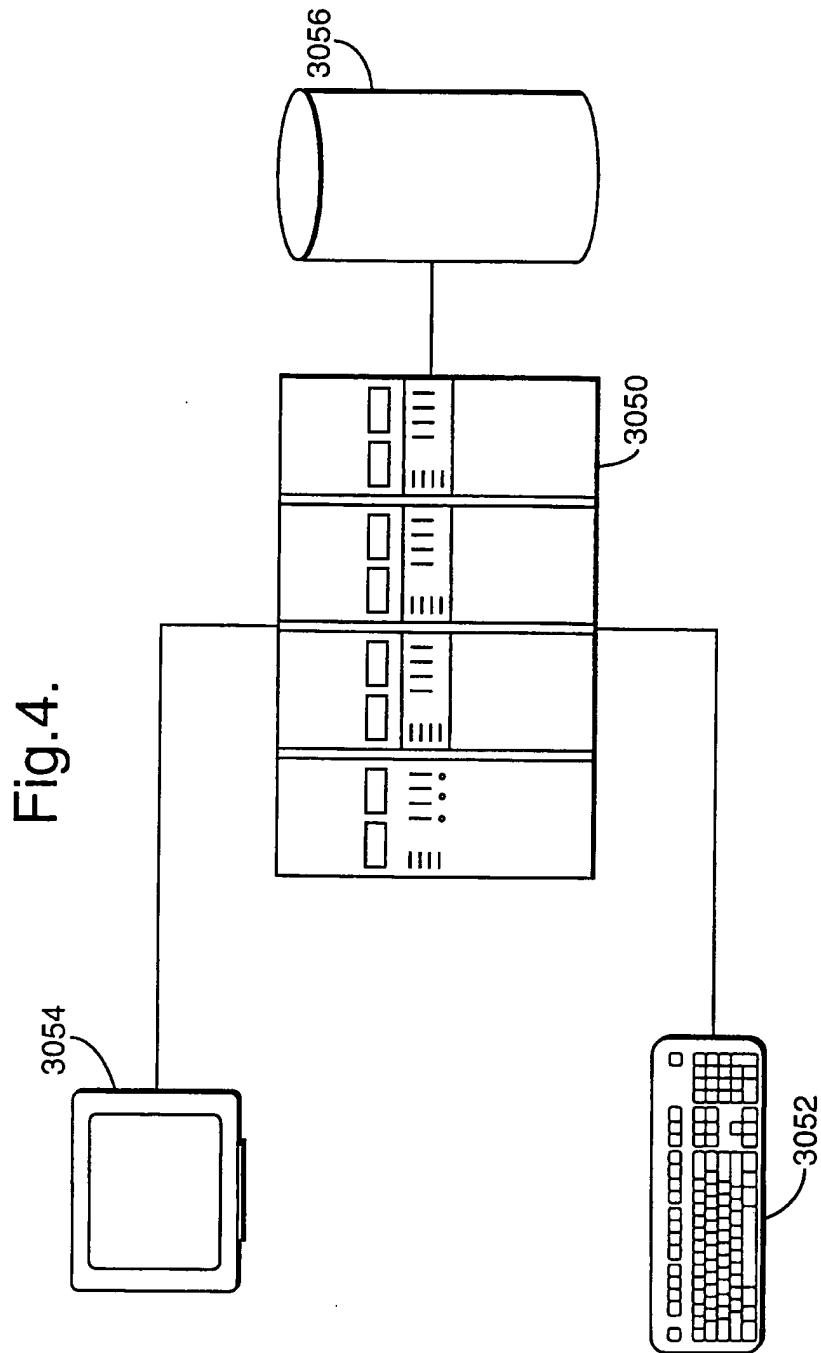
Fig.2.



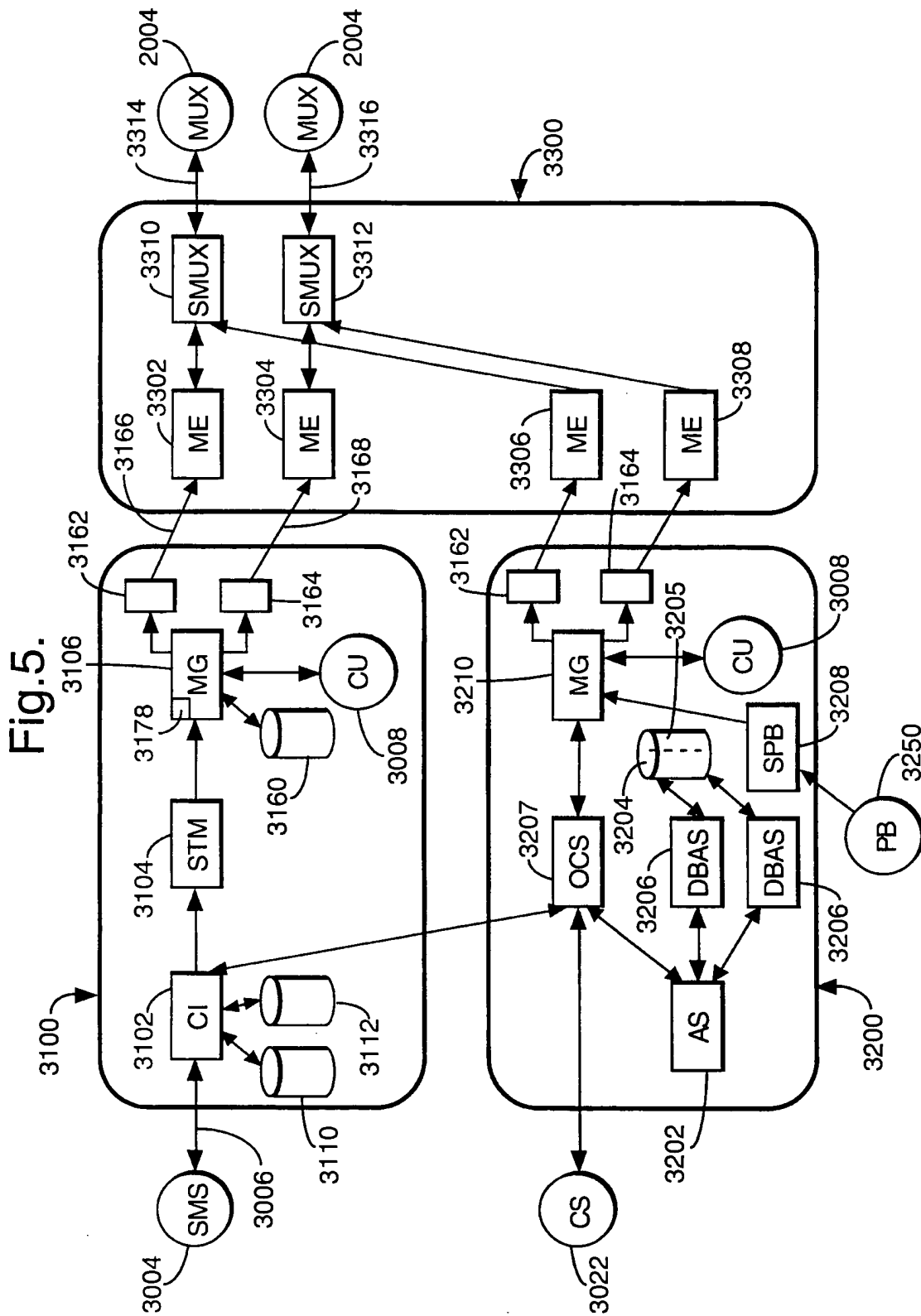
SUBSTITUTE SHEET (RULE 26)

Fig.3.

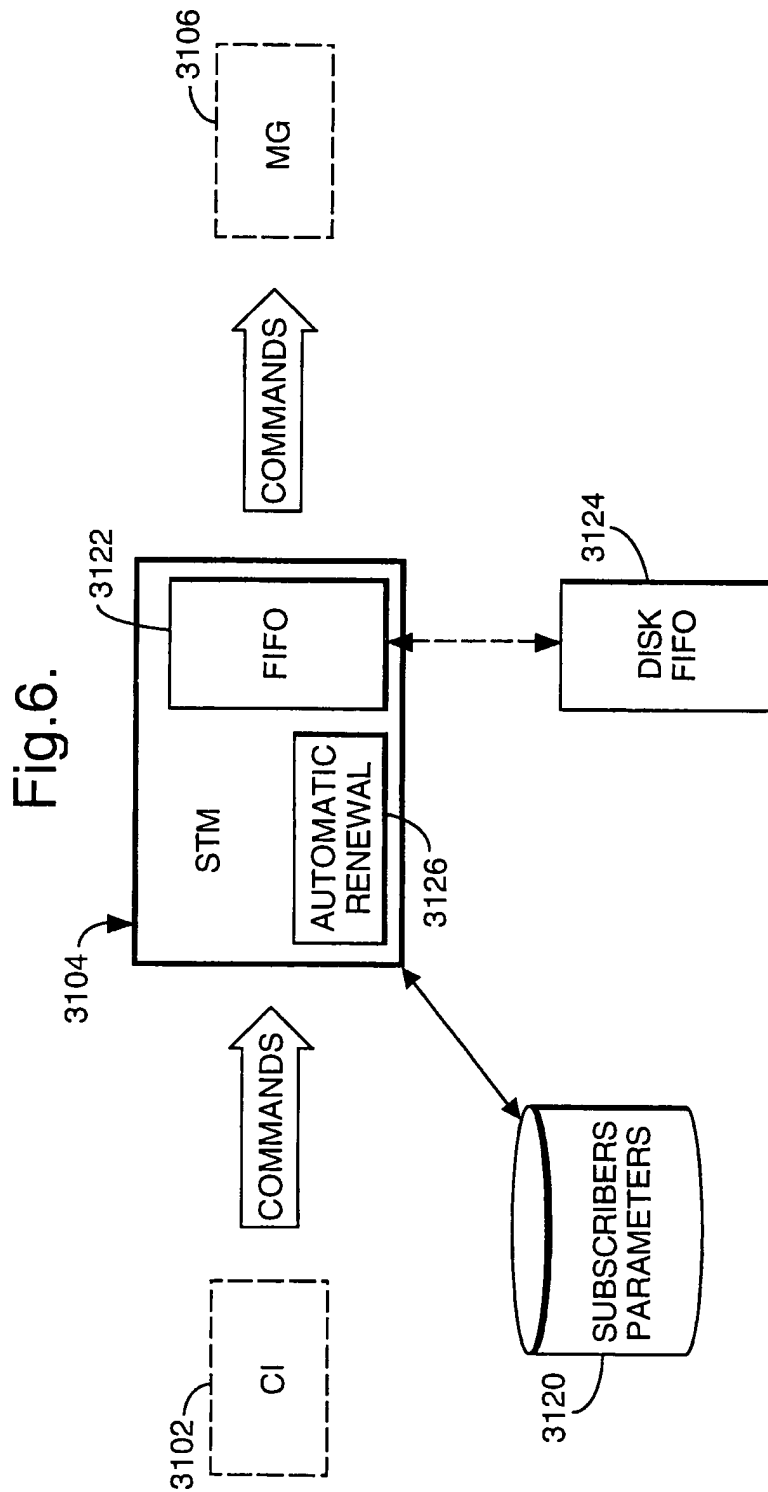




SUBSTITUTE SHEET (RULE 26)

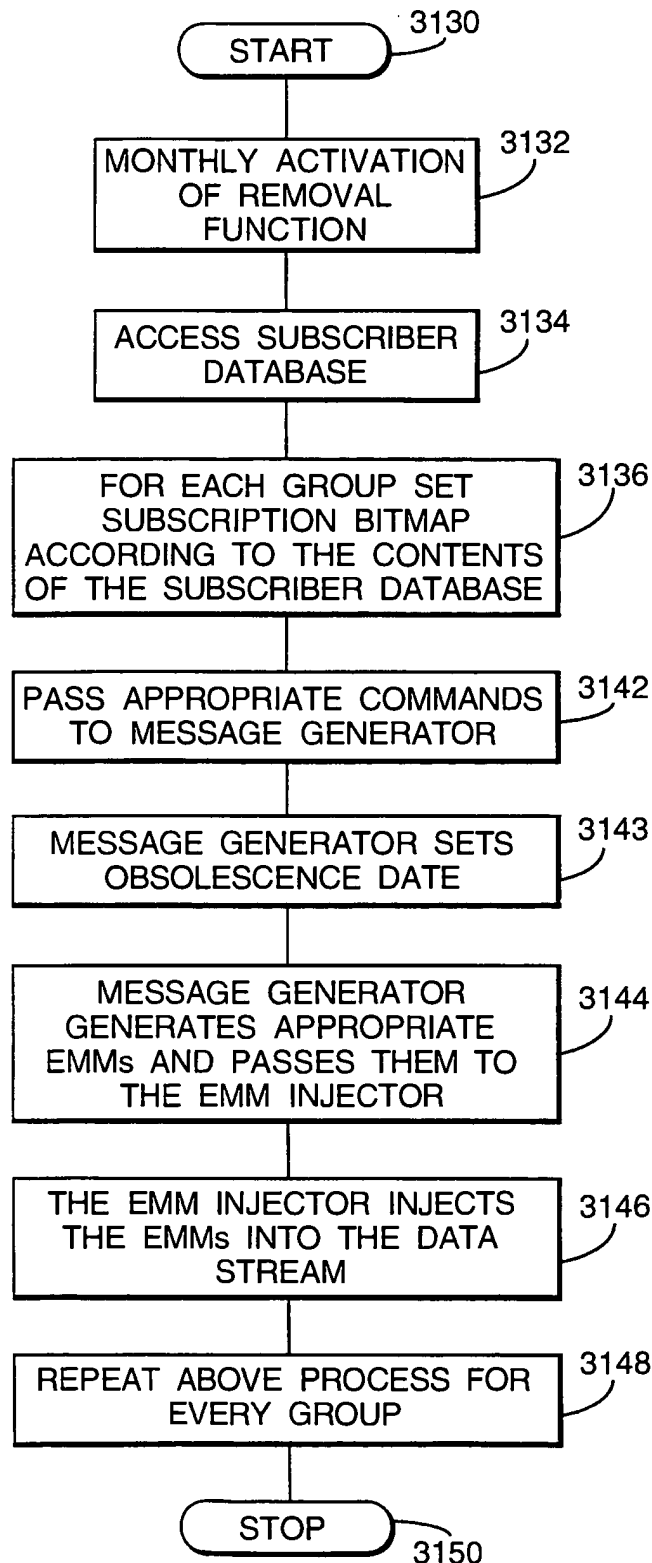


SUBSTITUTE SHEET (RULE 26)



7/17

Fig.7.



SUBSTITUTE SHEET (RULE 26)

Fig.8.

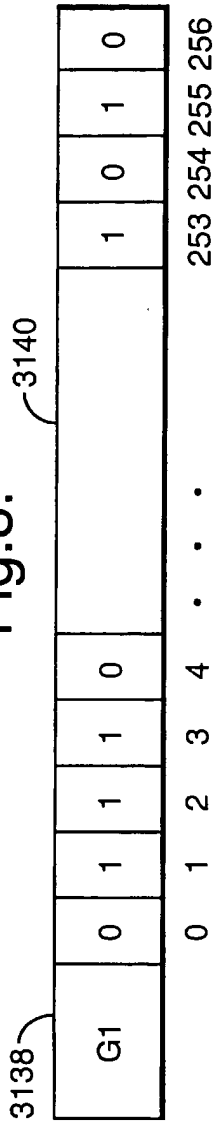


Fig.9.

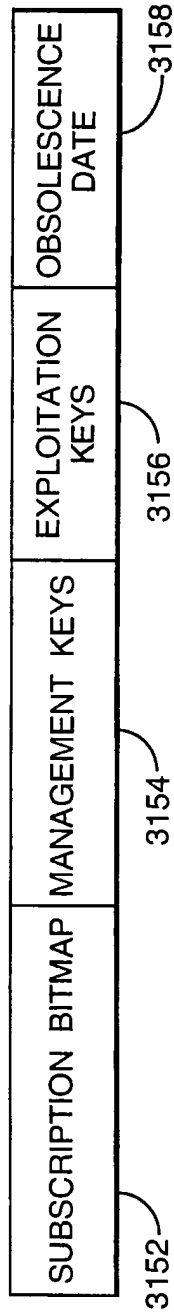
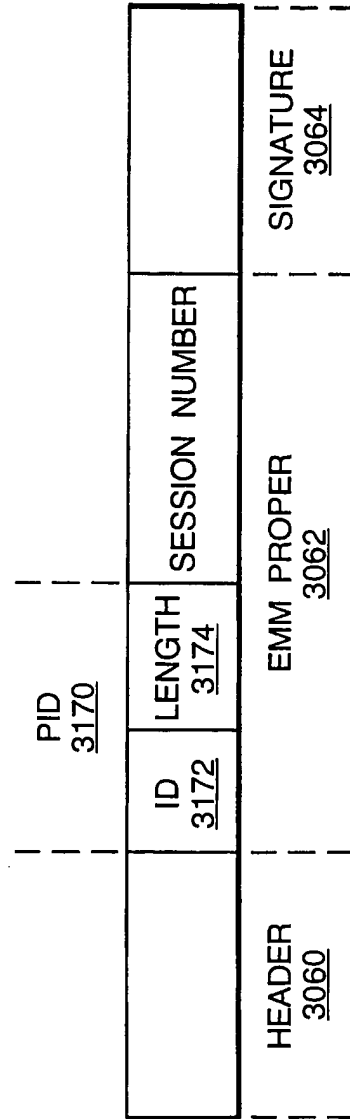
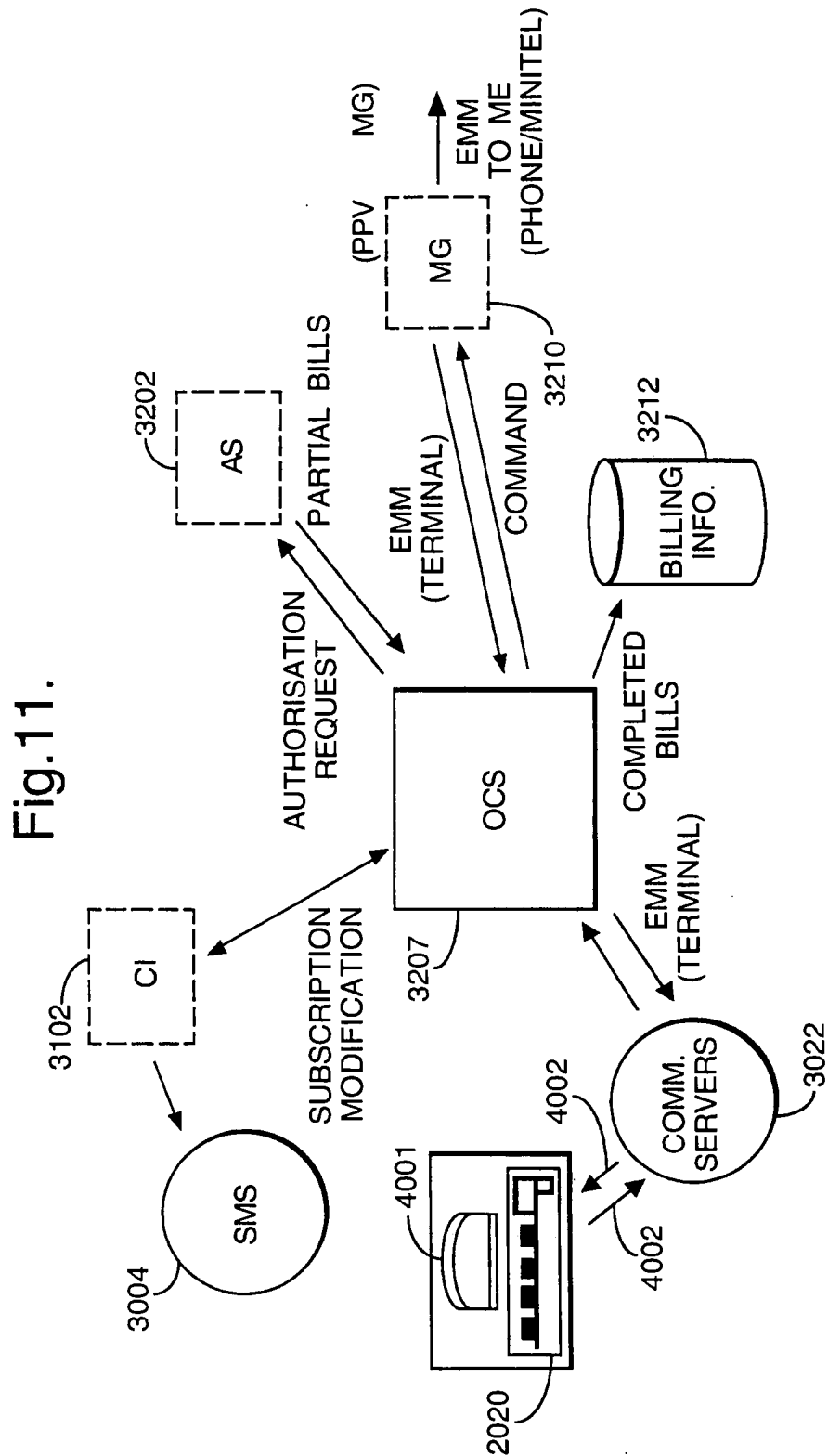


Fig.10.

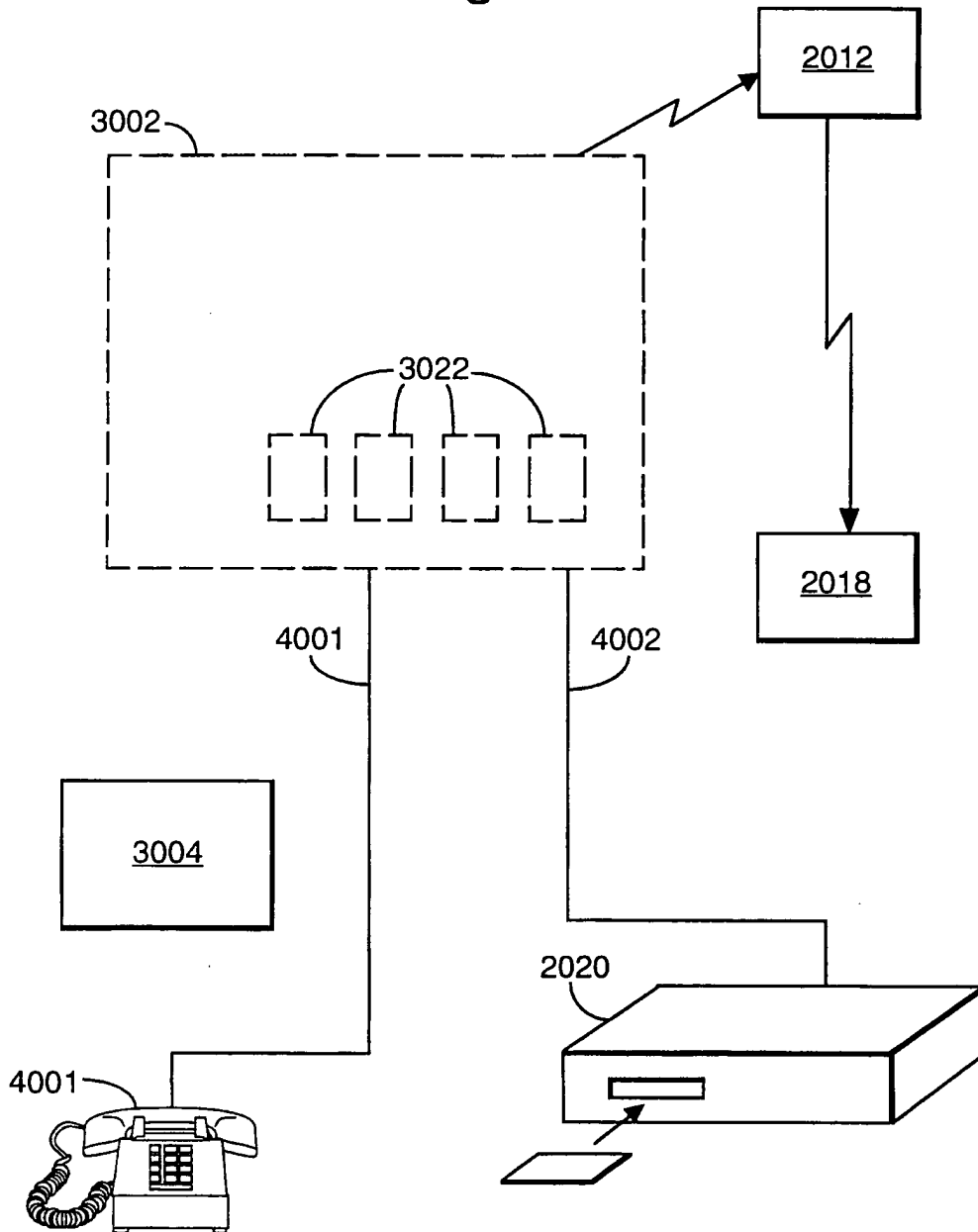


9/17



SUBSTITUTE SHEET (RULE 26)

Fig.12.



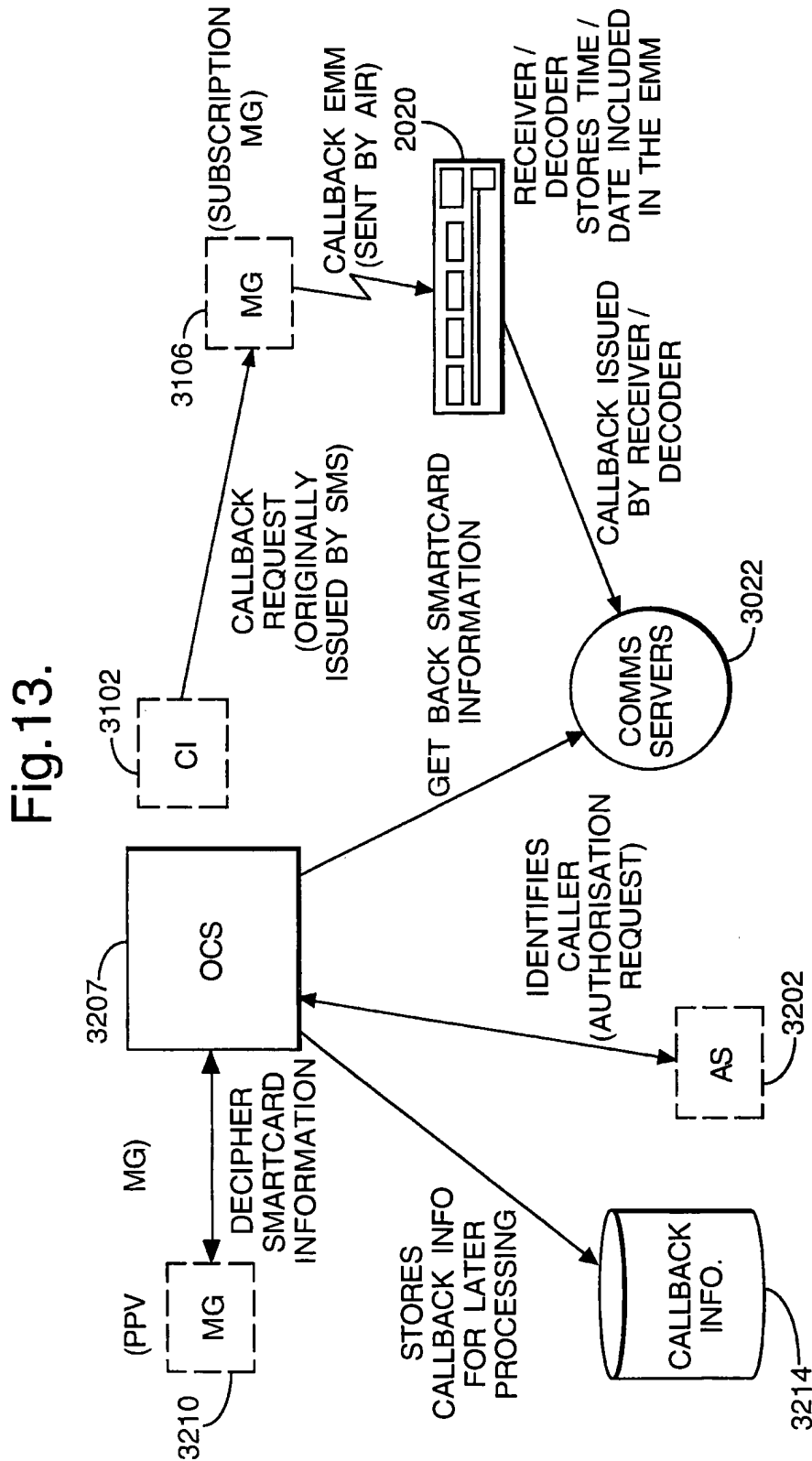
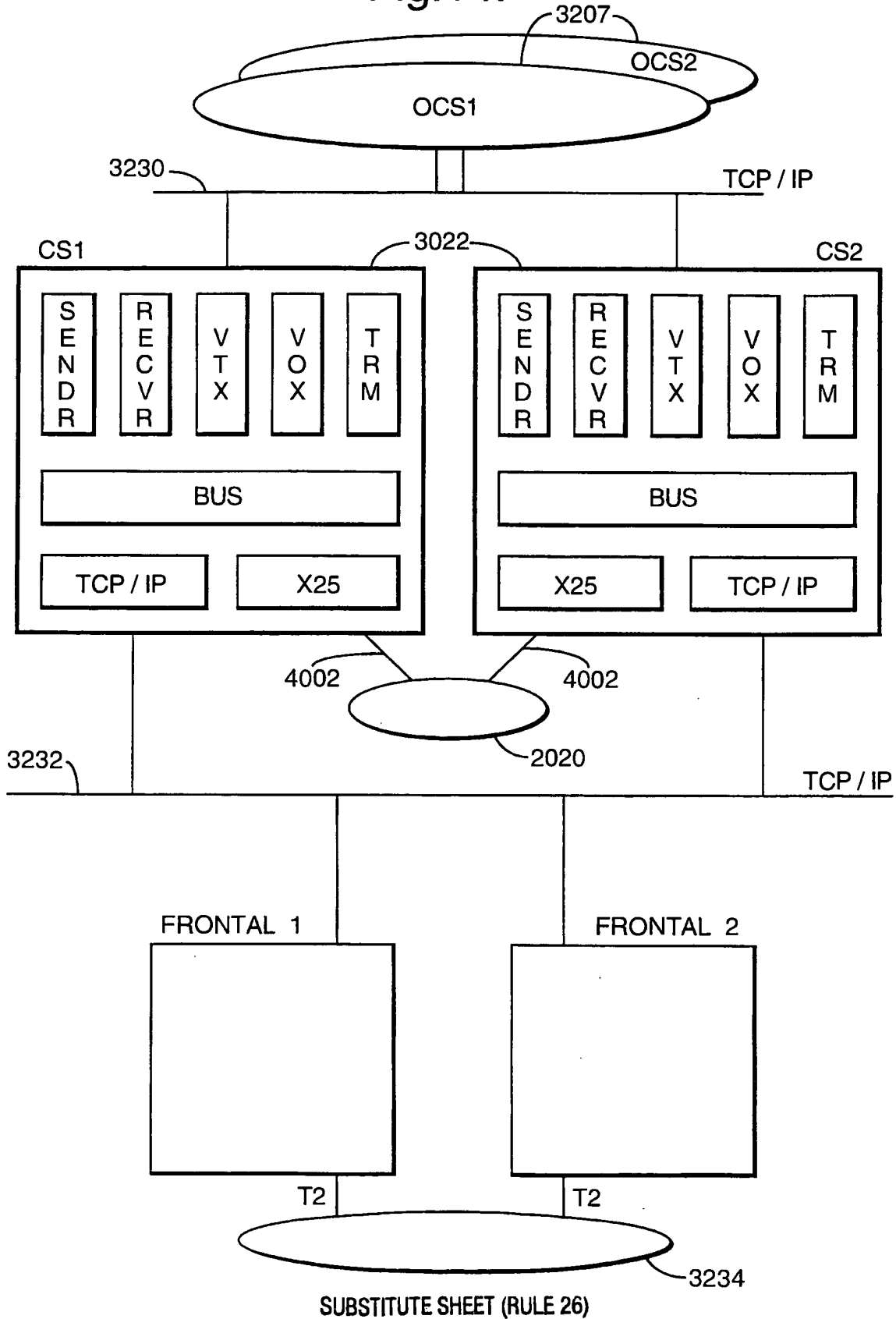


Fig.13.

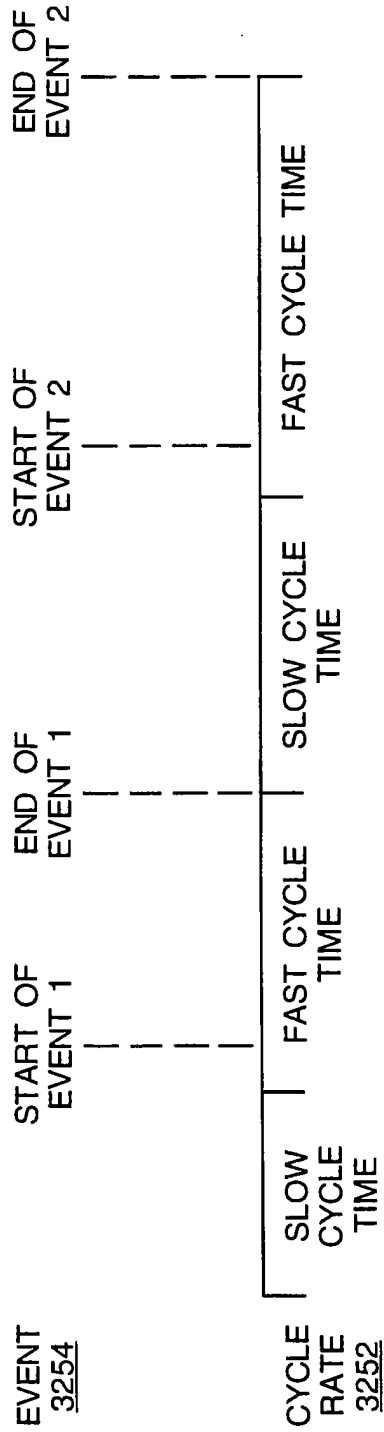
12/17

Fig.14.



SUBSTITUTE SHEET (RULE 26)

Fig.15.



SUBSTITUTE SHEET (RULE 26)

Fig.16.

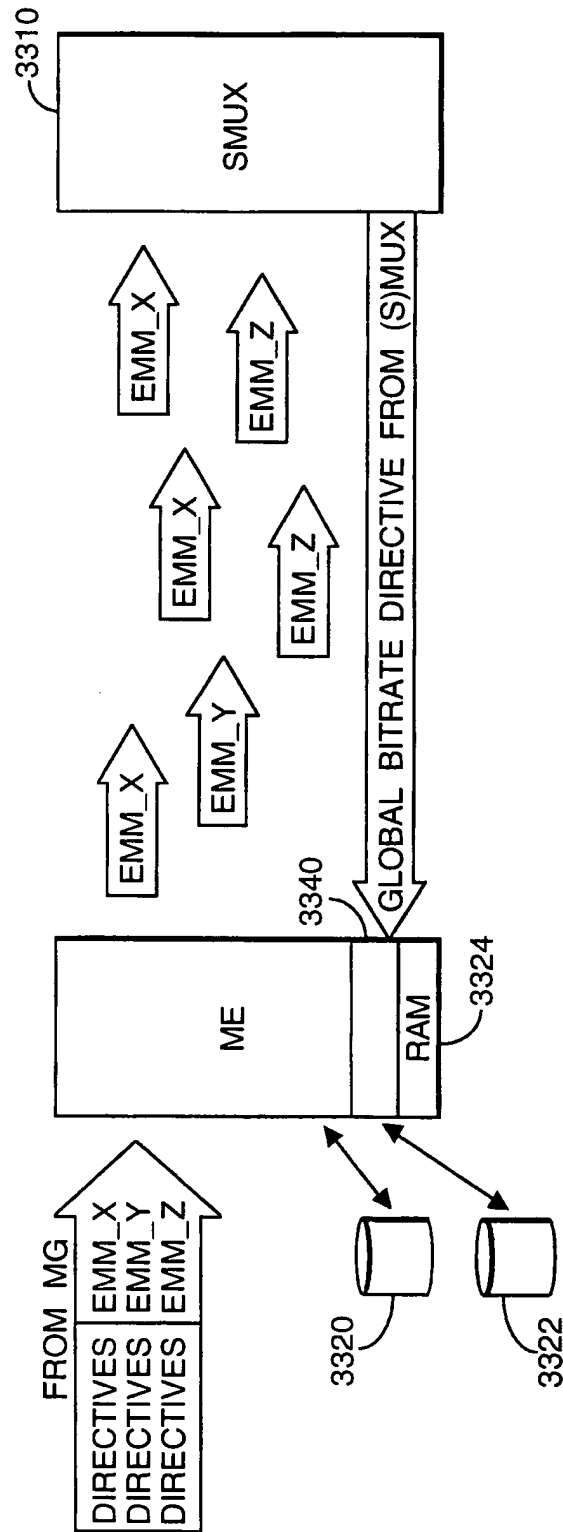


Fig.17.

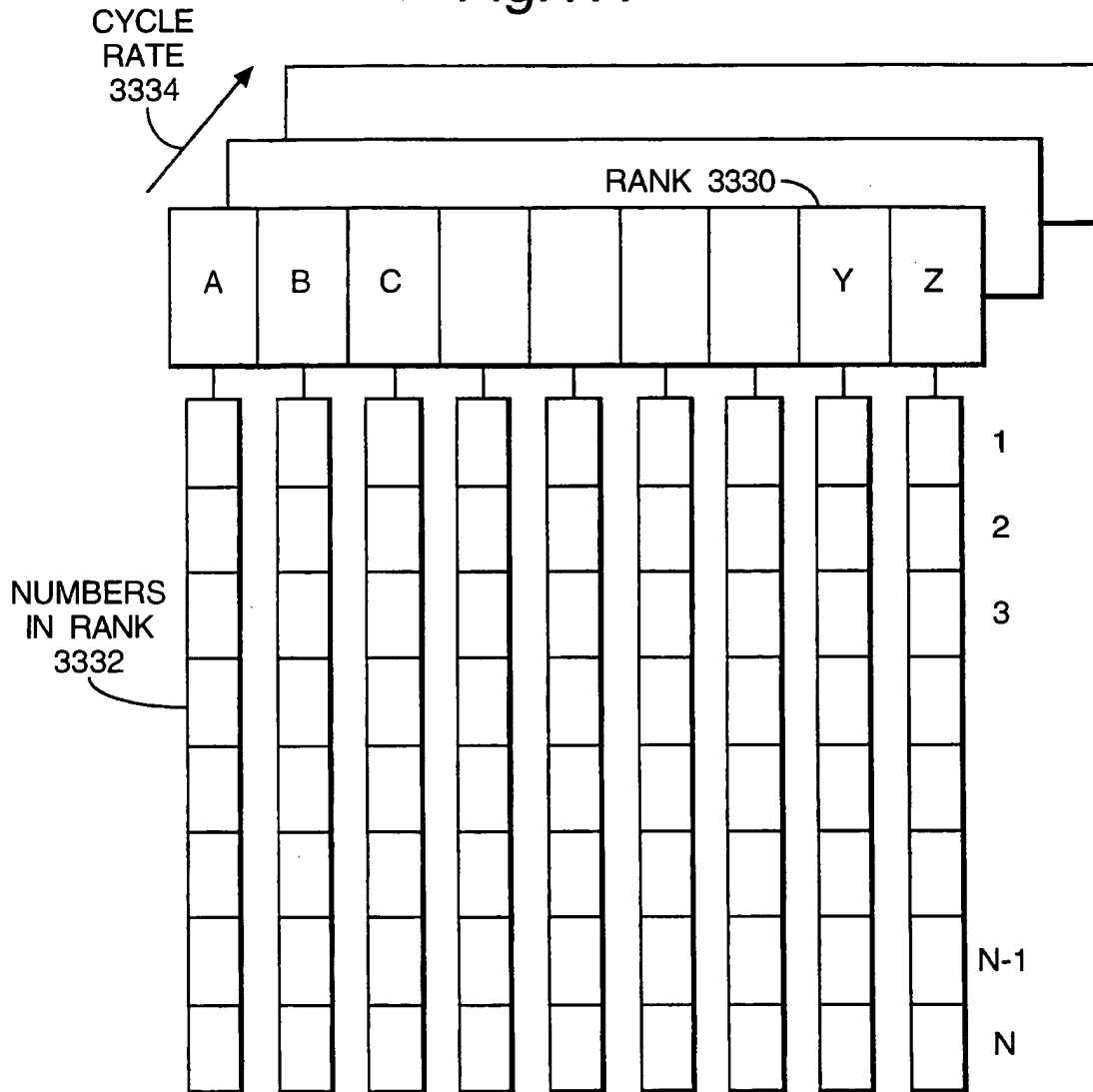


Fig.18.

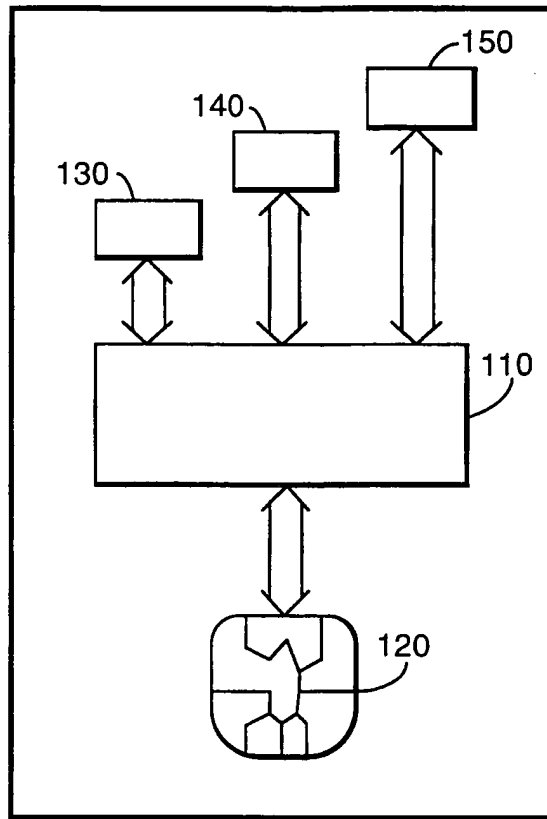
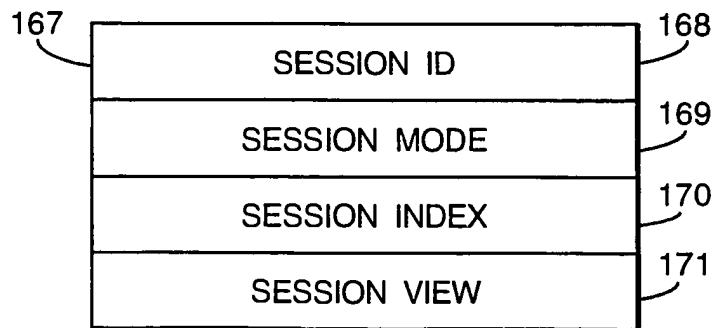
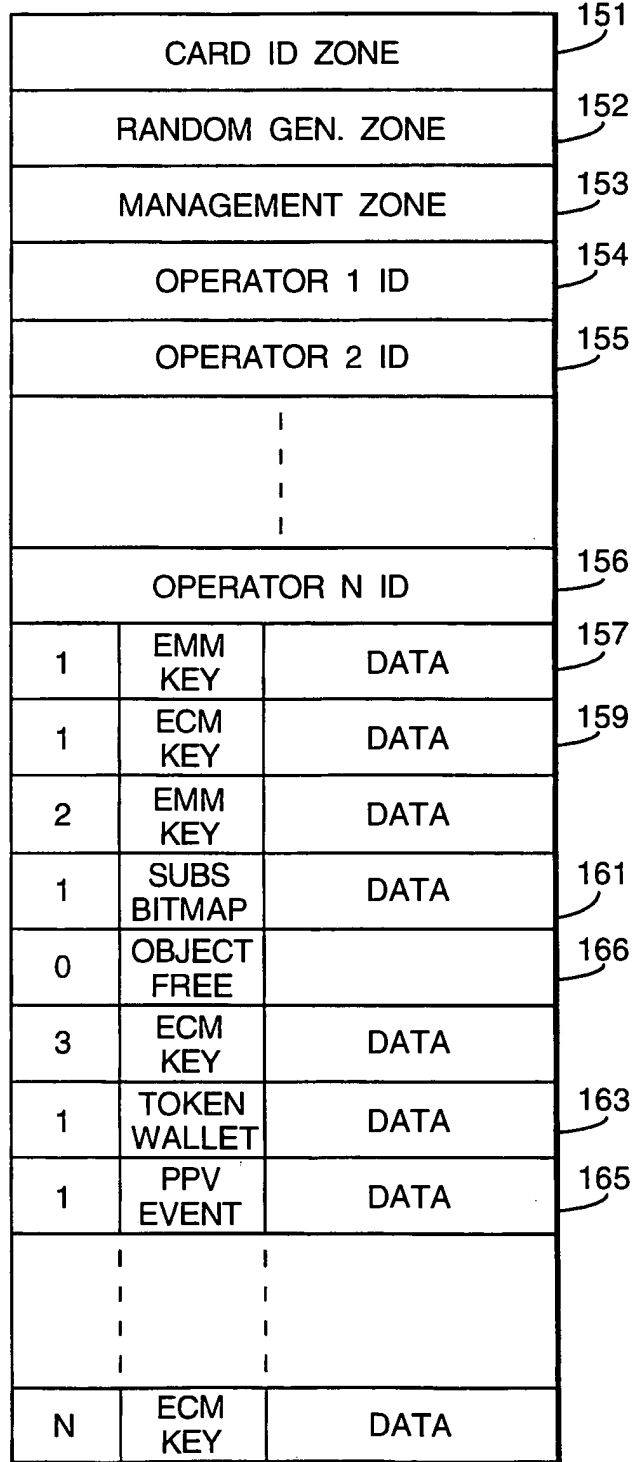


Fig.20.



SUBSTITUTE SHEET (RULE 26)

Fig.19.



SUBSTITUTE SHEET (RULE 26)

INTERNATIONAL SEARCH REPORT

Intern: al Application No
PCT/EP 97/02108

A. CLASSIFICATION OF SUBJECT MATTER IPC 6 H04N7/16 H04N7/167		
According to International Patent Classification (IPC) or to both national classification and IPC		
B. FIELDS SEARCHED		
Minimum documentation searched (classification system followed by classification symbols) IPC 6 H04N		
Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched		
Electronic data base consulted during the international search (name of data base and, where practical, search terms used)		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	"FUNCTIONAL MODEL OF A CONDITIONAL ACCESS SYSTEM" EBU REVIEW- TECHNICAL, no. 266, 21 December 1995, pages 64-77, XP000559450 see the whole document ---	1-12, 15-19, 21,22
X	WO 94 14284 A (DISCOVERY COMMUNICAT INC) 23 June 1994 see page 8, line 8 - page 14, line 23 see page 18, line 28 - page 21, line 19 see page 24, line 25 - page 29, line 31 see page 33, line 8 - line 17 see figures 1-11 --- -/--	1-12, 14-17
<input checked="" type="checkbox"/> Further documents are listed in the continuation of box C.		
<input checked="" type="checkbox"/> Patent family members are listed in annex.		
° Special categories of cited documents :		
A document defining the general state of the art which is not considered to be of particular relevance	*T* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention	
E earlier document but published on or after the international filing date	*X* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone	
L document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	*Y* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.	
O document referring to an oral disclosure, use, exhibition or other means	*&* document member of the same patent family	
P document published prior to the international filing date but later than the priority date claimed		
Date of the actual completion of the international search <div style="text-align: center; font-size: 1.2em;">11 November 1997</div>	Date of mailing of the international search report <div style="text-align: center; font-size: 1.2em;">18. 11. 97</div>	
Name and mailing address of the ISA European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Authorized officer <div style="text-align: center; font-size: 1.2em;">Van der Zaal, R</div>	

INTERNATIONAL SEARCH REPORT

International Application No
PCT/EP 97/02108

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 5 144 663 A (KUDELSKI ANDRE ET AL) 1 September 1992 see column 2, line 5 - line 23 see column 3, line 6 - column 4, line 65 see column 5, line 62 - column 8, line 58 see figures 1-11 -----	1-12

1

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No

PCT/EP 97/02108

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9414284 A	23-06-94	AU 5732994 A	04-07-94
		AU 5733094 A	04-07-94
		AU 5733194 A	04-07-94
		AU 5733294 A	04-07-94
		AU 5736394 A	04-07-94
		AU 5845894 A	22-06-94
		AU 5869894 A	04-07-94
		CA 2151458 A	23-06-94
		CN 1093211 A	05-10-94
		CN 1090451 A	03-08-94
		CN 1090452 A	03-08-94
		CN 1096151 A	07-12-94
		CN 1090453 A	03-08-94
		CN 1090454 A	03-08-94
		EP 0673578 A	27-09-95
		EP 0673579 A	27-09-95
		EP 0673580 A	27-09-95
		EP 0673581 A	27-09-95
		EP 0673582 A	27-09-95
		EP 0673583 A	27-09-95
		EP 0674824 A	04-10-95
		IL 107908 A	10-01-97
		IL 107909 A	15-04-97
		IL 107910 A	10-06-97
		IL 107912 A	18-02-97
		IL 107913 A	15-04-97
		JP 8510869 T	12-11-96
		JP 8506938 T	23-07-96
		JP 8506939 T	23-07-96
		JP 8506940 T	23-07-96
		JP 8506941 T	23-07-96
		JP 8506942 T	23-07-96
		NZ 259146 A	26-05-97
		NZ 259147 A	26-05-97
		NZ 259148 A	26-11-96
		WO 9413107 A	09-06-94
		WO 9414279 A	23-06-94
		WO 9414280 A	23-06-94
		WO 9414281 A	23-06-94
		WO 9414282 A	23-06-94

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No

PCT/EP 97/02108

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9414284 A		WO 9414283 A	23-06-94
		US 5559549 A	24-09-96
		US 5600364 A	04-02-97
		US 5659350 A	19-08-97

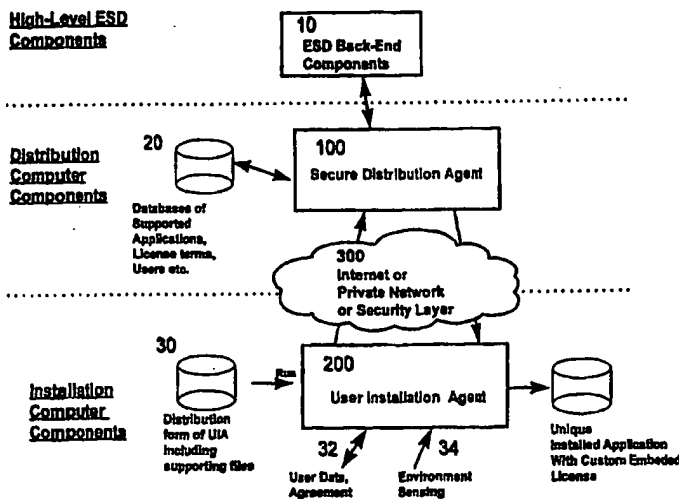
US 5144663 A	01-09-92	AU 599646 B	26-07-90
		AU 7157887 A	22-10-87
		DE 3751410 D	24-08-95
		DE 3751410 T	11-04-96
		EP 0243312 A	28-10-87
		EP 0626793 A	30-11-94
		ES 2076931 T	16-11-95
		JP 2610260 B	14-05-97
		JP 63023488 A	30-01-88
		JP 2520217 B	31-07-96
		JP 5244591 A	21-09-93



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

<p>(51) International Patent Classification ⁶ : G06F 1/00</p>	<p>A1</p>	<p>(11) International Publication Number: WO 98/45768 (43) International Publication Date: 15 October 1998 (15.10.98)</p>
<p>(21) International Application Number: PCT/CA98/00241 (22) International Filing Date: 18 March 1998 (18.03.98) (30) Priority Data: 08/831,696 10 April 1997 (10.04.97) US (71) Applicant: NORTHERN TELECOM LIMITED [CA/CA]; Station A, P.O. Box 6123, Montreal, Quebec H3C 3J5 (CA). (72) Inventors: LAROSE, Gordon, Edward; 2417 Baseline Road, Ottawa, Ontario K2C 0E3 (CA). ALLAN, David, Ian; 852 Forest Street, Ottawa, Ontario K2B 5P9 (CA). (74) Agents: MCGRAW, James et al.; Smart & Biggar, 900-55 Metcalf Street, P.O. Box 2999, Station D, Ottawa, Ontario K1P 5Y6 (CA).</p>		<p>(81) Designated States: AU, CA, CN, JP, European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE). Published <i>With international search report.</i></p>

(54) Title: METHOD AND SYSTEM FOR NETWORKED INSTALLATION OF UNIQUELY CUSTOMIZED, AUTHENTICABLE, AND TRACEABLE SOFTWARE APPLICATIONS



(57) Abstract

A method to create, distribute and install on an installation computer a uniquely customised instance of a software application that is authenticable and traceable to a particular user. A secure distribution agent resident on a distribution computer collects identifying information, and calculates a cryptographic signature of the software application and identifying information. The identifying information and cryptographic signature are embedded in the software application by the secure distribution agent. The software application with embedded data is transmitted via a distribution channel to the installation computer. A user installation agent resident on the installation computer manages the installation of the software application with embedded data on the installation computer. Prior to installation, the user installation agent may use the cryptographic signature to verify that the software application, and the identifying information, are authentic and have not been tampered with.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Larvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

- 1 -

METHOD AND SYSTEM FOR NETWORKED INSTALLATION OF
UNIQUELY CUSTOMIZED, AUTHENTICABLE, AND TRACEABLE
SOFTWARE APPLICATIONS

FIELD OF THE INVENTION

5 This invention relates to a method and system for the electronic distribution and installation to users via a network of software applications that are uniquely customized, authenticable and traceable to the individual user.

BACKGROUND OF THE INVENTION

10 With the increasing importance and reliance on networked computer environments such as the Internet, Electronic Software Distribution (ESD) is assuming an increased importance as a means of distributing software applications to users. The on-line infrastructures currently in place enable
15 users to purchase and install software applications without the need for physical delivery of shrink-wrapped software. Typically, a software publisher will prepare a master of the software application for electronic distribution. A customer will then go on-line and submit an order to purchase the
20 software application, which will be received and fulfilled by the publisher. The customer will then download the software application and install it to his/her own computer.

 A disadvantage of the current on-line infrastructure is that it delivers software applications to users in a form
25 that is identical with those found in retail stores and catalogues. Absent cryptographic protection, users can freely share the distribution form of the software amongst themselves.

 Even where cryptographic protection are present, the potential for unauthorized copying is still significant because
30 all the users possess identical copies (necessarily having identical encryption schemes) of a software application. There is in all such cases a single underlying decryption key, and in most cases this key, or an equivalent variant of it, is entered by the user, who can then share it with other users who can use
35 it to obtain unlicensed usage of the program. There exist today bulletin boards and Internet sites devoted to the sharing of such keys, which are visited by persons interested in

- 2 -

obtaining unpaid for usage of programs by applying such keys to copies of the applications they have obtained.

Further, even where more subtle anti-piracy schemes are in place in a software application, it is not uncommon for software "hackers" to produce "crack" programs which can be used to process a freely-distributed, limited functionality version of a program to produce a revised, fully-functional version of the same program which can be used without purchasing a license. Even the most ingenious forms of single-key mass distribution, which might involve input of one-time-only responses to a dynamic challenge to infer the key, are vulnerable to a "crack" which simply causes the application of the "true" universal decryption key. Although such "crack" involves more technical sophistication than sharing of keys as above, the distribution channels and potential effect on the product's revenues are very similar.

In addition, software applications distributed by conventional ESD techniques provide no means to police their own integrity to prevent unauthorized tampering.

Portland Software has produced an electronic software distribution system sold under the trade-mark ZipLock™ that packages software for electronic distribution over the Internet. The ZipLock™ system discloses a system that distributes, from a secure server to a client resident on the user's computer, a standard executable software application that is protected by means of a cryptographic key. Data input by the users is transmitted to the secure server and is used to construct a customized digital licence certificate that is transmitted to the user in a separate computer file. The Ziplock™ system does not provide a mechanism to detect tampering done to the executable software application itself, nor does it provide traceability if the digital licence certificate is not included with an unauthorized redistribution of the software application.

The prior art discloses a number of other systems and methods to protect unauthorized use of software electronically distributed to users. In Choudhury U.S. Pat. No. 5,509,074,

- 3 -

there is disclosed a method of protecting electronically published materials using cryptographic protocols. A first described embodiment requires special purpose hardware to decrypt the document that is transmitted to the user. This
5 eliminates the method from general use with personal computers used by the general public. In a second method, there is no requirement for special purpose hardware. In this method, the publisher modifies the inter-line or inter-word spaces of the document to make each document unique for each user. The
10 unique document is then encrypted and transmitted to the user's computer. Upon receipt of the encrypted document, the user's computer will prompt the user to enter his/her secret key which is used to decrypt the document for viewing. The method disclosed by this reference does not prevent piracy, it only
15 discourages piracy by making the pirated document traceable to the user. In addition, this reference pertains only to data files, not to the protection of executable files of any type.

In Cane U.S. Pat No. 5,416,840, there is disclosed a method and system for protecting computer program distribution
20 in a broadcast medium, such as radio frequency public broadcast or computer network. In this reference, the method involves encrypting at least a portion of a computer program, the user being supplied with a password for use in decrypting the computer program so that the computer program can be installed
25 and used. A unique password is generated and transmitted to the user for subsequent use in decrypting the selected software program contained on the medium. While there is disclosed a method and system for the generation, transmission and use of unique passwords that cannot be shared among different users of
30 the software application, this reference requires the user to own proprietary hardware that eliminates it from general use with personal computer owned by the general public.

In Yuval U.S. Pat No. 5,586,186, there is disclosed a method and system for controlling unauthorized access to
35 software distributed to users. The main components of the system are an encryptor, a user key generator, and a decryptor. The encryptor generates encryption and decryption keys,

- 4 -

encrypts the software using the encryption keys, and stores the encrypted forms of the software of the broadcast medium, such as CD ROM. The user key generator generates a unique key using numeric representations of identifying information supplied by users and the decryption keys. The decryptor is responsible for decrypting the encrypted forms of the software using the identifying information supplied by the user, and the unique user keys. The decryption method disclosed by this reference enables a large number of different but logically similar keys to be used as decryption keys, each of which is unique to a particular user. However, this reference does not disclose a means to customize a software application with user-specific data such that the software application itself can be authenticated. Furthermore, this reference does not prevent piracy by sharing of keys; it only discourages it through traceability of keys.

SUMMARY OF THE INVENTION

The present invention pertains to a method for the electronic distribution of a software application from a distribution computer to an installation computer comprising the steps of receiving at said distribution computer identifying information, embedding said identifying information in said software application at said distribution computer to form an identifiable software application, generating a cryptographic signature for said identifiable software application, embedding said cryptographic signature in said identifiable software application to form an identifiable and authenticable software application, and transferring said identifiable and authenticable software application from said distribution computer to said installation computer.

The method and system of the present invention discloses an on-line software customization, delivery and installation scheme. Instead of distributing a software application to a user that results in the installation of a totally generic, untraceable executable file on the installation computer, the method and system disclosed herein discloses a means to create, distribute and install on an

- 5 -

installation computer a uniquely customised instance of a software application that is authenticable and traceable to a particular user.

The method and system disclosed herein provides for a user installation agent (UIA) resident on an installation computer to establish a connection through a distribution channel to a secure distribution agent (SDA) resident on a distribution computer. The UIA and/or SDA prompt the user to input identifying information that, together with business related information such as licensing terms, etc., is used to create a unique data set that is embedded in the desired software application by the SDA. By the use of a cryptographic hash algorithm, and private/public key cryptography wherein a private key is only known to the SDA, a cryptographic signature of the desired software application and embedded data set is calculated and also embedded into the software application. The software application with embedded data and cryptographic signature is transmitted via a distribution channel to the installation computer where it is installed on the installation computer. Optionally, the installation computer may use the cryptographic signature to verify that neither the software application, nor the embedded data have been tampered with. Public key(s) used to decrypt the cryptographic signatures may be transmitted to the installation computer with the software application, or by any other means, such as e-mail, Internet bulletin boards, etc. Following installation, the embedded data and cryptographic signature are used in a variety of ways, such as to provide a means to trace the software application to the user, to police the continued integrity of the software application, to ensure that license conditions continue to be met, to perform virus checking, or automatic upgrading of the software application itself.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 is a block diagram of a system overview showing the various inputs and components of the system and method of the present invention;

- 6 -

Figure 2 is a data flow diagram of the structure and operation of the Secure Distribution Agent employed by the present invention;

Figure 3A is a block diagram showing details of the construction of the aggregate distribution file using a one-step cryptographic process;

Figure 3B is a block diagram showing details of the construction of an aggregate distribution file using a two-step cryptographic process;

Figure 3C is a block diagram showing details of the construction of an aggregate distribution file using a cryptographic process that is a variant of the two-step cryptographic process shown in Figure 3B;

Figure 4 is a block diagram of the structure and operation of the User Installation Agent employed by the present invention;

Figure 5 is a block diagram showing the means of extracting and authenticating embedded data from an installed distribution file;

Figure 6 is a flow chart of a first embodiment of the present invention that authenticates embedded data by means of a common encryption key;

Figure 7 is a flow chart of a second embodiment of the present invention that authenticates embedded data by means of a unique per-user encryption key; and,

Figure 8 is a block diagram showing the various uses of the installed software application delivered to the user by means of the present invention.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENT

Figure 1 shows the various inputs and components of the system and method of the present invention. At the top level is shown a representation of the Electronic Software Distribution (ESD) back-end components 10, which include clearinghouses of software, software manufacturers, publishers, credit card servers, etc., all of which interact with a Secure Distribution Agent (SDA) 100 resident on a distribution computer that forms an essential part of the present invention.

- 7 -

The SDA 100 interfaces with these ESD back-end components 10 via the Internet or private computer network to provide payment methods support, loading of software applications from publishers, etc. The exact nature of the ESD back-end
5 components 10 may vary without affecting the method and system of the present invention.

The SDA 100 is comprised of a system of co-operating software programs, which run in a secure environment. The nature of the secure environment is immaterial to the invention
10 as long as it ensures the ability to protect the privacy of user data, authentication of users and possibly other third-parties, and suitable limitations on the operations which can be accessed externally. This environment might or might not be physically separated from an installation computer. The
15 structure and operation of the SDA 100 is more fully described in Figure 2.

One of the inputs to the SDA 100 is a set of databases 20 of supported software applications, license terms, licensed users, etc. The SDA 100 transmits and receives
20 relevant data to/from the databases 20 prior to, and during the operation of the present invention. The exact nature and content of the databases 20 is not an essential feature of the invention.

A distribution channel 300 is illustrated in Figure 1
25 that can comprise computer networks such as the Internet, or private network, or a security layer as required to maintain security if the SDA 100 were located in close proximity with a user installation agent (UIA) 200. Alternatively, it may contain some combination of these elements. The distribution
30 channel 300 is used to connect the UIA 200 to the SDA 100 (and thus connect the distribution computer to the installation computer) so that information may be exchanged between these two agents, and so that an aggregate distribution file 170 (shown in Figure 2) can be distributed from the SDA 100 to the
35 UIA 200. Though the distribution channel 300 is illustrated between the SDA 100 and the UIA 200, the system of the present

- 8 -

invention does not require that the SDA 100 be physically distant from the UIA 200.

At a user's end is the UIA 200 which is an installation/automatic upgrade software program resident on the installation computer. This program is used to communicate via the distribution channel 300 to the SDA 100, and to perform the required operations, more fully described below, on the installation computer. Though normally one UIA 200 would be required for each supported software application, persons skilled in the art would be familiar with the capability to develop a UIA 200 that would support multiple software applications. Also shown in Figure 1 is a distribution form 30 of the UIA 200, including support files. The nature of the distribution form 30 of the UIA 200 is immaterial to the operation of the present invention. Any of CD ROM, World Wide Web (WWW) download, floppy diskette, etc. could be used.

The UIA 200 accepts data 32 input from the user, such as name, address, payment options, etc., as well as data pertaining to the acceptance of an end-user license. Environment sensing data 34 such as speed of CPU, size of hard disk, speed of modem, etc. may also be input to the UIA 200 for processing. The identifying information processed by the UIA 200 may include any information pertaining to the purchaser, the seller, the installation agent, date, serial number, license specifics, etc. This data may be used for the automatic registration of the desired software application with a publisher or its commercial proxies.

As noted above, the identifying data 32, 34 constitutes identifying information concerning the user, its computer, etc. The identifying data 32, 34 is processed by the UIA 200 and transmitted to the SDA 100 via the distribution channel 300. Of course, it is understood that the identifying information does not necessarily have to be transmitted to the SDA 100 by means of the distribution channel 300. For example, the identifying information may be locally entered into the SDA 100 by an agent using information received verbally, in writing, or in some other non-electronic manner. The SDA 100

- 9 -

combines the identifying data 32, 34 with the data stored in the databases 20 to produce an aggregated distribution file 170 that is uniquely customized, authenticable, and traceable to the user. The aggregate distribution file 170 is transmitted
5 via the distribution channel 300 to the UIA 200. The output from the UIA 200 is a uniquely customized software application 15 (to be referred to below as an "installed aggregate distribution file") installed on the installation computer, with identifying information embedded therein.

10 Though the description of the present invention implies that the "user" is an individual user of the software application 15 to be installed on a personal computer, persons skilled in the art will appreciate that the present invention would also operate in the context of a networked end-user
15 environment, where the "user" was a network administrator responsible for installing software on a central server for use by a number of end users.

 Figure 2 is a data flow diagram of the structure and operation of the SDA 100 employed by the present invention. An
20 original distribution file 130 is shown as an input to a conversion program 110. In the envisioned implementation, the original distribution file 130 is input to the SDA 100 by the databases 20 shown in Figure 1. It is understood that the original distribution file 130 does not necessarily have to be
25 input to the SDA 100 by the databases 20, since the original distribution file 130 may already be resident on the distribution computer containing the SDA 100. The conversion program 110 has, as additional inputs, the data 140 to be embedded in the distribution file 130, and required
30 public/private cryptographic key pairs 150. The embedded data 140 is produced by a user interaction program 120 which interacts with the user through the UIA 200 to receive identifying data 32, 34 (shown in Figure 1) as well as data from the databases 20 of supported software applications,
35 license terms, licensed users, etc.

 While the embedded data 140 can be of any form and content, it is anticipated that the embedded data 140 will

- 10 -

contain information enabling the software application 15 to be traceable to an individual user and license transaction. For example, the embedded data 140 can include a unique serial number used to identify the aggregate distribution file 170 to be distributed to the user. The would eliminate serial number fraud that is common in the software industry, whereby current software applications can only perform simple validity checks, which can be fooled by widespread fraudulent re-use of a single valid serial number. The embedded data 140 may take the form of a complete license agreement customized to the individual user, including user name, address, software serial no., license terms, etc. Records of the user information collected by the user interaction program 120 may be kept by the databases 20.

15 The output of the conversion program 110 is an aggregate distribution file 170 which contains both the contents of the original distribution file 130, the embedded data 140, as well as a cryptographic signature of the embedded data 140 and the original distribution file 130. The aggregate distribution file 170 is then transmitted via the distribution channel 300 to the UIA 200. The UIA 200 then installs the aggregate distribution file 170 on the installation computer. Once the aggregate distribution file 170 is installed, it takes the form of an installed aggregate distribution file 15.

25 By means of its connection with the UIA 200, the SDA 100 can negotiate arbitrary license terms with the user, display an End-User License Agreement (EULA), confirm acceptance of that agreement, and automatically perform on-line registration of the software based on the already-established identity of the user and the specific license terms. Subject to commercial and legal considerations, an SDA 100 could offer different pricing and license terms, and possibly different executable versions, to users in different countries, for example. In addition, differential pricing based on attributes of the installation computer such as CPU power could be provided.

- 11 -

The SDA 100 does not require intelligence within itself for functions such as establishing that a user's stated address and credit-card number are valid, consistent, and within a given geographical area. Such functions may be
5 undertaken by the high-level ESD components 10 illustrated in Figure 1.

Figure 3A shows the procedure for constructing the aggregate distribution file 170 in greater detail. For the sake of illustration, the original distribution file 130 is
10 assumed to have a structure including header information and different types of internal sections for code, static data and so on, such as a Windows™ 'Portable Executable' (PE) program file. One of ordinary skill in the art can appreciate that the method and system of the present invention can be applied to a
15 number of different file formats. Similarly, the inputs 140, 151 to the conversion program 110, and output 170 from the conversion program 110 are illustrated to be computer files, but they could be in-memory images, streams from other processors, etc.

20 A typical sequence of steps run by the SDA 100 to construct the aggregate distribution file 170 is described below.

1. The conversion program 110 is run, as a result of the user interaction program 120 having determined that a
25 conversion is required i.e. that a delivery of a particular aggregate distribution file 170 according to the method of the present invention is authorized, and that the required embedded data block 140 has been constructed. All subsequent steps are executed by the conversion program 110 unless otherwise
30 indicated.

The object is to obtain what is often referred to as a "digital signature", or "cryptographic signature" which inherently has two aspects:

(i) By the use of a cryptographic hash algorithm, the
35 production of a cryptographic fingerprint that uniquely corresponds to the data "ed" 130, 140; and

- 12 -

(ii) Protection of that cryptographic fingerprint by encrypting it with a private key, such that the recipient of the cryptographic fingerprint may, by using a public key and the cryptographic algorithm, verify that the data "ed" 130, 140 is intact, without having the ability to generate a new cryptographic fingerprint, and plausibly change the data.

These two steps are essential to realize the advantage of the present invention, since without both steps a third party may intervene and alter data without the recipient being able to detect it. This procedure is to be distinguished from simply encrypting the data "ed" 130, 140, which is a step that is not necessary to the operation of the present invention since there is no way to plausibly alter the data 130, 140 without such alterations being detectable.

2. The input/output logic 111 of the conversion program 110 reads in the desired original distribution file 130, its corresponding cryptographic private key 151, and the data to be embedded 140. Though not required by the conversion program 110, a public key 152 may be passed through in order that it may be added to the aggregate distribution file 130. Utilizing cryptographic hash algorithms 112 and Public-Private key (PPK) encryption algorithms 113, a cryptographic signature 174 is produced. The basic steps of this process are:

2.1 Apply a one-way hash function "hf" to the data "ed" 130, 140 producing a cryptographic fingerprint "edh", that is $edh = hf(ed)$. The requirements for this cryptographic fingerprint are as follows: (i) that it produce a reasonably compact result i.e. $size(edh) \ll size(ed)$, and preferably a fixed-length result; (ii) that the fingerprint alone cannot be used to ascertain the original data block back i.e. there is no back-hash function "bhf" such that $bhf(edh) = ed$; (iii) that it be extremely sensitive to changes in "ed"; say, that a single-bit change in "ed" changes on average 50% of the bits in "edh", and (iv) that it is extremely difficult to

- 13 -

construct a false embedded data block "fed" which produces the same fingerprint as "ed", that is $hf(ed) = hf(fed)$. There are a number of algorithms which satisfy these requirements, such as MD5 (Message Digest 5) and SHA (Secure Hash Algorithm). Other algorithms that also meet the above criteria that may be employed by the present invention.

2.2 Encrypt the cryptographic fingerprint "edh" using the private key 151 "prk" and a public/private encryption function "ppef" to produce a cryptographic signature "edf" 174, that is: $edf = ppef(prk, edh)$. The requirements for the encryption function "ppef" are as follows: (i) that it produce a result not substantially larger than its input; (ii) that it effectively protect relatively short data sets, since "edh" will be bytes long rather than kilobytes long; (iii) that is computationally infeasible to use the public key 151 ("puk") and the cryptographic signature "edf" 174, or multiple instances of "edf" 174 (which will be visible on the installation computer) to infer the private key "prk", that is, there is no cracking function "cf" such that: $puk = cf(edf, puk)$; (iv) that there is no conceivable means of replicating the behaviour of "ppef" using "prk" without in fact possessing both "ppef" and "prk". In principle, "ppef" can be inferred from its corresponding decryption function, so "prk" is the important secret in practical terms; (v) that the corresponding public-key decryption function "ppdf" have acceptable performance on a typical installation computer for the pertinent file sizes. Note that if a specific ppef/ppdf is chosen for security reasons and does not yield acceptable performance, the encryption could be applied to only a portion of the selected files and still offer the same benefits; (vi) that it be suitable (preferably, via established cryptanalysis) for specific application in this domain i.e. digital signatures. There are a number of algorithms which might satisfy these requirements, such RSA and those of Rabin and ElGamal. The

- 14 -

careful selection of implementation parameters can help attain required security and performance.

3. The cryptographic signature 174 from step 2.1, and the data to be embedded 140 are inserted into the original distribution file 130 to produce the aggregate distribution file 170. This insertion is not a simple copying of bits into the middle of a file, since it must be compliant to the format requirements of the particular file types. For example, headers may have to be updated to identify the new data etc.

10 The system and method of the present invention does not require that the embedded data 171 or the cryptographic signature 174 be positioned in any particular manner in the aggregate distribution file 170. What is necessary is that: (i) the software on the installation computer, and the UIA 200 in particular, be able to locate the embedded data 171 and cryptographic signature 174, and (ii) that the aggregate distribution file 170, after it is installed on the installation computer, be able to perform its intended function; for example, if it is an executable file, that it still conform to structural and other platform requirements so it can load and run on the installation computer as it might have before the conversion process. For example, if the file were in a format common to current computers containing an Intel™ microprocessor and running a Microsoft™ Windows™ operating system, the conversion program 110 could inspect the "header" section of the original distribution file 130 to determine where there were sections containing static data so as to avoid sections containing executable code. A static data section would be selected and a suitable location for the embedded data block 171 and cryptographic signature 174 would be found or created. This would be done by, for example, (i) determining that an existing static data block had unused capacity sufficient to add the data, (ii) allocating a new static data block, or (iii) expanding an existing static data block.

The method illustrated in Figure 3A discloses a one-step process wherein cryptographic signature 174 is ascertained

- 15 -

for the original distribution file 130 and the embedded data 140. An optional method, such as that illustrated in Figure 3B, would be to employ a two-step process wherein a cryptographic signature 172 of the embedded data 171 is first produced using the same algorithm described in step 2 above. This embedded data cryptographic signature 172 is then itself embedded into the original distribution file 130. The original distribution file 130, embedded data 171, and embedded data cryptographic signature 172 are then input to the second cryptographic step, wherein an overall cryptographic signature 176 is ascertained using the same algorithm described in step 2 above. The benefit of the two step process is that it augments the capabilities of the system and method of the present invention to authenticate and detect tampering in the software application installed on the installation computer. For example, separate cryptographic public/private key pairs could be provided for the two cryptographic signatures 172, 176. Furthermore, the two-step process allows the embedded data 171 to be extracted and authenticated, even if the original file contents 173a, 173b have been corrupted.

Another alternative is to construct the aggregate distribution file 170 using a variation of the two-step cryptographic process wherein a first cryptographic signature 175 is made of only the original file contents 173a, 173b, and a second cryptographic signature 172 is made of the embedded data 171. This is illustrated in Figure 3C. This scheme has all the advantages of the two-step process illustrated in Figure 3B, and also allows for separate authentication of the embedded data 171 and the original file contents 173a, 173b. This would allow a user to verify that original distribution file 130 provided by the publisher had not been altered by the on-line installation process disclosed by the present invention.

One of ordinary skill in the art will appreciate that any of the cryptographic signatures 172, 174, 175, 176 shown in Figures 3A, 3B and 3C do not have to be produced using the same set of cryptographic public/private key pairs, or even the same

- 16 -

cryptographic algorithms. As well, it is not necessary that the cryptographic signatures 172, 174, 175, 176 be calculated each time an aggregate distribution file 170 is distributed to a user. The SDA 100 could maintain a database of

5 partially-precomputed signatures to speed up the related calculations. The availability of cryptographic hardware support such as RSA co-processors in the installation computer, could be used to attain good responsiveness with maximal security. As well, it is not essential that the aggregate

10 distribution file 170 be constructed in its entirety by the SDA 100. What is necessary is that the aggregate distribution file 170 be derivable in its entirety by the UIA 200.

Figure 4 illustrates the structure and operation of the UIA 200 which consists of a transient installation index

15 204, a transient installation input fileset 205, and a UIA proper executable software program 203. One skilled in the art will appreciate that there are many ways in which the UIA program 203 could be implemented. Since a significant part of the UIA's 200 functionality involves user interaction and

20 dialog with the SDA 100, options for the implementation of the UIA 200 include either making it an adjunct to a World Wide Web (WWW) browser, or implementing it as a stand-alone program which itself embeds or invokes already-present browser capability on the installation computer.

25 A typical execution sequence of the UIA 200 is described below:

1. After the UIA program 203 and its support data 204, 205 have been copied onto the installation computer, the user runs the UIA program 203. Note that the UIA program 203 could

30 also have been initiated remotely e.g. sent as an active program within a browser framework by a WWW server. Unless otherwise stated, all subsequent steps are executed by the UIA program 203.

2. The installation index 204 and installation input

35 fileset 205 are read by the installation computer to determine the particular default SDA 100 appropriate for the installation

- 17 -

of the desired software application (known as the "installed aggregate distribution file") 15.

3. The installation computer is examined to determine the probable means of establishing communications with the SDA 100, for example, the presence of TCP/IP network interfaces, modems etc. If no such means are found, the program optionally assists the user to find parameters which will work properly, then fails with a warning. This is because access to the SDA 100 is essential for operation of the invention.

10 4. The user 1 is prompted with the default data from steps (2) and (3) above, i.e. informed where the UIA program 203 will look for the desired SDA 100, and over what sort of distribution channel 300. The user 1 is then given an opportunity to change this information, either for commercial 15 reasons (e.g. maybe an SDA has changed names or locations), or for technical reasons (e.g. the user does not have working TCP/IP connectivity and wants to use a straight modem link, perhaps via an 800 number.)

5. Via the distribution channel 300, the UIA program 203 20 establishes contact with SDA 100. If this cannot be done, the UIA program 203, after optionally helping the user determine parameters which will work properly, fails with a warning. While the security of the distribution channel 300 is optional to the operation of this invention, it is expected that the 25 distribution channel 300 will support appropriate protocols to protect the SDA 100 from fraud. A common protocol supporting authentication and privacy, such as Secure Sockets Layer (SSL) is appropriate.

6. The UIA program 203 acts as an intermediary between 30 the user and the SDA 100, enabling the user 1 to establish any legitimate agreement which the SDA 100 supports with respect to the desired installed aggregate distribution file 15. The UIA program 203 also has the ability to determine whether the available system resources of the installation computer meet 35 the requirements of the desired installed aggregate distribution file 15.

- 18 -

There are no technical limits to the variety of options that can be displayed to the user, the questions the user might be asked, the data that might be gathered about the installation computer, etc. Since the SDA 100 is being run
5 throughout the data gathering, data embedding, software distribution and software installation process, the system and method of the present invention can employ various levels of cryptography without the user ever being informed of the cryptographic keys, or any information from which they could be
10 derived. This is unlike other electronic delivery systems which typically require subsequent off-line entry of 'secret keys' or derivatives thereof which have been explicitly divulged to the user. Of course, public keys used for the authentication of cryptographic signatures are an exception in
15 that the user may be able to determine them easily, however this is not a security issue since they have no fraudulent application.

7. Assuming that the user 1 meets all the criteria set out by the SDA 100, the SDA 100 will determine a specific
20 set of files that must be transmitted to the UIA 200 to complete installation on the installation computer, notably including at least one aggregate distribution file 170 (shown in Figures 3A-3C). It is immaterial to the system and method of the present invention what is the nature of the agreement
25 entered into between the user 1 and SDA 100, or how it is validated. That is the responsibility of the SDA 100 and its subtending commercial systems 10, if any. Most importantly, the UIA 200 does not and cannot itself decide whether an agreement has been reached between the user and the SDA 100.
30 The UIA 200 does not have, and should not have, access to all the information required to complete the installation, except through interaction with the SDA 100.

8. The SDA 100 transmits an index of the required distribution files to the UIA 200 via the distribution channel
35 300. The UIA 200 uses this index to augment its own local index 204 forming a complete index for the upcoming installation.

- 19 -

9. The SDA 100 constructs one or more aggregate distribution files 170 and any other files required for the installation, and transmits these files to the UIA 200 via the distribution channel 300.

5 10. Using its local index and support files 204, 205 the UIA program 203 completes the installation of the installed aggregate distribution file 15 in a manner compliant with the platform of the installation computer. In particular, the UIA 200 installs the aggregate distribution file 170 such that the
10 cryptographic signature 174 and the embedded data 171 are unaffected. Once the aggregate distribution file 170 is installed on the installation computer, it is referred to as an installed aggregate distribution file 15. The UIA program 203 will also perform other system updates 212 as necessary, such
15 as updating the operating system registry (in the case of Windows 95™), and installing any additional application files. Other optional operations, such as leaving an appropriate 'uninstall' utility, may also be involved.

20 12. If an error should occur, the UIA program 203 may signal the SDA 100 to re-initiate the installation. If no error has occurred, the UIA program 203 signals the SDA 100 that all required data has been received. This could, for example, be used as the trigger signal for the SDA 100 to commit to a financial transaction. Leaving the financial commit
25 to this late part of the process minimizes the probability of the user being charged for a software application which has not been successfully installed, thus reducing one cause of customer frustration.

30 13. The UIA program 203 deletes any transient files, indices etc. that it might have placed on the installation computer.

14. The UIA program 203 disconnects from the SDA 100 and the distribution channel 300 and exits.

35 Upon successful completion of the optional authentication procedures described in further detail below, the user can then run the installed aggregate distribution file 15 on the installation computer. It should be understood that

- 20 -

the authentication procedures described below can be done either before or after the installation is completed.

The method and system of the present invention would diminish disputes arising from software which is purchased but not successfully installed. The UIA 200 can detect and warn the user if the installation computer had inadequate resources to run the desired software application, before any financial transaction has been made. Further, the final financial commitment to purchase the software application by the user can be done late in the installation process so that the probability of the financial transaction being successful, but the installation itself failing, would be low.

One of ordinary skill in the art will appreciate that the UIA 200 may be distributed to users in a mass-produced media form containing the original distribution file 130, or a derivative thereof not subject to successful fraudulent re-use through simple copying. In this scenario, the SDA 100 would transmit to the UIA 200 only the incremental information which the UIA 200 would require to construct the aggregate distribution file 170 and complete the installation. Any attempts to pirate the software application can be defeated by ensuring that the distribution form of the UIA 200 contains an incomplete set of executable files, thereby requiring essential data from the SDA 100 to be capable of executing on the installation computer.

Figure 5 illustrates the means of authenticating and extracting user data from an installed aggregate distribution file 15 to verify that neither the original file contents 173a, 173b, nor the embedded data 171 have been tampered with. This step is optional to the operation of the present invention because the installed aggregate distribution file 15 may be run by the user without authentication. It should be understood that the authentication procedures described below can be done either before or after the installation is completed. If authentication is done prior to installation on the installation computer, then the following procedures are

- 21 -

directed by the UIA 203 to the aggregate distribution file 170, instead of the installed aggregate distribution file 15.

The process illustrated in Figure 5 is in relation to an installed aggregate distribution file 15 constructed using the two-step process illustrated in Figure 3B. The principles of authenticating and extracting user data from an installed aggregate distribution file 15 constructed using the one-step cryptographic process illustrated in Figure 3A, or the variant of the two-step process illustrated in Figure 3C, are the same as those described below, with appropriate modifications, depending on the nature of the cryptographic signatures to be compared.

Though a separate authentication and reading program 400 is shown performing the functions of authentication and reading of embedded data 171, a person skilled in the art will appreciate that these functions need not be embodied in such a stand-alone program, and could be incorporated as functions of other programs, such as the UIA 200, a license-checker, a virus-checker, a program loader, a copy program, etc. A typical execution sequence of the authentication and reading program 400 is described below:

1. The authentication and reading program 400 is run, either by a user or by automatic invocation from another program such as the UIA 200. Unless otherwise indicated, the following steps are all executed by authentication and reading program 400.

2. Determine which installed aggregate distribution file 15 to process, either by prompting the user or having this passed as a parameter by the UIA 200. Also determine (if derivable therefrom in the particular implementation, as opposed to being contained in the file itself), which particular public key 152 is applicable to this installed aggregate distribution file 15.

3. Open the installed aggregate distribution file 15 in question and check that it meets the applicable format requirements. For example, a given implementation might support executable (EXE) and dynamic link library (DLL) files

- 22 -

in the 'PE' format for Intel™ processors. If the installed aggregate distribution file 15 fails these basic checks, or is not found, the authentication and reading program 400 fails with an appropriate warning.

5 4. Examine the file to determine the location of the overall cryptographic signature 176, the embedded data cryptographic signature 172, and the embedded data 171. The installed aggregate distribution file 15 can be formatted in various ways to support this, such as including pointers to
10 these sections in the file header. If applicable in the particular implementation, (i.e. the public key 152 is included in the file as opposed to being otherwise determined the authentication and reading program 400), find and extract the required public key 152.

15 If any of the above steps fail, the authentication and reading program 400 fails with an appropriate warning.

 5. Use the public key 152 to decrypt the overall cryptographic signature 176 into its unencrypted form 176a (the decrypted remote overall fingerprint).

20 6. Using the same known cryptographic signature algorithm as was employed by the SDA 100, calculate a local version 176b (the locally calculated overall fingerprint) of the overall cryptographic signature. This calculation will necessarily exclude the overall cryptographic signature 176 itself i.e.
25 cover all parts of the installed aggregate distribution file 15 except 176, in order that the locally calculated overall fingerprint 176b will not depend on itself.

 7. Compare the locally calculated overall fingerprint 176b to the decrypted remote overall fingerprint 176a. If they
30 differ, the authentication and reading program 400 will fail with a warning that the installed aggregate distribution file 15 has been corrupted. At this point, the UIA 200 may be invoked to contact the SDA 100 to re-acquire the installed aggregate distribution file 15.

35 8. Extract the embedded data 171 and present it graphically to the user, if the program has been user-invoked,

- 23 -

or pass it in message form to the invoker routine, if software-invoked.

9. Use the public key 152 to decrypt the embedded data cryptographic signature 172 into its unencrypted form 172a (the
5 decrypted remote embedded data fingerprint).

10. Calculate a local version 172b (the locally calculated embedded data fingerprint) of the embedded data cryptographic signature 172 using the same known cryptographic signature algorithm as the SDA 100 used.

10 11. Compare the locally calculated embedded data fingerprint 172b to the decrypted remote embedded data fingerprint 172a. If they differ, the authentication and reading program 400 will fail with a warning that the embedded data 171 has been corrupted.

15 A similar procedure of comparison would be followed in respect of the cryptographic signature 174 if the one step process illustrated in Figure 3A had been followed. As well, a similar procedure of comparison would be followed in respect of the original file contents cryptographic signature 175 if the
20 variant of the two-step cryptographic process illustrated in Figure 3C had been undertaken.

Figure 6 is a flow-chart of a summary of the procedures described in relation to Figures 2, 3A, 3B, 3C, 4 and 5. It should be noted that the public key 152 used to
25 authenticate the integrity of the installed aggregate distribution file 15 could be delivered to the UIA 200 by any means since it is not a secret and might be useful for more than one purpose. For example, the public key may be embedded in the aggregate distribution file 170, it may be explicitly
30 sent to the user as a separate file or message, or it may be obtained automatically by the installation computer from a network trusted authority (e.g. Verisign™ Inc.)

Figure 7 is a flow-chart of another set of procedures that may be employed in accordance with the present invention,
35 whereby the original file contents 173a, 173b are encrypted using a unique private key calculated by the SDA 100 for this particular transaction. A record of this unique private key is

- 24 -

kept by the SDA 100, and the corresponding unique public key is transmitted with the aggregate distribution file 170 via the distribution channel 300 to the UIA 200. The UIA 200 will decrypt the aggregate distribution file 170 using the public
5 key. For security reasons, it is preferred that this public key not be permanently stored on the installation computer. Instead, the unique public key would exist only in the computer's Random Access Memory (RAM) for the duration of the installation. This makes usable redistribution of the
10 aggregate distribution file 170 practically impossible.

Although the present invention has been described with reference to the preferred embodiments, one of ordinary skill in the art will recognize that a number of variations, alterations and modifications are possible. In Figure 8 there
15 is an illustration showing the various uses of the installed aggregate distribution file 15. After installation and authentication by the UIA 200, the installed aggregate distribution file 15 may run normally without making use of the embedded data 171 in any way. To ensure licence compliance,
20 the installed aggregate distribution file 15 may also be run in association with a license-enforcement program that verifies that any license terms comprising part of the embedded data 171 are being complied with. The embedded data 171 and
25 cryptographic signatures 172, 174, 175, 176 (depending on the manner in which the aggregate distribution file 170 was constructed) may also be used as an input to a virus checker that may perform an integrity check on the installed aggregate distribution file 15 by using the public key 152 and the same known cryptographic signature algorithm as was employed by the
30 SDA 100. Each time the installed aggregate distribution file 15 is run, the authentication and reading program 400 shown in Figure 5 may also be run, either by itself, or in association with an authenticating loader that would reject tampered files, and would not permit a tampered installed aggregate
35 distribution file 15 to be run. The embedded data 171 may also be used simply for display to the user.

- 25 -

The method and system disclosed herein can also be used to upgrade an installed aggregate distribution file 15 present on an installation computer. In this case, the UIA 200 and the SDA 100 would verify of the license status of the installed aggregate distribution file 15 present on the installation computer, and then invoke the method and system disclosed herein to construct, deliver and install an upgraded version of the installed aggregate distribution file 15 to the installation computer. The capability to invoke the upgrading feature of the present invention could be done at the request of the user, or it could be invoked automatically upon detection by the UIA 200 of the availability of a new version of the original distribution file 130.

The uniqueness of the installed aggregate distribution file 15 can be used to restrict its operation to a specific central processing unit (CPU) on the installation computer. The identification of the CPU for these purposes would be done by the UIA 200 during the stage of gathering data 32, 34 for transmission to the SDA 100.

The SDA 100 and UIA 200 disclosed herein are not restricted to being invoked at the time of installation or upgrading of the installed distribution file 170. For example, in a computer game environment, the SDA 100 and UIA 200 could be invoked when the user reaches a certain point in the game, giving the user the option to purchase additional functionalities or levels for the game.

This disclosure does not presuppose that the UIA 200 does not possess added intelligence to increase the functionality of the present invention. For example, the UIA 200 may possess the intelligence to find and recognize separate Personal Digital Certificates on the installation computer which establish his identity for purposes sufficient to authorize all, or part of, the transaction in question. Such Personal Digital Certificates and their method of application would conform to established standards such as those used by commercial certificate provider Verisign™ Inc. In addition, the UIA 200 could possess the intelligence to find and

- 26 -

recognize digital "coupon" certificates which establish that the user has some specific privilege, such as an entitlement to a specific price for a piece of software, or one which establishes his membership in a specific group, such as a company. In addition, the UIA 200 could locate pre-existing files installed according to the method of the present invention, and examine the embedded data 171. If the UIA 200 determines that there is license information present which may affect the terms of the transaction, or which may indicate a user's likely interest in, for example, an upgrade, the UIA 200 can transmit this information to the SDA 100 so that it can suitably mediate the transaction, advertise an upgrade, etc. A typical example of this would be examining a word-processing application installed in accordance with the present invention to determine that the user is entitled to a free upgrade, which the present invention can then proceed to install.

In another set of variations of the invention, the installed aggregate distribution file 15 is one which uses the principles of Nortel Algorithmic Authorization (NAA), as disclosed in U.S. Patent Application No. 08/674,037 to add robust self-policing of its own integrity. In a first variation, the run-time NAA algorithms, which already have the capability of using the installed aggregate distribution file's own code as an input required for proper operation, and thus of forcing catastrophic failure in the event of tampering, have the scope of this input expanded to include an in-memory copy of one or more of the data items added by the SDA 100, such as the overall cryptographic signature 176.

In a second variation, the "launch stub" component could go further, extracting and decoding the embedded data 171 in the installed aggregate distribution file 15, and comparing the license terms therein (e.g. a specific CPU identified by, say, a certain physical Media Access Control address on a network card) to those it found by examining its current environment. In accordance with the principles of Nortel Algorithmic Authorization, the "launch stub" would not have to "decide" whether to proceed, since such decision-

- 27 -

points are obvious attack points for 'hackers' wishing to defeat security mechanisms. Rather, it could modify data upon which proper program operation depended, in such a way that the program would continue to run properly only if said data
5 corresponded to the proper environment per the license. As for the first variation, the application would have been pre-constructed for the specific instance, as per the patent-pending technology, in such a way that its proper flow of control used input data that was initially 'incorrect' in
10 just such a way as to be 'corrected' only by application of the correct license data, or a simple derivative thereof.

The invention disclosed herein does not necessarily alter the functionality of the installed form of the installed aggregate distribution file 15, it only adds information and
15 authenticability to it. However, there are a number of means by which the behaviour of the installed aggregate distribution file 15 can be mediated in new ways enabled by this invention. In one variation, the SDA 100 would have access to a variety of executable forms for a given program, or
20 to software routines which would dynamically construct variant forms, in order to produce a program which meets particular customer function/cost requirements, and/or which actively binds itself to very specific license terms. For example, in the Microsoft Windows™ environment, different
25 behaviour could be embodied by different Dynamic Link Libraries (DLLs) which could be selectively included.

In another variation, the initial executable form of the program file would have specific functional and license-binding choices built-in, and the SDA 100 would inject
30 (possibly authenticable) data into the executable file which caused it to exhibit the desired behaviour. In yet another variation, the SDA 100 could make use of routines with detailed knowledge of specific program structures in order to add variant code to a pre-existing executable program which was not
35 explicitly designed to accommodate such variation.

The described embodiments of the present invention focus on a single "core" file of a specific file type as the

- 28 -

cornerstone of a software application's installation and security. However the method of the present invention may certainly be applied to more than one file or file type in a particular case. For example, all of the static files
5 associated with an installed software application could receive embedded information such that they were all authenticable and associable with the particular application and installation instance.

We Claim:

1. A method for the electronic distribution of a software application from a distribution computer to an installation computer comprising the steps of:
 - 5 a. receiving at said distribution computer identifying information;
 - b. embedding said identifying information in said software application at said distribution computer to form an identifiable software application;
 - 10 c. generating a cryptographic signature for said identifiable software application;
 - d. embedding said cryptographic signature in said identifiable software application to form an identifiable and authenticable software application;
 - 15 and
 - e. transferring said identifiable and authenticable software application from said distribution computer to said installation computer.
2. The method of claim 1, wherein the step of generating
20 a cryptographic signature for said identifiable software application includes the steps of
 - a. applying a one-way hash function "hf" to the identifiable software application "ed" producing a hash result "edh", where $edh = hf(ed)$; and
 - 25 b. encrypting the hash result "edh" using a cryptographic key to obtain a cryptographic signature.

- 30 -

3. The method of claim 2, wherein the one-way hash function is generated using any one of a Message Digest 5 (MD5) algorithm and a Secure Hash Algorithm (SHA) algorithm.
4. The method of claim 2 or 3, wherein the step of
5 encrypting the hash result "edh" includes the step of using a public/private encryption function "ppef" and a private encryption key "prk" to encrypt the hash result "edh" to produce a cryptographic signature "edf" where $edf = ppef(prk, edh)$.
- 10 5. The method of claim 4, wherein the public/private encryption function "ppef" is generated using any one of an RSA algorithm, a Rabin algorithm and an ElGamal algorithm.
6. The method of claim 1, 2, 3, 4 or 5, wherein the
15 distribution computer and the installation computer are connected by the Internet.
7. The method of claim 1, 2, 3, 4, 5, or 6, wherein the identifying information received at said distribution computer is transmitted from said installation computer.
8. A method of receiving electronically at an
20 installation computer a software application distributed from a distribution computer comprising the steps of:
- a. receiving an identifiable and authenticable software application from the distribution computer, the identifiable and authenticable software
25 application having embedded therein the identifying information and a cryptographic signature of the identifiable and authenticable software application; and
 - b. installing the identifiable and authenticable
30 software application at the installation computer.

- 31 -

9. The method of claim 8, wherein prior to the step of receiving an identifiable and authenticable software application from the distribution computer, the installation computer transmits identifying information to the distribution
5 computer.

10. The method of claim 8 or 9, wherein prior to the step of installing the identifiable and authenticable software application, the installation computer authenticates the integrity of the software application.

10 11. The method of claim 10, wherein the installation computer uses the cryptographic signature to authenticate the integrity of the software application.

12. A method for the electronic distribution of a software application from a distribution computer to an
15 installation computer comprising the steps of:

a. receiving identifying information at said distribution computer;

b. embedding said identifying information in said software application at said distribution computer to
20 form an identifiable software application;

c. generating a cryptographic signature for said identifiable software application;

d. embedding said cryptographic signature in said identifiable software application to form an
25 identifiable and authenticable software application;

e. transferring said identifiable and authenticable software application from said distribution computer to said installation computer; and

- 32 -

f. installing said identifiable and authenticable software application at said installation computer.

13. The method of claim 12, wherein the distribution computer and the installation computer are connected by the Internet.
5

14. The method of claim 12 or 13, wherein the identifying information received at said distribution computer is transmitted from said installation computer.

15. A software distribution computer for distributing an identifiable and authenticable software application to a user comprising:
10

a. a communications link between said software distribution computer and said user;

b. a storage device for storing a software application for distribution;
15

c. a communications interface in communication with said link, for receiving identification data from said user, and for transferring said identifiable and authenticable software application to said user;

d. means for embedding identification data received from said installation computer in said software application to form an identifiable software application;
20

e. means for generating a cryptographic signature for said identifiable software application; and
25

f. means for embedding said cryptographic signature in said identifiable software application to form said identifiable and authenticable software application.

- 33 -

16. A software installation computer for receiving an identifiable and authenticable software application distributed by a distribution computer:
- 5 a. a communications link between said software installation computer and said software distribution computer;
 - b. a storage device for storing identification data, and for storing an installed software application;
 - 10 c. a computer communications interface in communication with said link, for transferring said identification data, and for receiving said identifiable and authenticable software application, the identifiable and authenticable software application having embedded therein the
 - 15 identification data, and a cryptographic signature of the identifiable and authenticable software application;
 - 20 d. means for installing said identifiable and authenticable software application on said computer storage device.

17. A software distribution computer for distributing an identifiable and authenticable software application from a distribution computer to an installation computer comprising:
- a distribution computer;
 - 25 an installation computer;
 - a communications link between said installation computer and distribution computer;
 - said distribution computer comprising:

- 34 -

a. distribution computer storage device for storing a software application for distribution;

5 b. a distribution computer communications interface in communication with said link, for transferring an identifiable and authenticable software application to said installation computer and for receiving identification data from said installation computer;

10 c. means for embedding identification data received from said installation computer in said software application to form an identifiable software application;

d. means for generating a cryptographic signature for said identifiable software application; and

15 e. means for embedding said cryptographic signature in said identifiable software application to form an identifiable and authenticable software application;

said installation computer comprising:

20 a. an installation computer storage device for storing said identification data, and for storing an installed software application;

25 b. an installation computer communications interface in communication with said link, for transferring said identification data to said distribution computer and for receiving said identifiable and authenticable software application from said distribution computer; and,

d. means for installing said software application on said installation computer storage device.

High-Level ESD Components

FIG. 1

1/10

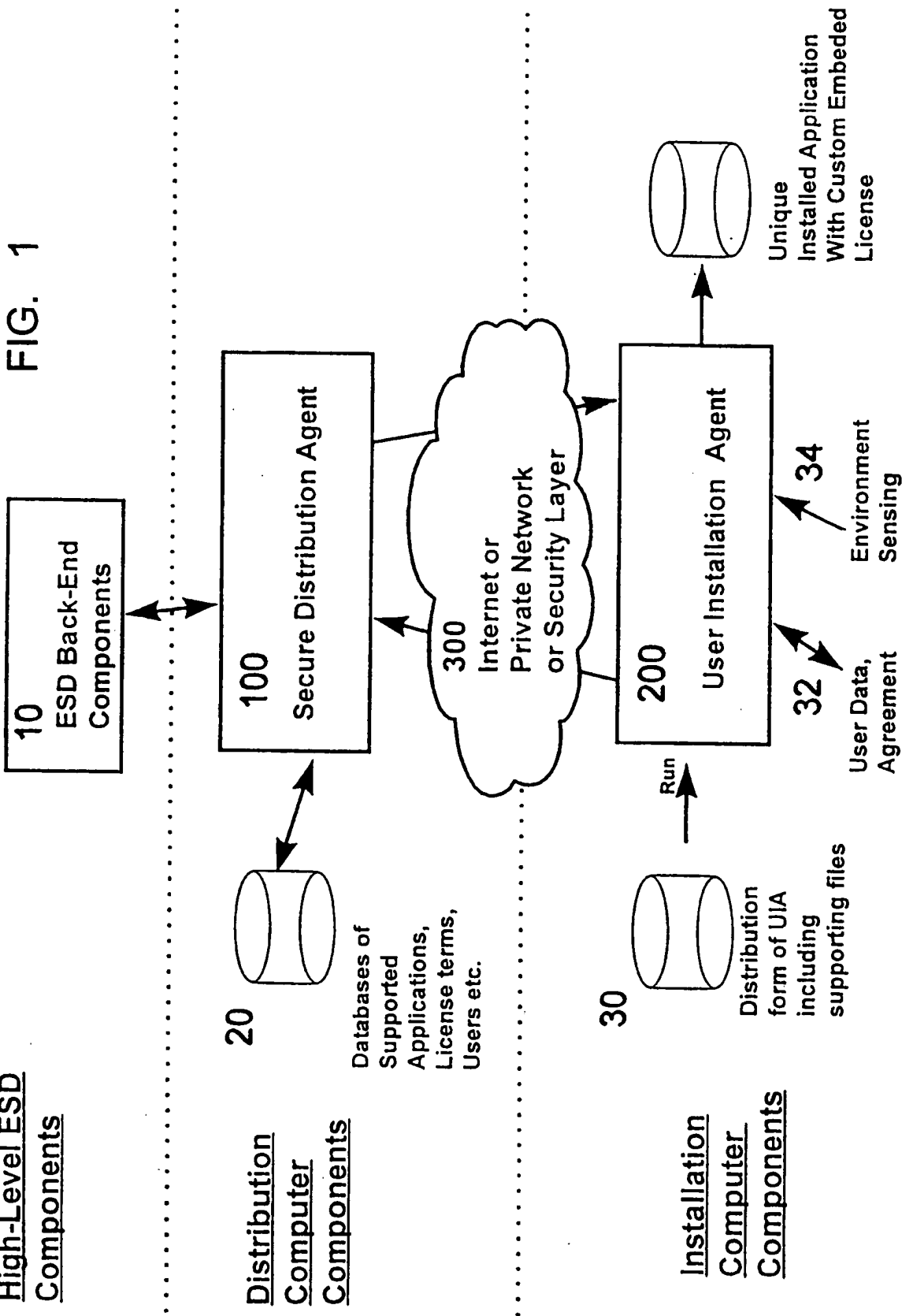
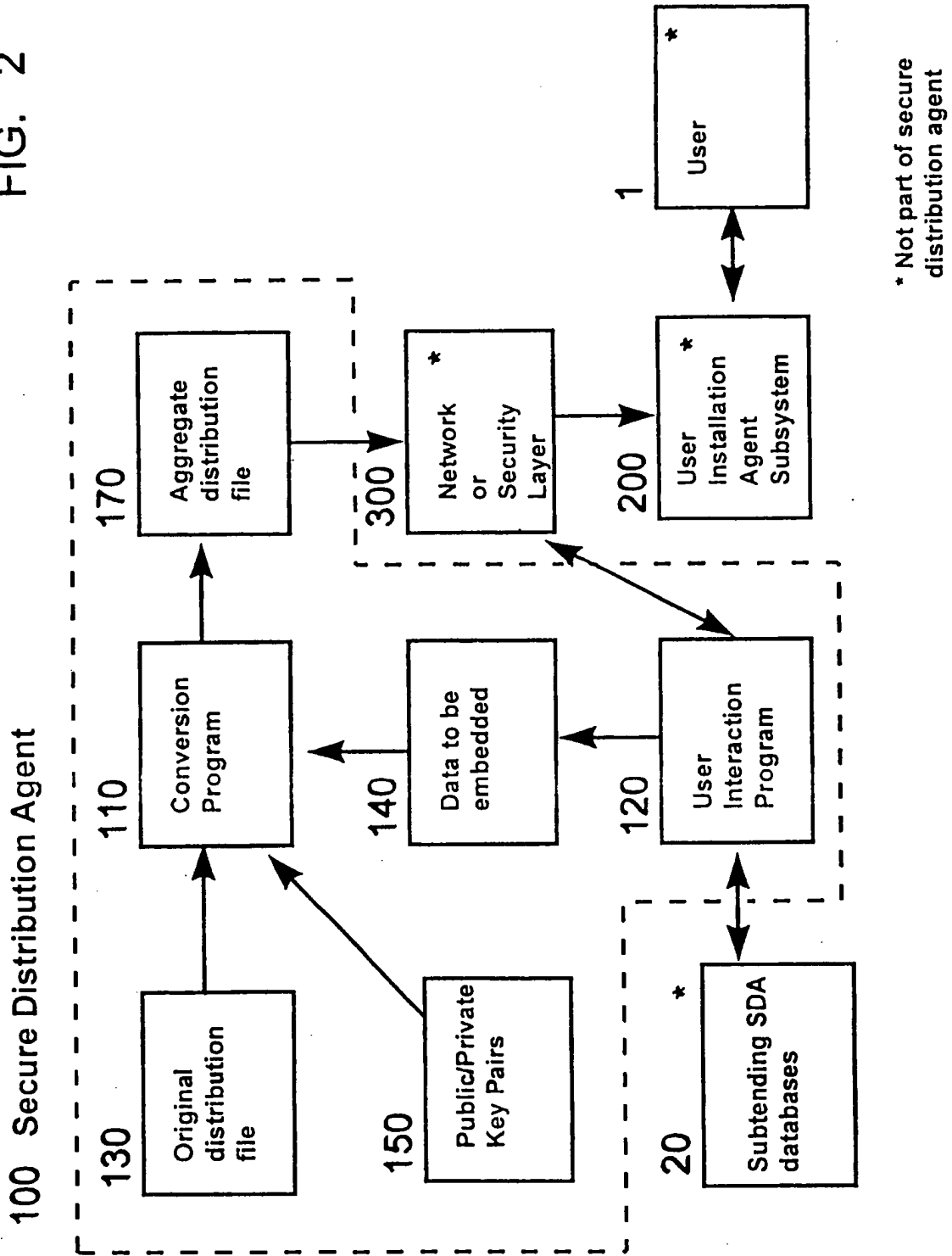


FIG. 2



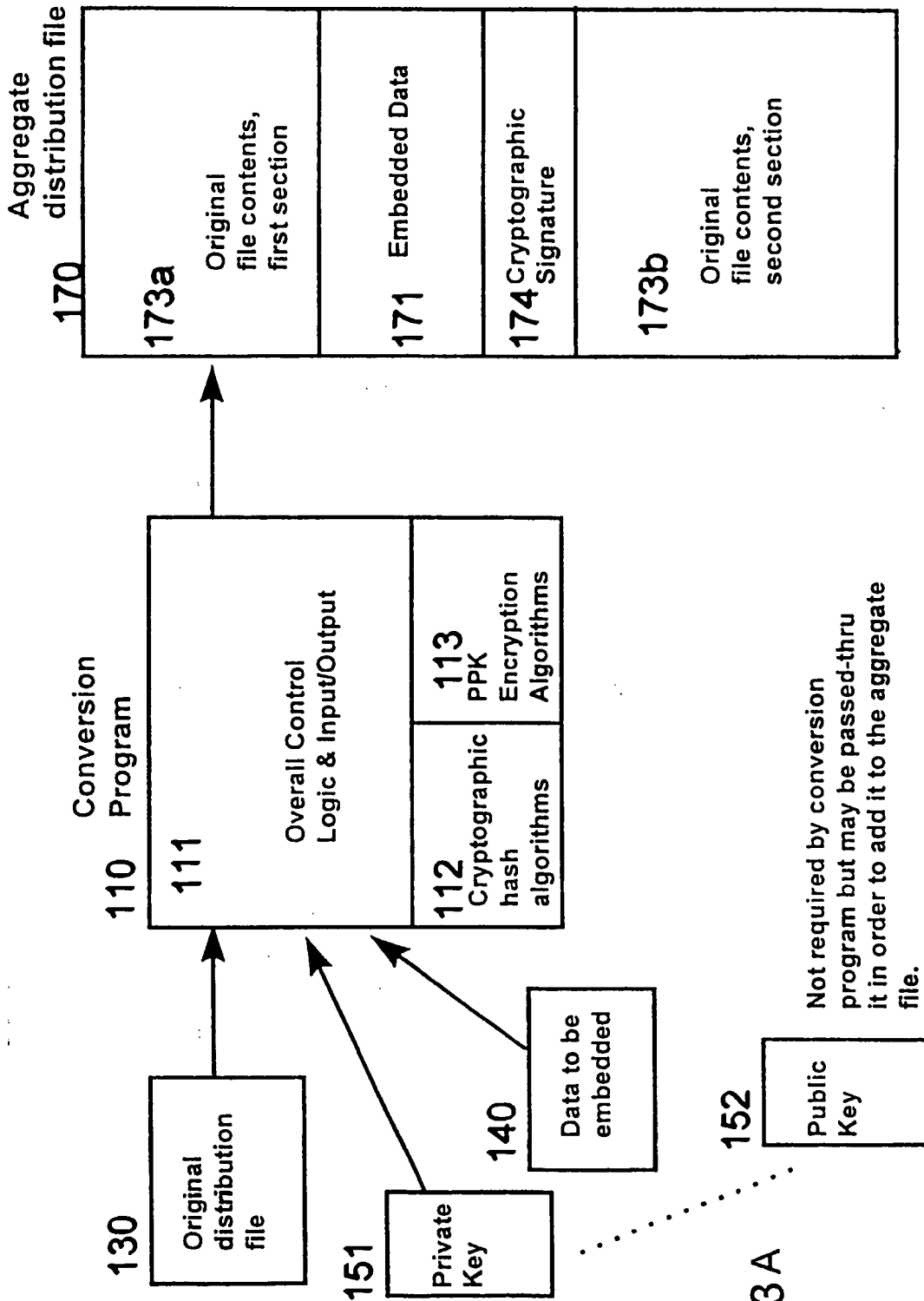


FIG. 3A

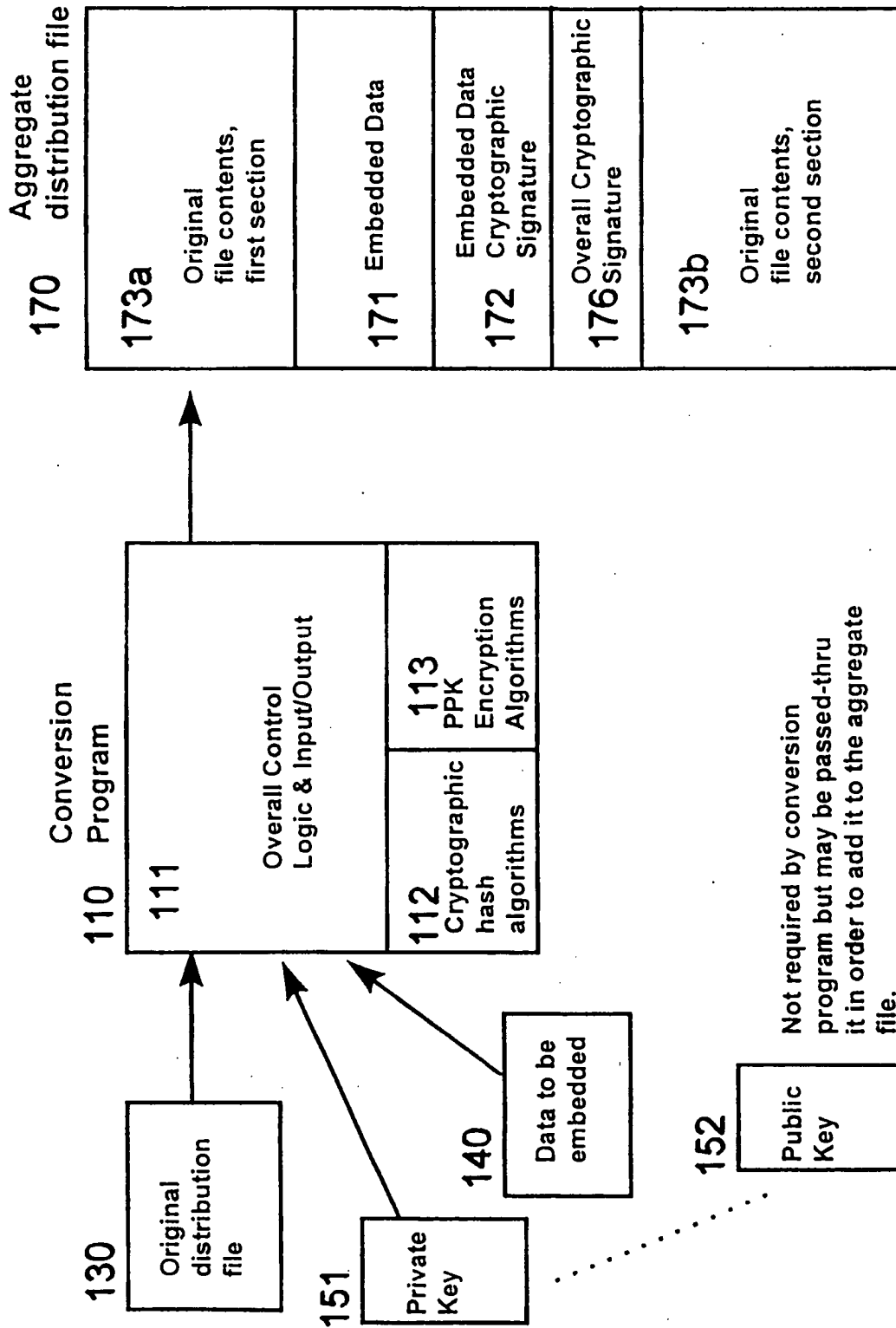


FIG. 3B

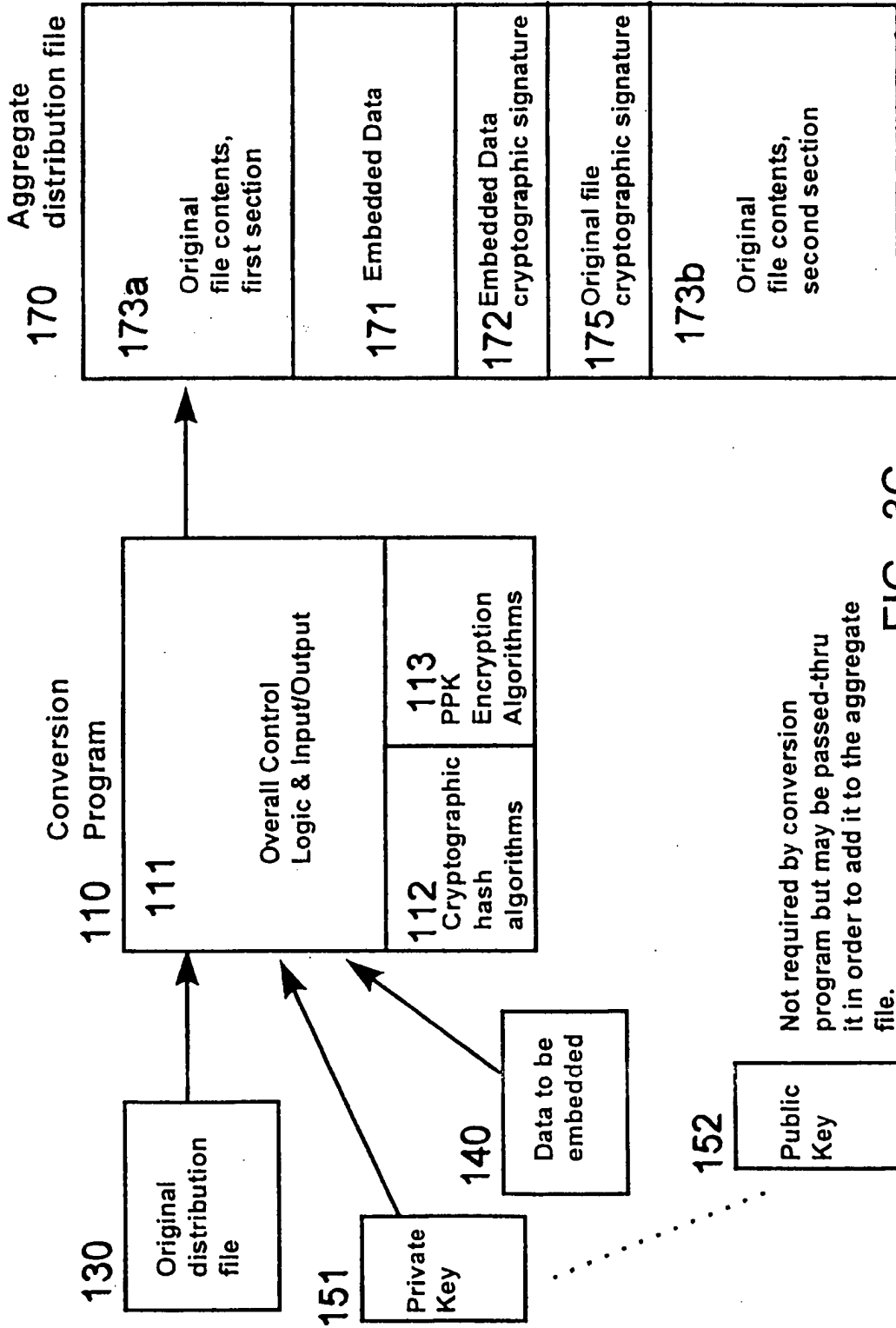


FIG. 3C

6/10

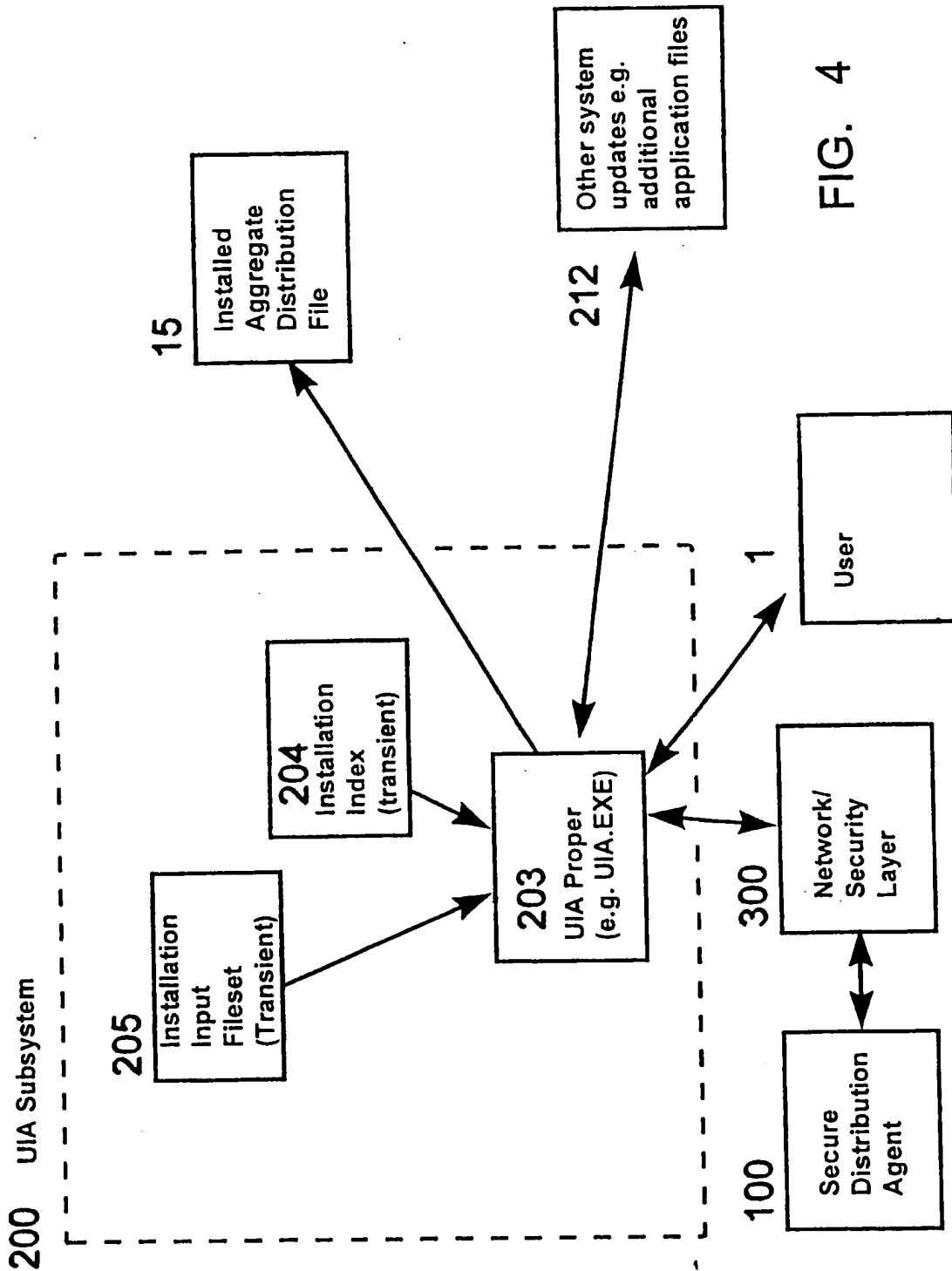
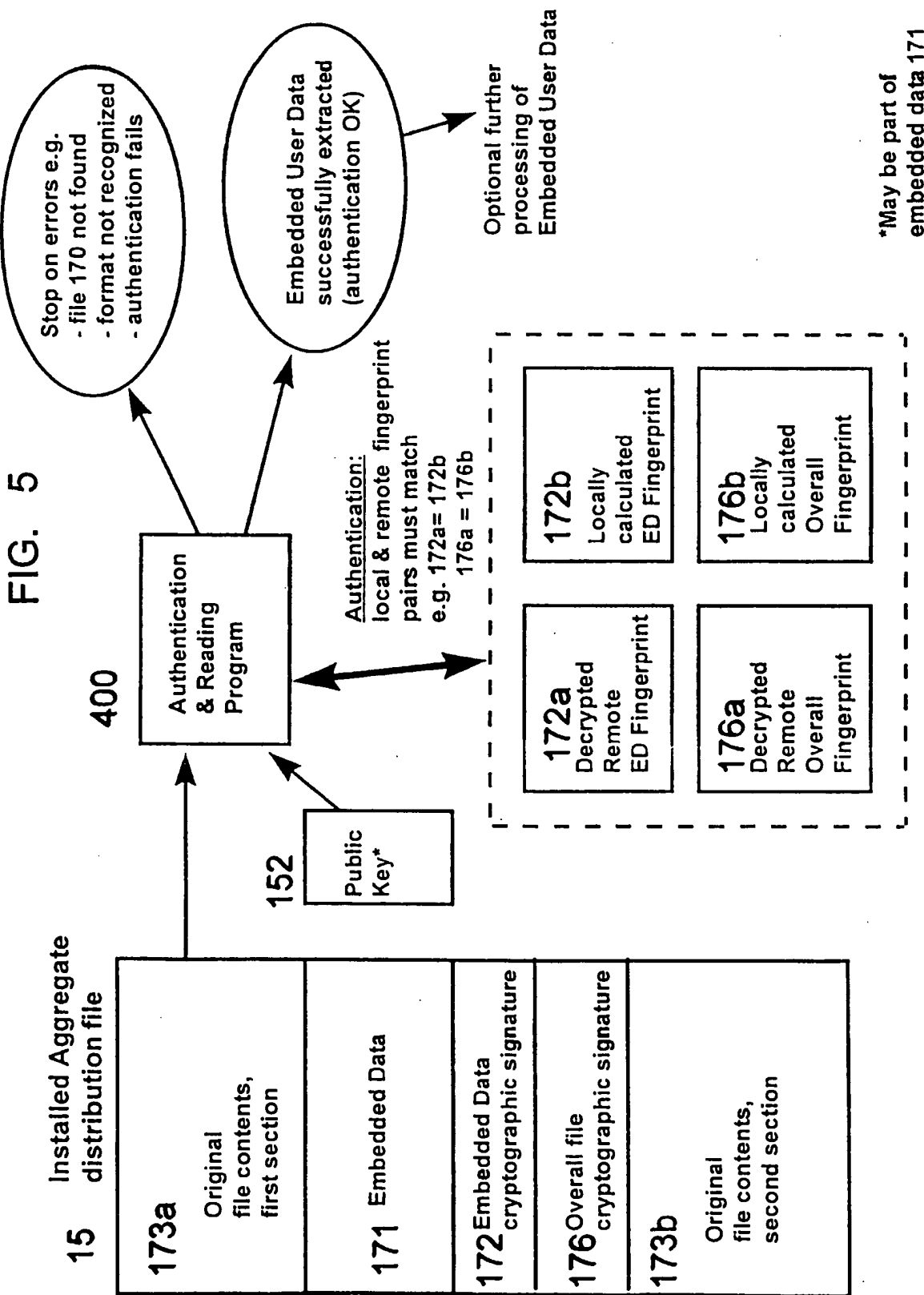


FIG. 4

7/10



8/10

FIG. 6

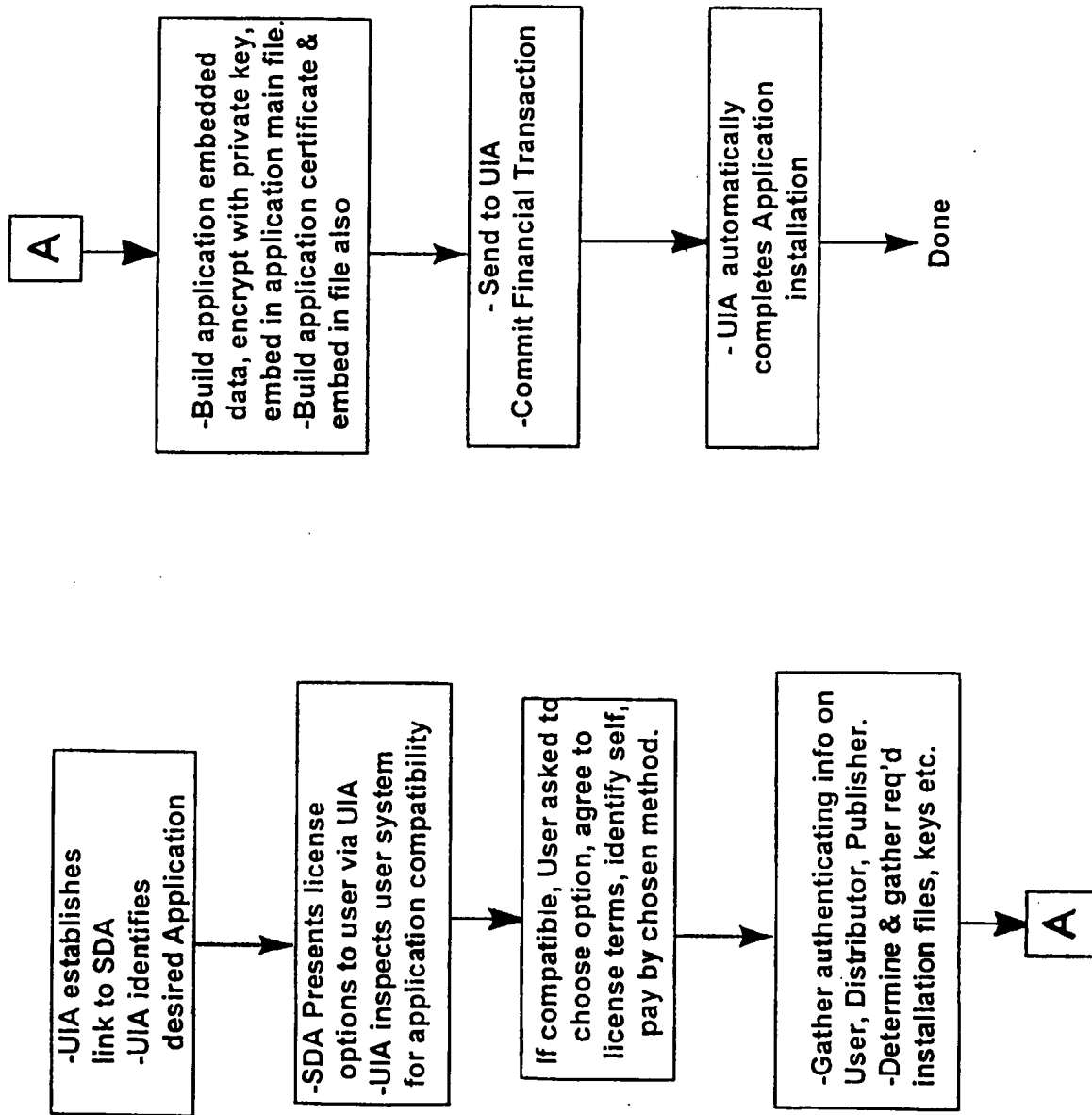
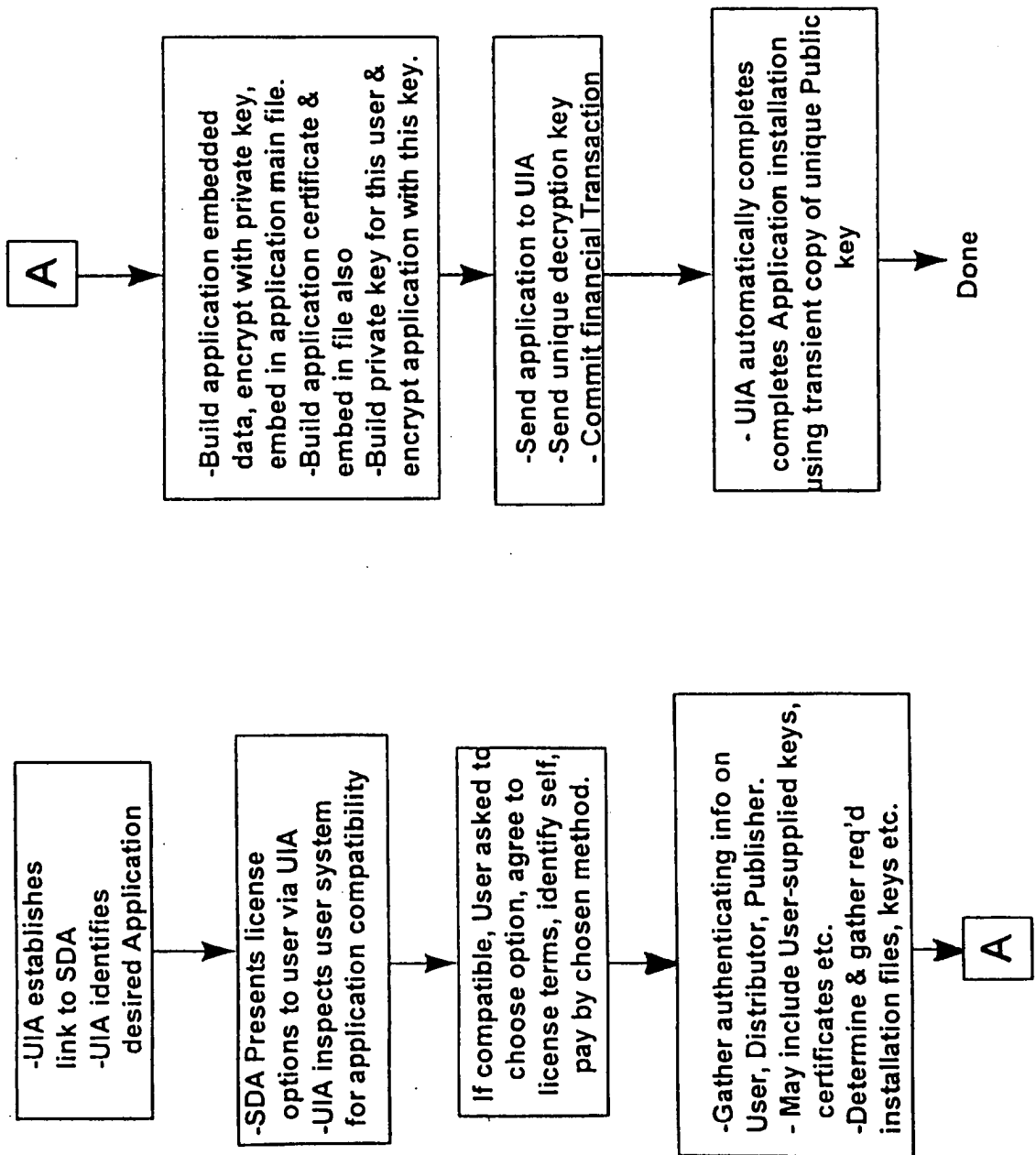
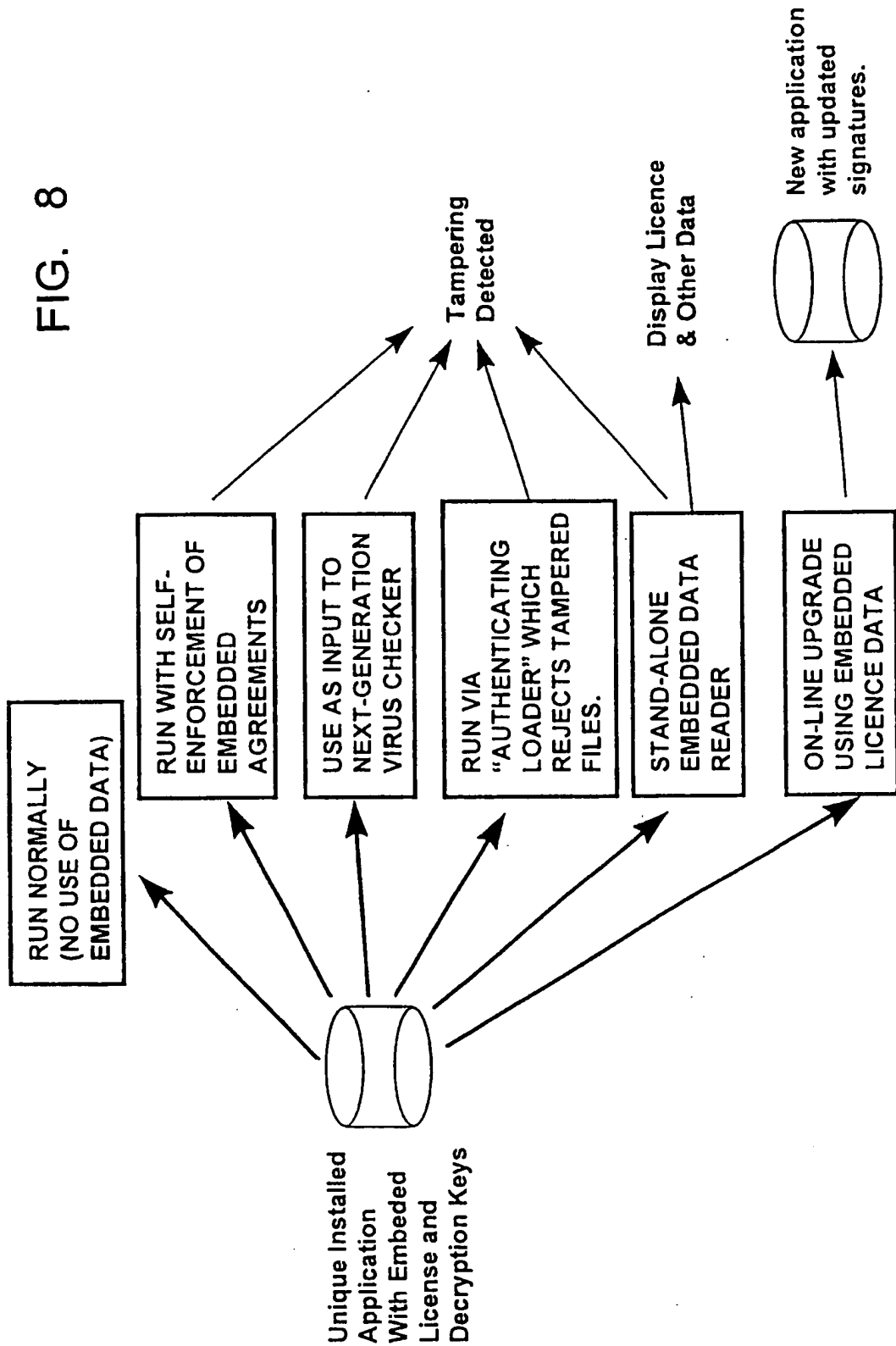


FIG. 7



10/10

FIG. 8



INTERNATIONAL SEARCH REPORT

International Application No
PCT/CA 98/00241

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6 G06F1/00

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED
Minimum documentation searched (classification system followed by classification symbols)
IPC 6 G06F

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	EP 0 686 906 A (SUN MICROSYSTEMS INC) 13 December 1995	1-6, 8, 10-13, 15-17
A	see column 4, line 1 - column 5, line 5; figures 4,6A,6B	7,9,14
A	<p style="text-align: center;">---</p> LEIN HARN ET AL: "A SOFTWARE AUTHENTICATION SYSTEM FOR INFORMATION INTEGRITY" COMPUTERS & SECURITY INTERNATIONAL JOURNAL DEVOTED TO THE STUDY OF TECHNICAL AND FINANCIAL ASPECTS OF COMPUTER SECURITY, vol. 11, no. 8, 1 December 1992, pages 747-752, XP000332279 see page 750, left-hand column, paragraph 2 see page 750, left-hand column, paragraph 8 - right-hand column, paragraph 5 <p style="text-align: center;">---</p>	1-17
	-/--	

<input checked="" type="checkbox"/> Further documents are listed in the continuation of box C.	<input checked="" type="checkbox"/> Patent family members are listed in annex.
<p>* Special categories of cited documents:</p> <p>"A" document defining the general state of the art which is not considered to be of particular relevance</p> <p>"E" earlier document but published on or after the international filing date</p> <p>"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)</p> <p>"O" document referring to an oral disclosure, use, exhibition or other means</p> <p>"P" document published prior to the international filing date but later than the priority date claimed</p>	
<p>"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention</p> <p>"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone</p> <p>"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.</p> <p>"&" document member of the same patent family</p>	
Date of the actual completion of the international search 8 July 1998	Date of mailing of the international search report 15/07/1998
Name and mailing address of the ISA European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Authorized officer Moens, R

1

INTERNATIONAL SEARCH REPORT

International Application No

PCT/CA 98/00241

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 5 509 074 A (CHOUDHURY ABHIJIT K ET AL) 16 April 1996 cited in the application see abstract ---	1-17
A	EP 0 717 337 A (IBM) 19 June 1996 see column 1, line 1 - column 2, line 40 -----	1, 8, 12, 15-17

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No

PCT/CA 98/00241

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
EP 0686906 A	13-12-1995	US 5724425 A JP 8166879 A	03-03-1998 25-06-1996
US 5509074 A	16-04-1996	CA 2137065 A EP 0665486 A JP 7239828 A	28-07-1995 02-08-1995 12-09-1995
EP 0717337 A	19-06-1996	JP 8221268 A US 5745678 A	30-08-1996 28-04-1998



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

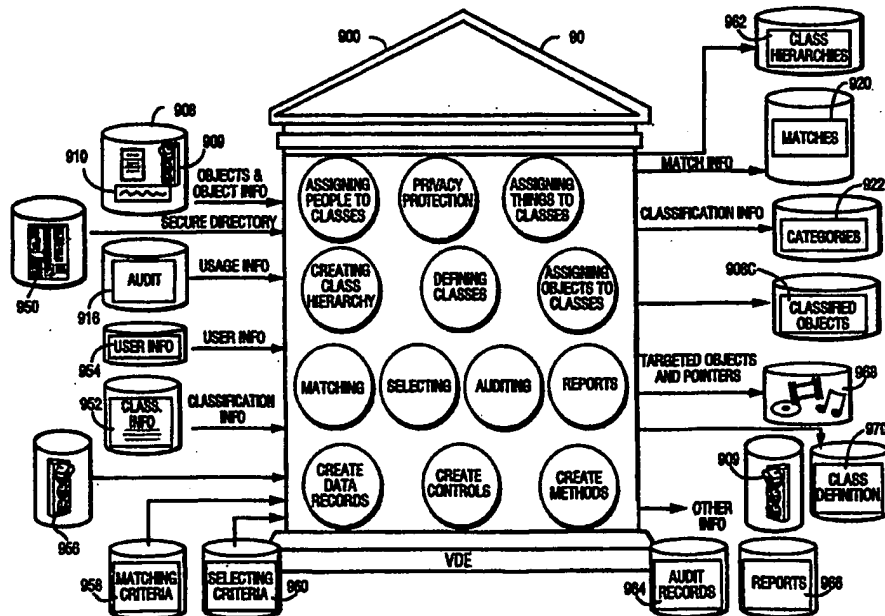
<p>(51) International Patent Classification ⁶ : G06F 17/60</p>	<p>A2</p>	<p>(11) International Publication Number: WO 99/24928 (43) International Publication Date: 20 May 1999 (20.05.99)</p>
<p>(21) International Application Number: PCT/US98/23648 (22) International Filing Date: 6 November 1998 (06.11.98) (30) Priority Data: 08/965,185 6 November 1997 (06.11.97) US (71) Applicant: INTERTRUST TECHNOLOGIES CORP. [US/US]; 460 Oakmead Parkway, Sunnyvale, CA 94086 (US). (72) Inventors: SHEAR, Victor, H.; 5203 Battery Lane, Bethesda, MD 20705 (US). VAN WIE, David, M.; Apartment 216, 965 East El Camino Real, Sunnyvale, CA 94087 (US). WEBER, Robert, P.; 215 Waverley Street #4, Menlo Park, CA 94025 (US). (74) Agent: FARIS, Robert, W.; Nixon & Vanderhye P.C., 8th floor, 1100 N. Glebe Road, Arlington, VA 22201 (US).</p>		<p>(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, HR, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).</p> <p>Published <i>Without international search report and to be republished upon receipt of that report.</i></p>

(54) Title: SYSTEMS AND METHODS FOR MATCHING, SELECTING, NARROWCASTING, AND/OR CLASSIFYING BASED ON RIGHTS MANAGEMENT AND/OR OTHER INFORMATION

(57) Abstract

Rights management information is used at least in part in a matching, narrowcasting, classifying and/or selecting process. A matching and classification utility system comprising a kind of Commerce Utility System is used to perform the matching, narrowcasting, classifying and/or selecting. The matching and classification utility system may match, narrowcast, classify and/or select people and/or things, non-limiting examples of which include software objects. The Matching and Classification Utility system may use any pre-existing classification schemes, including at least some rights management information and/or other qualitative and/or parameter data indicating

and/or defining classes, classification systems, class hierarchies, category schemes, class assignments, category assignments, and/or class membership. The Matching and Classification Utility may also use at least some rights management information together with any artificial intelligence, expert system, statistical, computational, manual, or any other means to define new classes, class hierarchies, classification systems, category schemes, and/or assign persons, things, and/or groups of persons and/or things to at least one class.



FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

**SYSTEMS AND METHODS FOR MATCHING,
SELECTING, NARROWCASTING, AND/OR
CLASSIFYING BASED ON RIGHTS
5 MANAGEMENT AND/OR OTHER
INFORMATION**

FIELDS OF THE INVENTIONS

The inventions relate to electronic rights and transaction management. More particularly, the inventions relate to automated
10 systems, methods and techniques for efficiently matching, selecting, narrowcasting, categorizing and/or classifying in a distributed electronic rights and/or other event and/or transaction management environment. For example, the inventions provide electronic computer based systems, methods and techniques for matching,
15 classifying, narrowcasting, and/or selecting digital information describing people and/or other things. This matching, classifying, narrowcasting, and/or selecting can be based, at least in part, on elements of rights management information and/or one or more other categories of information -- wherein such information is used for
20 efficient, trusted event management assuring the execution of one or more controls related to, including, for example, consequences of processing such digital information describing people and/or other things. The present inventions also provide systems and methods for efficiently determining class hierarchies, classification schemes,
25 categories, and/or category schemes and/or the assignment of objects, persons and/or things to said class hierarchies, classification schemes,

categories, and/or category schemes using at least some rights management information.

BACKGROUND AND SUMMARY OF THE INVENTIONS

5 The modern world gives us a tremendous variety and range of options and choices. Cable and satellite television delivers hundreds of different television channels each carrying a different program. The radio dial is crowded with different radio stations offering all kinds of music, news, talk, and anything else one may care to listen
10 to. The corner convenience store carries newspapers from around the country, and a well stocked newsstand allows you to choose between hundreds of magazines and publications about nearly every subject you can think of. Merchandise from all corners of the world is readily available at the shopping mall or by mail order. You can pay by
15 check, in cash, or using any number of different kinds of credit cards and ATM cards.

 This tremendous variety is good, but it also presents problems. Sometimes, it is hard or inefficient for us to find what we want and need because there are too many things to evaluate and choose from,
20 and they are often located in too many places. We can waste a lot of time searching for the things we need or want at the right price, with the rights features, and at a particular time.

 Sometimes, we never find things that satisfy what we feel we need or want. This happens when we don't know what to look for,

how to look for it, or don't have the necessary assistance or tools to search successfully. For example, we may not know the best way of looking for something. Sometimes, we know what we are looking for but can't express or articulate it in ways that help us look. And
5 sometimes, we don't even know what we are looking for. You may know you need something, know its missing, but never really know how to communicate to others what you are looking for. For example, someone who speaks only English may never find resources using Japanese or Spanish. In general, we often don't have the time
10 or resources to look for all the things that would give us the most benefit or make us the most satisfied.

It's Hard To Find Mass Media Things You Want Or Need.

Figure 1A shows, as one example, how frustrating it can be to
15 find anything to watch on the hundreds of television channels that may be available. The man in Figure 1A spends a lot of time "channel surfing," trying to find something he is interested in watching. He may be moderately interested in golf, but may not like the particular golf tournament or golf players being broadcast at 7
20 o'clock on a particular channel. After flipping through other channels, he might think an action movie looks interesting only to find out after watching it for a while that he isn't really interested in it after all. A documentary on horses also seems interesting at first, but he finds it boring after watching it awhile because it doesn't give him
25 the kind of information he is interested in. The whole process can be

frustrating and he may feel he wasted a lot of time. Figure 1B shows the man getting so frustrated at the wasted time and energy that he thinks that maybe watching television is just not worth it . What the man really needs is a powerful yet efficient way to find those things
5 that most satisfy his desires -- that is, match his needs and/or his interests.

Our Mail Overloads Us With Things We Don't Want or Need

The same thing can happen with information sent to us in the
10 mail. It can be fun to receive some kinds of mail, such as personal letters, or magazines and catalogs on topics of personal interest. Certain other mail, such as bills, may not be fun but are usually important. Unfortunately, our mailboxes are typically overflowing with yet another kind of mail commonly referred to as "junk mail."
15 The person in Figure 2 finds his mailbox stuffed to the overflowing point with mail he never asked for and has absolutely no interest in. Most of this junk mail ends up unread and in the trash. However, it can take a long time to sort through all this mail to be sure you are only throwing out only the junk mail and not the good mail you are
20 interested in or need. For example, it's sometimes hard to distinguish credit card bills from offers for new credit cards you don't need or want. Wouldn't it be useful if your mail could be automatically "cleaned" of the mail you had no interest in and you received only the mail you wanted or needed?

Sorting through things to identify things you might want, then selecting what you actually want, can be a frustrating and time consuming experience. For example, it wastes the time of the person who receives the junk mail, and it also wastes the time, money and
5 effort of the people who spend their money to send mail to people hoping that they will buy their products.

As frustrating as finding and selecting may be to consumers, they often create even greater problems for businesses and people who want to locate or provide information, goods and services. It is
10 often said, that in the world of business, "Information is Power" and "efficiency is the key to success." To find or sell the most relevant or useful information and to provide the ability to most efficiently allow business to operate at its best, we need easy-to-use tools that can help us navigate, locate, and select what matches our interests. In the
15 modern world, it is often difficult to find out what different people like, and to supply people with the opportunity to select the best or most satisfying choices.

Past attempts outside the computer world to match up people with information, goods and/or services have had limited success.
20 For example, attempts to "target" mass mailings may increase the chance that they will go to people who are interested in them, but the entire process is still very wasteful and inefficient. It is considered a good success rate to match the interests of only a few percent of the recipients of "junk" mail. Telemarketing campaigns that use the

telephone to reach potential consumers can be very expensive, very annoying to consumers who are not interested in the products being marketed, and very costly and inefficient. A much more ideal situation for all concerned is enabling businesses to send information
5 only to individual consumers likely to find the information interesting, desirable, convincing, and/or otherwise useful. That way, businesses save time and money and consumers aren't unproductively hassled by information, phone calls, junk mail, junk e-mail and the like. However, right now it is extremely difficult to accomplish this
10 goal, and so businesses continue to annoy consumers while wasting their own time, money, and effort.

Because of the Vast Amount of Information Available, Even Systems that Provide a High Degree of Organization May Be Difficult to Use or Access

15 You can find yourself wasting a lot of time finding things -- even in places where finding things is supposed to be easy. For example, a library is a place where you can find all sorts of useful information but can also waste a lot of time trying to find what you are looking for. Modern libraries can be huge, containing tens or
20 even hundreds of thousands or millions of different books, magazines, newspapers, video tapes, audio tapes, disks, and other publications. Most libraries have an electronic or manual card catalog that classifies and indexes all of those books and other materials. This classification system is useful, but it often has significant
25 limitations.

For example, normally a card catalog will classify materials based only on a few characteristics (for example, general subject, author and title). The boy in Figure 3 is looking for information on American League baseball teams during World War II for a high school report. The card catalog led to the general subject of baseball and other sports, but, looking at the catalog, he can't identify any books that seem to provide the specific information he wants to see, so he must rely on books classified as "histories of sports" or "histories of baseball." He can spend lots of time looking through the books on the shelves, going back to the card catalog, and going back to the shelves before he finds a reference that's reasonably helpful. He may need to go ask an expert (the librarian) who is familiar with the books the library has on sports and may know where to look for the information. Even then, the boy may need to flip through many different books and magazines, and look in many different places within the library before he finds the information he is looking for.

Finding Products You Want or Need Can Be Very Difficult and Time Consuming

The same kind of frustrating experience can happen when you shop for a particular kind of item. While some people enjoy shopping, and have fun seeing what is in various stores, many people dislike spending time shopping, searching for the best or most affordable item. And sometimes even people who like to shop don't have the time to shop for a specific item.

For example, the man in Figure 4 goes into a shopping mall looking for a tie to fit very tall people. He didn't wear a tie to work that day, but, at the last minute, an important meeting was scheduled for later that day and he needs to dress up. The shopping mall has a large variety of stores, each selling a range of merchandise. But the man may only have a short time to look. For example, he may be on his lunch break, and needs to get back to work soon. He can't spend a lot of time shopping. He may therefore need to rely on tools to help him identify where he wants to buy the tie. Perhaps he uses a mall directory that classifies the different stores in terms of what kinds of merchandise they sell (for example, clothing, books, housewares, etc.). Perhaps he asks at the malls help desk staffed by "experts" who know what is available in the shopping mall. But even these resources may not tell him where to buy Italian silk ties that are discounted and cost \$20. So he does the best he can with the available resources.

These Problems Are Worse in the Digital World

The electronic or digital world offers a rapidly growing, vast array of electronically published products and services. For example, computer superstores have a dizzying array of different software products. Furthermore, music is now published primarily in digital form on optical disks, and video will soon be published that way too. And, of particular interest related to certain of the inventions described by this document, the Internet now has millions of home pages with an overwhelmingly variety and quantity of digital

information, and, these millions of home pages, in turn, point or "link" to millions of other web pages as well.

Today, for example, you can use the Internet to:

- 5 • read electronic newspapers, books and magazines and see them on your computer screen;
- get music in electronic form and play it using your computer;
- send and receive electronic mail all over the world;
- 10 • download reports and other information compiled by governments, companies, industries, universities, and individuals;
- watch videos and animations;
- play games with "cyber-friends" located around the world;
- 15 • chat with individuals and groups who share at least some interests in common;
- participate in "virtual reality" worlds, games, and/or experiences;
- (offer to) buy, and/or (offer to) sell nearly anything;
- 20 and
- conduct electronic transactions and commerce.

Today on the Internet and you can also find nearly anything and everything you can possibly imagine, although finding exactly what you really want may be time consuming and frustrating. This is

because the Internet and World Wide Web provide perhaps the best example of an environment that is particularly hard to navigate. There are an overwhelming number of choices -- too many to easily relate to or understand -- and many of which are terribly hard to find, even using the various Web searching "engines." The Internet is particularly exciting because it has the potential to provide to nearly everyone access to nearly every kind of information. Information can also come from an almost limitless variety of sources. But today, so much information on the Internet is superficial or useless, and too many choices can be more a curse than a blessing if you don't have meaningful, easy ways to eliminate all but a relatively few choices. And the situation will only become much worse as more Web sites appear, and as digital information is distributed in "objects" or "containers" providing enhanced security and privacy but possibly more difficult access and identifiability.

As time passes, more and more valuable and desirable information will be available in digital containers. However, unless tools are developed to solve the problem, there will be no efficient or satisfying means to sort through the potentially trillions of digital containers available on tens of millions of Web pages, to find containers satisfying a search or fulfilling an information need. Furthermore, existing information searching mechanisms typically provide no way to readily perform a search that matches against underlying commercial requirements of providers and users.

It Will Be Difficult to Find Rights Management Scenarios Matching Your Requirements

If, for example, you have an auto repair newsletter and you want to create an article containing information on auto repair of Ford
5 Bronco vehicles, you may wish to look for detailed, three dimensional, step-by-step "blow-up" mechanical images of Ford Bronco internal components. Perhaps these are available from hundreds of sources (including from private individuals using new, sophisticated rendering graphics programs, as well as from
10 engineering graphics firms). Given the nature of your newsletter, you have decided that your use of such images should cost you no more than one penny to redistribute per copy in quantities of several thousand -- this low cost being particularly important since you will have numerous other costs per issue for acquiring rights to other
15 useful digital information products which you reuse and, for example, enhance in preparing a particular issue. You therefore wish to search and match against rights management rules associated with such products -- non-limiting examples of which include:

- cost ceilings,
- 20 • redistribution rights (e.g., limits on the quantity that may be redistributed),
- modification rights,
- class related usage rights,
- category related usage rights,

- sovereignty based licensing and taxation fees,
- import and export regulations, and
- reporting and/or privacy rights (you don't want to report back to the product provider the actual identity of your end users and/or customers.

5
10 If you can't match against your commercial requirements, you may be forced to waste enormous amounts of time sifting through all of the available products matching Ford Bronco internal components - - or you may settle for a product that is far less than the best available (settling on the first adequate product that you review).

Computers Don't Necessarily Make It Easier to Find Things

15 Anyone who has ever used the Internet or the World Wide Web knows that networks, computers and electronics, when used together, do not necessarily make the overall task of finding information easier. In fact, computers can make the process seem much worse. Most Internet users will probably agree that trying to find things you are interested on the Internet can be a huge time drain. And the results can be very unsatisfactory. The rapid growth rate of information available on the Web is continually making this process of finding desired information even harder. You can spend many hours looking for information on a subject that interests you. In most cases, you will eventually find some information of value -- but even using today's advanced computer search tools and on-line directories, it can

take hours or days. With the advent of the technology advances developed by InterTrust Technologies Corp. and others, publishers will find it far more appealing to make their valuable digital information assets available on-line and to allow extractions and
5 modifications of copyrighted materials that will vastly expand the total number of information objects. This will enormously worsen the problem, as the availability of valuable information products greatly expands.

It Is Usually Hard to Find Things On the Internet

10 There are many reasons why it is difficult to find what you want on the Internet. One key reason is that, unlike a public library, for example, there is no universal system to classify or organize electronic information to provide information for matching with what's important to the person who is searching. Unlike a library, it
15 is difficult on the Internet to efficiently browse over many items since the number of possible choices may be much larger than the number of books on a library shelves and since electronic classification systems typically do not provide much in the way of physical cues. For example, when browsing library shelves, the size of a book, the
20 number of pictures in the book, or pictures on magazine covers may also help you find what you are interested in. Such physical cue information may be key to identifying desired selections from library resources. Unfortunately, most digital experiences typically do not provide such cues without actually loading and viewing the work in
25 digital form.

Thus, another reason why the electronic or digital world can make it even harder to find information than ever before has to do with the physical format of the information. The digital information may provide few or no outward cues or other physical characteristics that could help you to even find out what it is – let alone determine whether or not you are interested in it, unless such cues are provided through special purpose informational (for example, graphical) displays. On the Internet, everyone can be an electronic publisher, and everyone can organize their offerings differently -- using visual cues of their own distinctive design (e.g., location on a web page, organization by their own system for guiding choices). As one example, one publisher might use a special purpose graphical representation such as the video kiosk to support an electronic video store. Other publishers may use different graphical representations altogether.

Historically, there has been no particular need for consistent selection standards in conventional, non-electronic store based businesses. Indeed, it is often the unique display and choice selection support for customers' decision processes that make the difference between a successful store and a failure. But in the electronic world--where your choice is not among a few stores but rather is a choice among potentially thousands or even millions of possibly useful web sites and truly vast numbers of digital containers -- the lack of a consistent system for describing commercially significant variables that in the "real" world may normally be provided by the

display context and/or customized information guidance resource (catalog book, location of goods by size, etc.) seriously undermines the ability of digital information consumers to identify their most desirable choices.

5 Adding to this absence of conventional cues, the enormity of available choices made available in cyberspace means that the digital information revolution, in order to be practical, must provide profoundly more powerful tools to filter potentially desirable opportunities from the over abundance of choices. In sum, the
10 absence of the ability to efficiently filter from a dimensionally growing array of choices, can completely undermine the value of having such a great array of choices.

 In the "real" world, commercial choices are based on going to the right "store" and using the overall arrays of available information
15 to identify one's selection. However, as information in digital and electronic form becomes more and more important, the problem of relating to the vast stores of information will become a nightmare. For example, picture yourself in a store where each shopping aisle is miles long, and each item on the shelf is packaged in the same size
20 and color container. In an actual store, the product manufacturers put their products into brightly colored and distinctively shaped packages to make sure the consumer can readily find and select their product. These visual cues distinguish, for example, between a house brand

and a specific name brand, between low fat and regular foods, and between family size and small size containers.

On the Internet, a digital "store" is likely to be many stores with vast resources integrating products from many parties. If you were
5 limited to conventional classification and matching mechanisms, you would be unable to sift through all the material to identify the commercially acceptable, i.e., an item representing the right information, at the right price, providing license rights that match your interests. Certainly, if each digital package looks the same, you
10 are at a loss in making reasonable decisions. You can't tell one from another just by looking at it.

While information written on the "outside" of a digital package may be useful, you simply don't have the time to read all the packages, and anyway, each packager may use different words to
15 describe the same thing and the descriptions may be difficult to understand. Some people may write a lot of information on the outside of their package, and others may write little or nothing on the outside of the package. If there is no universal system agreed upon by everyone for defining what information should be written on the
20 outside of the package and how it should be formatted, using such a store would be painfully difficult even if you could limit the number of choices you were evaluating.

**There is a Need For Efficient and Effective Selection
Based, at Least in Part, on Rights Management
Information**

5 Unlike a real store where all breakfast cereals are shelved
together and all soft drinks are in the same aisle, there may be no
single, universal way to display the organization of all of the
information in a "digital store" since, by its nature, digital information
frequently has many implications and associated rules. For example,
there now exist highly developed rights management systems such as
10 described in U.S. Patent application Serial No. 08/388,107 of Ginter
et al., filed 13 February 1995, for "Systems And Methods For Secure
Transaction Management And Electronic Rights Protection (hereafter
"Ginter et al") – the entire disclosure (including the drawings) of
which is expressly incorporated into this application as if expressly
15 set forth herein. Many rules associated with any given piece of digital
information may, combinatorially, given rise to many, very different,
commercial contexts that will influence the use decisions of different
potential users in many different ways (e.g., cost, auditing, re-use,
redistribution, regulatory requirements, etc.).

20 No readily available systems developed for the digital
information arena provide similarly satisfying means that describe the
many commercial rules and parameters found in individual custom
catalogs, merchandise displays, product specifications, and license
agreements. Further, no readily available mechanisms allow

"surfing" across vast choice opportunities where electronic matching can single out those few preferred items.

As one example, picking an appropriate image may involve any or all of the following:

- 5 • price,
- republishing (redistribution) rights,
- rights to extract portions,
- certified usable in certain sovereignties (e.g.,
 pornographic content not allowed in Saudi Arabia),
- 10 • size,
- format, etc.,
- use and reuse administrative requirements (e.g., which
 clearinghouses are acceptable to rightsholders, what is
 the requirement for reporting usage information – is the
15 name of your customer required, or only the use class(es)
 or none -- is advertising embedded), and
- other features.

No previously readily available technology allows one to efficiently make selections based on such criteria.

20 By their nature, and using the present inventions in combination with, amongst other things, "Ginter et al", the packages in a digital store may be "virtual" in nature -- that is, they may be all

mixed up to create many, differing products that can be displayed to a prospective customer organized in many different ways. This display may be a "narrowcasting" to a customer based upon his matching priorities, available digital information resources (e.g., repository, property, etc.) and associated, available classification information. In the absence of an effective classification and matching system designed to handle such information, digital information of a particular kind might be just about anywhere in the store, and very difficult to find since the organization of the stores digital information resources have not been "dynamically" shaped to the matching interests of the potential customer.

These Inventions Solve These Problems

The present inventions can help to solve these problems. It can give you or help you to find the things you like, need or want. For example, it can deliver to you, (including narrowcasting to you), or help you to find:

- things that match your interests;
- things that match your lifestyle;
- things that match your habits;
- things that match your personality;
- things you can afford and/or accept your preferred payment method;
- things that help you in your work;
- things that help you in your play;

- things that help you to help others;
- things that other people who are similar to you have found helpful,
- things that fulfill the commercial objective or
- 5 requirements of your business activities; and
- things that will make you happy and fulfilled.

The present inventions can expand your horizons by helping you to find interesting or important things, things that you enjoy, things that optimize your business efficiency, and things that help you

10 make the best digital products or services you can -- even if you didn't know precisely what or how to look for what you may need. It can also help you by allowing things you didn't know existed or know enough to look for -- but that you may be interested in, want or need -- to find you.

15 **The Present Inventions Can Use "Metaclasses" to Take Multiple Classifications Into Account**

In some areas, multiple classifications may already exist and thus it is important for a consumer to be able to find what he or she is looking for while taking into account not only that there may be

20 multiple classifications, but also that some classifications may be more authoritative than others. For example, Consumer Reports may be more authoritative on certain topics than more casual reviews published, for example, in the local weekly newspapers.

As another example, consider a book that rates restaurants according several factors, including, for example, quality, price, type of food, atmosphere, and location. In some locations there may be many guides, but they may review different sets of restaurants. One
5 guide may rate a particular restaurant highly while one or more others may consider it average or even poor. Guides or other sources of ratings, opinions, evaluations, recommendations, and/or value may not be equally authoritative, accurate, and/or useful in differing circumstances. One consumer may consider a guide written by a
10 particular renowned expert to be more authoritative, accurate, and/or useful than a guide reflecting consumer polls or ballots. However, another consumer may prefer the latter because the second consumer may perceive the tastes of those contributing opinions to be closer to his or her own tastes than those of the experts.

15 In accordance with the present inventions, a person may be able to find a restaurant that meets specified criteria – for example, the highest quality, moderately priced Cantonese and/or Hunan Chinese food located in Boston or Atlanta – while weighting the results of the search in favor of reviews from travel books rather than from the local
20 newspapers. As this example indicates, the searching may be according to class of authoritative source (and/or classes sources considered authoritative by the consumer) instead of weighting individual reviewers or sources. Thus in accordance with the present inventions, search may be performed at least in part based on classes
25 of classes, or "metaclasses."

The Present Inventions Can Make Choices Easier

One simple way to look at some examples of the present inventions is as a highly sensitive electronic "matchmaker" that matches people or organizations with their best choices, or even
5 selects choices automatically. The present inventions can match people and/or organizations with things and/or services, things with other things and/or services, and/or even people with other people. For example, the matching can be based on profiles that are a composite of preference profiles of one or more specific users, one or
10 more user groups, and/or organizations -- where the contribution of any given specific profile to the composite profile may be weighted according to the specific match circumstances such as the type and/or purpose of a given match activity.

Figure 5 shows a simplified example of an electronic
15 matchmaker that can match up two people with like interests. Sarah loves hiking, country and western music, gardening, movies and jogging. Mark loves movies, hiking, fast cars, country and western music, and baseball. The electronic matchmaker can look at the interests, personalities and/or other characteristics of these two people
20 and determine that they are compatible and should be together -- while maintaining, if desired, the confidentiality of personal information. That is, unlike conventional matchmaking services, the present inventions can keep personal information hidden from the service provider and all other parties and perform matching within a

protected processing environment through the use of encryption and protected processing environment-based matching analysis.

For example, certain matching of facts that are maintained for authenticity may be first performed to narrow the search universe.

5 Then, certain other matching of facts that are maintained for secrecy can be performed. For example, matching might be based on shared concerns such as where two parties who have a given disability (such as cancer or HIV infection) that is certified by an authority such as a physician who is certified to perform such certification; or the same
10 income level and/or bank account (as certified by an employer and/or financial authority such as a bank). Some or all of such secret information may or may not be released to matched parties, as they may have authorized and/or as may have been required by law when a match is achieved (which itself may be automatically managed within
15 a protected processing environment through the use of controls contributed by a governmental authority).

Figure 5A shows an electronic matchmaker that matches an electronic publisher with mystery stories for his quarterly electronic mystery anthology, where the matching is based on price,
20 redistribution rights, editing rights, attribution requirements (attributing authorship to the author), third party rating of the writers quality, length of story, and/or the topical focus of the story (for example). Here, rule managed business requirements of publisher and writers are matched allowing for great efficiency in matching,

coordination of interests, and automation of electronic business processes and value chain activities.

The convenience of the "electronic matchmaker" provided in accordance with the present inventions extends to commerce in physical goods as well -- as illustrated in Figure 5b. In this non-limiting example, the electronic matchmaker is communicating to the consumer via the Internet and World Wide Web. The matchmaker has found the lowest quoted price for a Jeep sports utility model given, in this one example, a multitude of factors including:

- 10 • model,
- color,
- options package,
- availability, and
- discounts resulting from the consumer's membership in
- 15 certain classes (such as membership in the American Association of Retired Persons, membership in the American Automobile Association, and being a graduate of Stanford University).

Membership in these associations and alumni status may be conveyed or indicated by possession of a special electronic document called a "digital certificate," "membership card," and/or other digital credential that warrants or attests to some fact or facts.

Thus, the electronic matchmaker provided in accordance with these inventions can also match people with things. Figure 6 shows two people, Harry and Tim. Harry loves sports most of all, but also wants to know a little about what is going on in the business world.

5 The business world is most important to Tim, but he likes to keep up with the baseball scores. The electronic matchmaker in accordance with these inventions can learn about what Harry and Tim each like, and can provide information to a publisher so the publisher can narrowcast a newspaper or other publication customized for each of

10 them. A newspaper company can narrowcast to Harry lots of sports information in his newspaper, and it can narrowcast to Tim mostly business information in his newspaper. In another example, Harry's newspaper may be uniquely created for him, differing from all other customized newspapers that emphasize sports over business

15 information. But information that Harry and Tim respectively want to maintain as authentic or secret can be managed as such.

The electronic matchmaker can also match things with other things. Figure 7 shows how the electronic matchmaker can help a student put together a school project about big cats. The electronic

20 matchmaker can help the student locate and select articles and other material about various kinds of big cats. The electronic matchmaker can, for example, determine that different articles about tigers, lions and cheetahs are all about big cats – but that articles about elephants and giraffes are not about big cats. If there is a charge for certain

25 items, the electronic matchmaker can find only those items that the

student can afford, and can make sure the student has the right to print pictures of the big cats. The electronic matchmaker can help the student to collect this information together so the student can make a colorful poster about big cats.

5 The electronic matchmaker can match up all sorts of different kinds of things. Figure 8 shows the electronic matchmaker looking at three different objects. The matchmaker can determine that even though objects A and C are not identical, they are sufficiently similar that they should be grouped together for a certain purpose. The
10 electronic matchmaker can determine that for this purpose, object B is too different and should not be grouped with objects A and C. For a different purpose, the electronic matchmaker may determine that objects A, B and C ought to be grouped together.

15 **The Present Inventions Can Make Use of Rights Management Information**

 How does the electronic matchmaker find out the information it needs to match or classify people and things? In accordance with a feature provided by these inventions, the electronic matchmaker gets information about people and things by using automated,
20 computerized processes. Those processes can use a special kind of information sometimes known as rights management information. Rights management information may include electronic rules and/or their consequences. The electronic matchmaker can also use information other than rights management information.

An example of rights management information includes certain records about what a computer does and how it does it. In one simple example, records may give permission to read a particular news article if that the customer is willing to pay a nickel to purchase the article and that the nickel may be paid using a budget provided by a credit card company or with electronic cash. A customer might, for example, seek only news articles from providers that take electronic cash and/or process information with a certain information clearinghouse as described in U.S. Patent application Serial No. 08/699,712 to Shear et al., filed 12 August 1996, for "Trusted Infrastructure Support Systems, Methods And Techniques For Secure Electronic Commerce Electronic Transactions And Rights Management" (hereafter "Shear et al") – the entire disclosure (including the drawings) of which is expressly incorporated into this application as if expressly set forth herein.

The Present Inventions Can Maintain Privacy

Figure 9 shows one way in which the electronic matchmaker can get information about a person. In this example, the electronic matchmaker asks Jill to fill out a computer questionnaire about what she likes. The questionnaire can also ask Jill what information she wishes to be maintained as authentic, and what information (e.g., encrypted by the system) may be used for secure matching only within a protected processing environment and can not be released to another party, or only to certain specified parties. The questionnaire

answering process may be directly managed by a protected processing environment to ensure integrity and secrecy, as appropriate.

For example, the questionnaire may ask Jill whether she likes baseball and whether she is interested in volcanoes. The electronic matchmaker can also ask Jill if it is okay to look at records her computer maintains about what she has used her computer for in the past. These computer records (which the computer can maintain securely so that no one can get to them without Jill's permission) can keep a history of everything Jill has looked at using her computer over the past month and/or other time period – this process being managed, for example, through the use of a system such as described in the "Ginter et al."

Looking at Figure 10, Jill may have used her computer last week to look at information about baseball, volcanoes and Jeeps. With Jill's permission, the electronic matchmaker can employ a protected processing environment 154 (schematically shown here as a tamper-resistant "chip" within the computer – but it can be hardware-based, software-based, or a combination of hardware and software) to look at the computer's history records and use them to help match Jill up with other kinds of things she is or may be interested in. For example, the electronic matchmaker can let an electronic publisher or other provider or information gatherer (e.g., market survey conductor, etc.) know that Jill is interested in team sports, geology and sports utility vehicles with or without more revealing detail -- as managed

by Jill's choices and/or rights management rules and controls
executing in her computer's protected processing environment 154.
The provider can send information to Jill – either automatically or at
Jill's request – about other, related things that Jill may be interested
5 in.

Figure 11 shows an example of how rights management and
other information Jill's computer maintains about her past usage can
be useful in matching Jill up with things she may need or want. The
computer history records can, for example, show that Jill looked at
10 hockey information for three hours and football information for five
hours during the past week. They can indicate that Jill uses a
Discover credit card to pay for things, usually spends less than \$10 per
item, averages \$40 per month in such expenses, and almost never
buys new programs for her computer.

15 The electronic matchmaker can, with and subject to Jill's
permission, look at and analyze this information. As one example,
the electronic matchmaker can analyze relevant rules and controls
provided by third parties who have rights in such information --
where such rules are controlled, for example, by Jill's computer's
20 protected processing environment 154. It can also look at and
analyze Jill's response to computer questionnaires indicating that she
likes baseball and football. The electronic matchmaker can, based on
all of this information, automatically select and obtain videos and/or
other publications for Jill about team sports and that cost less than

\$10 and that accept payment using a Discover card, so that Jill can preview and select those in which she may have a particular interest and desire to acquire.

Figure 12 shows that the electronic matchmaker can take into
5 account computer history records for lots of different people. The electronic matchmaker can work with other rights management related computer systems such as "usage clearinghouses" (non-limiting examples of which are described in each of "Ginter et al" and "Shear et al") to efficiently collect rights management related
10 information. The ability to collect history records from many different people can be very useful. For example, this can allow the electronic matchmaker to distinguish between things that are very popular and things that are not so popular.

The present inventions provide great increases in efficiency and
15 convenience. It can save you a lot of time and effort. It can allow computers to do a lot of the work so you don't have to. It can allow you to compete with larger businesses -- and allow large business to function more efficiently -- by allowing the location of resources particularly appropriate for certain business activities. You can
20 delegate certain complex tasks to a computer, freeing you to be more productive and satisfied with electronic activities. These automated processes can be "smart" without being intrusive. For example, they can learn about your behavior, preferences, changing interests, and even your personality, and can then predict your future interests based

on your past behavior and interest expressions. These processes can ensure confidentiality and privacy – so that no one can find out detailed information about you without your consent. Across the full range of personal and business activities, the present inventions allow
5 a degree of basic efficiency, including automation and optimization of previously very time consuming activities, so that interests and possible resources are truly best matched.

The present inventions handle many kinds of important issues and addresses the widest range of information and rights and
10 automation possibilities. For example, the present inventions are capable of handling (but are not limited to):

- consumer information;
- computer information;
- business information;
- 15 • entertainment information;
- other content information;
- information about physical products;
- all other kinds of information.

It can reflect and employ all kinds of rights to optimize
20 matching processes, including:

- content rights;
- privacy rights;
- governmental and societal rights;
- provider rights;

- distributor rights;
- consumer rights;
- workflow rights;
- other value chain participant rights;
- 5 • work flow rights;
- business and personal rights and processes of all kinds.

It can employ all kinds of parameter information, including:

- budget,
- 10 • pricing
- redistribution
- location (of party, item, etc.)
- privacy
- identity authenticity and/or specificity
- 15 • any other parameter information.

Pricing (for example the price of a specific item) can be used in matching based upon price per unit and/ or total price for a volume purchase, price for renting, right to redistribute, cost for redistributing items, etc.

- 20 Privacy can be used for establishing matching contingent upon usage reporting requirements for viewing, printing, extracting, dedistributing, listening, payment, and/or requiring the reporting of

other information such as personal demographics such as credit worthiness, stored value information, age, sex, marital status, race, religion, and/or usage based generated profiling information based materially upon, for example, a users history of usage of electronic
5 content and/or commercial transactions, etc.

Identity can be used for matching based upon, for example, such as the presence of one or more specific, class, and/or classes of certificates, including, for example, specific participant and/or group of participant, including value chain certificates as described in
10 "Shear et al".

With the inventions described herein, commercial requirement attributes embodied in rules (controls and control parameter data) are employed in classification structures that are referenced by search mechanisms, either, for example, directly through reading rule
15 information maintained in readable (not encrypted) but authentic (protected for integrity) form, through reading rule information maintained securely, through processes employing a protected processing environment 154 of a VDE node, and/or through the creation of one or more indexes and/or like purpose structures, that,
20 directly, and/or through processes employing a protected processing environment 154, automatically compile commercial and other relevant (e.g., societal regulatory information such as a given jurisdiction's copyright, content access and/or taxation regulations) for classification/matching purposes.

The present inventions can employ computer and communication capabilities to identify information, including:

- topical classification such as described by conventional library classification systems,
- 5 • commercial characterizations -- including commercial parameter data such as pricing, size, quality, specific redistribution rights, etc.,
- creator (e.g., a publisher or manufacturer), distributor, societal, user, and other participant interests information,
- 10 • information generated by automated profiling of any and all of such parties or collections of parties,
- matching (including electronically negotiating a match) between the interests of any of such parties,
- where appropriate, the use of statistical procedures, expert systems, and artificial intelligence tools for
15 profiling creation and/or analysis, matching, and/or negotiation.

The present inventions thus provide for optimal user, provider, and societal use of electronic cyberspace resources (for example,
20 digital information objects available across the Internet, sent by direct broadcast satellite, transmitted over a cable TV system, and/or distributed on optical disk).

Of particular importance is the notion of classes of content, classes of users, and classes of providers. For example, the present inventions can make use of any/all of the following:

- 5 • topical identification, for example, such as
 information represented in typical library subject
 and/or author and/or catalog and/or keyword search
 and retrieval information systems;
- 10 • any commercial requirements, associated with the use
 of electronic information (and/or to products,
 including non-electronic products, and/or to any
15 service), including information embodied in
 encrypted rules (controls and/or parameter data)
 governing rights in electronic value chain and
 electronic interaction contexts, and further including
20 information guaranteed for integrity;
- any information descriptive of an available resource
 (which may include any information, product, and/or
 service, whether available in electronic and/or
25 physical forms) such as: the quality of a digital
 product as evaluated and ranked and/or otherwise
 specified by one or more third parties and/or
 independent third parties (e.g., Consumer Reports, a
 trusted friend, and/or a professional advisor), the size
 of a product, length in time in business of a service or
 in the market of a product, a product's or service's

market share, and/or subject governmentally and/or other societally imposed rules and/or integrity guaranteed descriptions, including any associated regulatory requirements, such as societal

5 requirements granting and/or reporting access to information, for example, information on how to create a nuclear bomb to a confidential government auditing agency (this allowing free access to information while protecting societal rights);

10 • any information descriptive of a user and/or department and/or organization and/or class of users and/or departments and/or organizations (including, for example, such descriptive information encrypted and/or guaranteed for integrity) wherein such

15 information may include, for example, name, physical and/or network and/or cyber-wide logical network location, organizational and/or departmental memberships, demographic information, credit and/or trustworthiness information, and profile preference

20 and usage history information, including any generated profile information reflecting underlying preferences, and/or classes based on said descriptive information and/or profiles.

Some Of The Advantageous Features And Characteristics Provided By The Present Inventions

The classification, matching, narrowcasting, analysis, profiling, negotiation, and selection capabilities of the present inventions

5 include the following capabilities (listed items are not mutually exclusive of each other but exemplary samples):

- 10 • Enables highly efficient provision of classes of information, entertainment, and/or services to classes of individuals and/or entities that have (and/or may obtain) the right(s) to such information and are likely to find identified information interesting, useful, and/or entertaining.
- 15 • The present inventions also provide systems and methods for efficiently determining class hierarchies, classification schemes, categories, and/or category schemes and/or the assignment of objects, persons and/or things to said class hierarchies, classification schemes, categories, and/or category schemes using at least some rights management information.
- 20 • Helps systems, groups, and/or individuals classify, locate, and/or obtain specific information and/or classes of information made available through so-called "publish and subscribe" systems and methods using, among other things, subject-based addressing and/or messaging-based protocol layers.
- 25

- Provides fundamentally important commercial and societal rules based filtering to identify desired electronic information and/or electronic information containers through the use of classification structures, profiling technology, and matching mechanisms that harness the vast information opportunities in cyberspace by matching the information needs of users against commercial and/or societal rules related to the use of available information resources, including, for example, commercial and/or societal consequences of digital information use imposed as provider requirements and specified through the use of, and enforced by the use of, a trusted rights management system such as described in “Ginter et al”.
- Enables content creators and/or distributors to efficiently "stock the shelves" of retail electronic content outlets and similar merchandisers (both electronic and hard goods) with products and/or services most likely to be purchased and/or used by the customers of such merchandisers. This includes both identifying and "stocking" the most desirable products and/or other user desired resources and optimally presenting such products and/or other

resources in a manner optimized for specific users and/or user classes.

- 5 • Matching may be based on history of matching, that is, matching derived at least in part from previous matching, one non-exhaustive example of which includes learned matching for increasing efficiency.
- 10 • Enables matching for value chains where the matching is against a plurality of co-participating value chain parties requirements and/or profiles against match opportunities, and/or matching by matches comprised of match input and/or aggregation of match rule sets of providers used to "dock" with one or more user needs, interests, requirements match sets.
- 15 • Helps match persons and/or things using fuzzy matching, artificial intelligence (e.g., expert systems), and other methods that that match using plural match sets from providers and/or receivers.
- 20 • Makes search easier by using smart agents that match at least in part using at least one class.
- 25 • Helps bring buyers and sellers together through cross matching, where both parties offer to provide and/or receive content and/or physical goods for consideration, including barter matching and negotiated barter and other kinds of matching.

- Helps potential customers find those members (e.g., objects such as digital information containers) of any one or more classes of content most useful, entertaining, and/or interesting to them.
- 5 • Facilitates organizations securely and efficiently acquiring and distributing for internal use certain classes of content available from external providers and/or more securely and/or efficiently managing classes of their own content, including being able to
- 10 authorize certain classes of employees to use specified classes of internal and/or external content.
- Efficiently supporting matching between users and digital information where participants in a chain of handling and control have specified rules and usage
- 15 consequences for such digital information that may depend on class membership, for example, on class(es) of content and/or class(es) of value chain participants and/or classes of electronic events, wherein such participants include, for example, users
- 20 and/or participants contributing rules and consequences.
- Enables first individuals and/or organizations to locate efficiently other individuals, organizations, products, and/or services who have certain characteristics that
- 25 corresponds to such first individuals' and/or

organizations' interests, including interests generated by profiling information locally gathered through local event auditing at a VDE installation.

- 5 • Facilitates businesses informing a customer about things of special interest to her or him, such as classes of goods, services, and/or content, including directing such information to a customer at least in part based on profiling information locally gathered at a VDE installation through local event auditing at a VDE
10 installation.
- Allows trading companies to match suppliers of certain classes of goods and/or services with those who desire to purchase and/or use those classes of goods and/or services, wherein such matches may
15 include fulling a commercial business interaction and may further include one or more sequences of matches and/or nested matches (a sequence and/or grouping of matches within a given organization or group, wherein such matches may be required to
20 occur in a certain order and/or participate along with other matches in a group of matches before a given match is fulfilled).
- Enhances equity portfolio management by making
25 easier for traders to identify those equities having certain desired characteristics, such as belonging to

the class of equities that will have the greatest positive effect on the value of the trader's portfolio given certain classes of information and assumptions. Such matches may take into account information external to the fulfilment of a given trade, for example, one or more certain other market or specific variable thresholds must be met before an equity is traded, such as a certain rise in the an index stock value of, and/or revenue of, certain one or more network hardware suppliers before a certain quantity of equity is purchased at a certain price for stock of a certain network hardware supplier raw network component manufacturer, and wherein, for example, such determinations can be performed highly efficiently at a user VDE installation as the point of control, where such node receives such trusted information in, for example, VDE containers, as is necessary for a control decision to occur to purchase such equity of such network hardware supplier raw component manufacturer.

- Makes easier automated foreign currency exchange by enabling currency traders to identify members of the class of possible trades and/or conversions that are likely to produce the best returns and/or minimize losses.

- Helps consumers and organizations manage their affairs more efficiently and effectively and helps providers of services by automatically matching users with services that meet certain specified criteria, such as, for example, U. S. and Swiss banks offering the highest interest rates on certain time based classes of bank deposit instruments.
5
- Enables distributors of software and other content to identify one or more classes of users who are most likely to be interested in purchasing or otherwise using certain classes of software.
10
- Enables rightsholders to employ rules and/or usage consequences dependent on membership in one or more classes where class membership may be indicated by possession of a special digital document called a "certificate."
15
- Enables rightsholders to employ rules and/or usage consequences at least partially dependent on roles and responsibilities within an organization, where those roles and responsibilities may be indicated by possession of a digital certificate, digital membership card, and/or other digital credential.
20
- Facilitates more efficient automation of manufacturing and other workflow processes by, for example, matching certain manufacturing steps and/or
25

processes with performance parameter data associated with available classes of equipment capable of performing those steps and/or processes.

- 5 • Makes easier the administration and enforcement of government and/or societal rights by, for example, providing matching means for automatically applying certain classes of tax rules to appropriate classes of sales and other transactions.
- 10 • Enables altering the presentation of information and/or other content depending on the matching between preferences of the user and one or more classes of content being presented.
- 15 • Enables processing or altering (narrowcasting) of an event (e.g., the presentation of information and/or other content), for example, dynamically adjusting the content of an event, in response to a matching among the preferences and/or reactions of a user and/or user group, one or more classes of content being processed through one or more events, one or more classes of 20 one or more users participating in and/or otherwise employing the one or more events, and/or event controls (i.e., rules and/or parameter data).
- 25 • Allows the rules and usage consequences and the presentation of information to vary according to the difficulty of the information, including, for example,

adjusting the difficulty of an electronic game so that it is neither too frustratingly difficult nor too easy to use.

- 5 • Enables a user to efficiently locate content in one or more particular classes, where class is defined at least in part by weighted topical classification, where, for example, a document or other object is classified in one or more categories where at least one category reflects the absolute or relative attention given to that class in the object being classified.
- 10 • Facilitates users' creation of a new document from parts of two or more documents, where at least one of such parts is identified and/or retrieved based upon matching the part's membership in one or more classes identified by trusted, commercial controls employed through the use of a rights management system.
- 15 • Enables users to search for, locate, and use only those parts of a document that belong to one or more specified classes, including those parts having certain commercial controls, for example, reflecting acceptable usage restrictions and/or pricing.
- 20 • Enhances search and retrieval by creating new classes of content descriptors that incorporate various

disparate standards for content description and/or location.

- Allows consumers to easily locate services having certain specified characteristics, for example, gambling services offering the most favorable odds and/or specified rules for a particular game or games.
- Helps consumers obtain certain classes of tickets to certain classes of events.

The above capabilities, and others described in this application, are often ideally managed by distributed commerce nodes of a distributed, rights management environment embedded in or otherwise connected to the operating system clients of a distributed computing environment such as described in “Ginter et al” and further described in “Shear et al”, and employing, for example, rules, integrity management, container, negotiation, clearinghouse services, and trusted processing capabilities described in “Ginter et al” and “Shear et al”.

The Present Inventions Make Use Of Many Kinds Of Information And/Or Data

As discussed above, these inventions provide, among other things, matching, classification, narrowcasting, and/or selection based on rights management and other information. In particular preferred examples, these matching, classification, narrowcasting, and/or

selection processes and/or techniques may be based at least in part on rights management information. The rights management information may be an input to the process, it may be an output from the process, and/or the process can be controlled at least in part by rights management information. Information in addition to, or other than, rights management information may also be an input, an output, and/or a basis for controlling, the process and/or techniques.

Rights management information may be directly or indirectly inputted to the matching, classification and/or selection process. For example, rights management controls, rules and/or their consequences may be an input. Examples of such controls and/or rules include object registration related control set data, user related control set data and/or computer related control set data. In addition or alternatively, information provided based on control sets or rules and their consequences may be inputted. The following are examples of such information that may be provided based, for example, on rules and consequences:

- information exhaust;
- user questionnaires,
- audit trail related information;
- aggregated usage data;
- information measuring or otherwise related to user behavior;
- information measuring or otherwise related to user preferences;

- information measuring or otherwise related to user personality;
- information measuring or otherwise related to group behavior;
- 5 • information measuring or otherwise related to group preferences;
- information measuring or otherwise related to group culture
- 10 • information measuring or otherwise related to organizational behavior;
- information measuring or otherwise related to organizational preferences;
- information measuring or otherwise related to organizational culture;
- 15 • information measuring or otherwise related to institutional behavior;
- information measuring or otherwise related to institutional preferences;
- 20 • information measuring or otherwise related to institutional culture;
- information measuring or otherwise related to governmental behavior;
- information measuring or otherwise related to governmental preferences;

- information measuring or otherwise related to governmental culture;
- information measuring or otherwise related to societal behavior;
- 5 • information measuring or otherwise related to societal preferences;
- information measuring or otherwise related to societal culture;
- object history related information;
- 10 • other types of information;
- any combinations of information including, some, all or none of the information set forth above.

The processes, techniques and/or systems provided in accordance with these inventions may output rights management
15 related information such as, for example:

- one or more control sets;
- various rules and/or consequences;
- information used by control sets;
- certificates;
- 20 • other rights management information.

In accordance with various preferred embodiments provided by these inventions, information other than rights management information may also be used, at least in part, as an input, output and/or to control the matching, classification, narrowcasting, and/or

selection processes, systems and/or techniques. Examples of such information include:

- content object information;
 - full text
 - 5 • portions of objects
 - portions of sub-objects
 - abstracts
 - metadata
 - other content object related information
- 10 • user information
 - census information
 - purchasing habits
 - credit and financial transaction related
 - information
 - 15 • governmental records
 - responses to questionnaires
 - survey results
 - other user information
- 20 • computer related information
 - identification information
 - configuration information
 - other computer related information
- combinations of information.

Matching/Classifying/Selection

Systems, methods and techniques provided in accordance with these inventions can classify a variety of types of things including, for example:

- 5 • people
- computers
- content
- events
- transactions
- 10 • objects of all types
- combinations of things;
- combinations of people and things.

The matching, classifying and/or selecting processes provided in accordance with these inventions are very flexible and useful. For
15 example, they may be used to associate people with information, information with other information, people with other people, appliances with people, appliances with information, and appliances with other appliances. The present inventions in their preferred examples can associate any kind of information, object or thing with
20 any other kind of information, object or thing.

Different Associations Between Classes and Rights

The processes, systems and/or techniques provided in accordance with these inventions can provide and/or take into account many different kinds of associations between classes and rights. For

example, they can look at what rights are available to a user, computer, data structure or any other object. They can also look to rights selected by an object (for example, the subset of rights a user has chosen or otherwise identified). Alternatively or in addition, they
5 can look to rights that have been exercised by a user or in conjunction with an object or other thing, and they can look to the consequences of exercising such a right(s).

**Embodiments in Accordance With the Present
Inventions Can Be Used to Define Classes Based on Uni-
10 Dimensional and/or Multi-Dimensional Attributes and/or
Characteristics**

Example processes, systems and/or techniques provided in accordance with these inventions can be used to define classes based on uni-dimensional and/or multi-dimensional attributes and/or
15 characteristics. Any one or more attributes can be used. The attributes and/or characteristics can be flexibly defined. They may define groups or classes containing elements sharing certain attributes in common. There can, for example, be a spectrum of classification that takes into account gray areas as to whether a particular person or
20 thing possesses a certain one or a number of particular attributes and/or characteristics. Or classification may have a higher degree of certainty or definition. For example, a process can test to determine whether particular people or things are inside or outside of particular classes or groups based on one or a number of attributes or
25 characteristics (for example, whether you live in Denver, are under the age of 25 and are single). In accordance with additional specific

features provided by these inventions, there may be a minimum number of different classes set up to "cover" a particular situation – with every person or thing either being within or outside of a given, disjoint class or group.

5 Preferred Examples In Accordance With The Present Inventions Are Extensible to Accommodate Changing Conditions

The systems, methods and/or techniques provided by these inventions are extensible to accommodate changing conditions. For example, they can be made to readily adapt to changes in rules, consequences, topics, areas and/or subjects pertaining to groups such as, for example categories, and any other variable. Furthermore, partially and/or entirely new variables may be introduced to one or more existing sets of variables -- for example, to extend or otherwise modify a model to account for additional variables, to apply a new strategy, to adapt to new network and/or installation circumstances, to adapt to new user factors, to change analysis and/or other processing characteristics, and so on.

20 Preferred Examples In Accordance With The Present Inventions Are Compatible With Pre-Existing or Any New Classification Techniques or Arrangements

The example systems, methods and/or techniques provided by these inventions can be made fully compatible with any classification and/or categorization means, method, process, system, technique, algorithm, program, and/or procedure, presently known or unknown,

for determining class and/or category structures, definitions, and/or hierarchies, and/or the assignment of at least one object, person, thing, and/or member to at least one class and/or category, that without limitation may be:

- 5 • implemented by computer and/or other means; and/or
- based upon discrete and/or continuous mathematics; and/or
- using nominal, ordinal, interval, ratio and/or any other measurement scale and/or measurement mode; and/or
- 10 • including parameter data; and/or
- entail linear and/or non-linear estimation methods; and/or
- any other methods.

For example, classification can be performed using any or all of
15 the following example classification techniques:

- Statistical techniques that identify one or more clusters of cases sharing similar profiles and/or features, including any of the family of cluster analysis methods, for example, those described in
20 Hartigan (Hartigan, J. A., Clustering Algorithms, New York: Wiley, 1975);
- Methods for numerical taxonomy, for example, as described, for example, by Sneath and Sokal (Sneath, Peter H.A. and Robert R. Sokal, Numerical

Taxonomy: The Principals and Practice of Numerical Classification, San Francisco: W.H. Freeman, 1973);

- 5 • Any of the methods for cluster analysis, factor analysis, components analysis, and other similar data reduction/classification methods, for example, those implemented in popular statistical and data analysis systems known to those skilled in the arts, for example, SAS and/or SPSS;
- 10 • Pattern classification techniques, including components analysis and neural approaches, for example, those described by, for example, Schurmann (Schurmann, Jurgen, Pattern Classification: A Unified View of Statistical and Neural Approaches, New York: John Wiley & Sons, 1966);
- 15 • Statistical techniques that identify one or more underlying dimensions of qualities, traits, features, characteristics, etc., and assign parameter data indicating the extent to which a given case has, possesses, and/or may be characterized by the
- 20 underlying dimension, factor, class, etc. and/or result in the definition of at least one class and/or the assignment of at least one case to at least one class, for example, as described by Harman (Harman, Harry H., Modern Factor Analysis, 3rd ed. rev., Chicago: University of Chicago Press), and/or as implemented
- 25

by SAS and/or SPSS and/or other statistical analysis programs.

- 5 • Statistical methods that employ fuzzy logic and/or fuzzy measurement and/or whose assignment to at least one class entails probabilities different from 1 or zero.
- 10 • Bayesian statistical classification techniques that use estimates of prior probabilities in determining class definitions and/or the assignment of at least one case to at least one class;
- 15 • Any statistical and/or graphical classification and/or data reduction method that uses rotation of reference axes, regardless of whether orthogonal or oblique rotations are used, for example, as described in Harman, and as implemented in SAS and/or SPSS and/or other statistical programs;
- 20 • Statistical methods for two and three way multidimensional scaling, for example, the methods described by Kruskal and Wish (Krusgal Joseph B. and Myron Wish, *Multidimensional Scaling*, Beverly Hills, CA: Sage Publications, 1978), and/or by Shepard, et al. (Shepard, Roger N., A. Kimball Romney, and Sara Beth Nerlove, *Multidimensional Scaling: Theory and Applications in the Behavioral Sciences*, New York: Seminar Press, 1972);
- 25

- Knowledge based approaches to classification, for example, as described by, for example, Stefik (Stefik, Mark, "Introduction to Knowledge Systems," San Francisco: Morgan Kauffman, 1995); and
- 5 • any other classification techniques or arrangements pre-existing or yet to be developed.

10 **Preferred Examples In Accordance With The Present Inventions Are Fully Compatible With A Wide Array of Technologies Including the Distributed Commerce Utility System and the Virtual Distribution Environment**

Systems, methods and/or techniques provided in accordance with these inventions build upon and can work with the arrangements disclosed in "Ginter et al"; "Shear et al"; and other technology related
15 to transaction and/or rights management, security, privacy and/or electronic commerce.

For example, the present inventions can make particular use of the security, efficiency, privacy, and other features and advantages provided by the Virtual Distribution Environment described in
20 "Ginter et al".

As another example, a matching and classification arrangement can be constructed as a distributed commerce utility system as described in "Shear et al". The present inventions can work with other distributed commerce utility systems, and can enhance or be a
25 part of other commerce utility systems.

By way of non-exhaustive, more specific examples, the present inventions can be used in combination with (and/or make use of) any or all of the following broad array of electronic commerce technologies that enable secure, distributed, peer-to-peer electronic rights, event, and/or transaction management capabilities:

- a "VDE" ("virtual distribution environment") providing, for example, a family of technologies by which applications can be created, modified, and/or reused;
- a standardized control and container environment which facilitates interoperability of electronic appliances and efficient creation of electronic commerce applications and models;
- a programmable, secure electronic transaction management foundation having reusable and extensible executable components;
- seamless integration into host operating environments of electronic appliances or direct employment of such technologies in electronic commerce applications;
- cyberspace digital content rights and transaction management control systems that may operate in whole or in part over Internets, Intranets, optical media and/or over other digital communications media;
- support of an electronic "world" within which most forms of electronic transaction such as content usage,

distribution, auditing, reporting, and payment activities can be managed;

- 5 • Transaction Operating Systems (operating systems that have integrated secure, distributed, and programmable transaction and/or event management capabilities);
- Rights Operating Systems (operating systems that have integrated, distributed, and programmable rights management capabilities);
- secure content container management;
- 10 • clearinghouse functions related to content usage;
- overall electronic commerce architectures that provide electronic commerce automation through the use of secure, distributed digital events management;
- the general enablement of traditional commerce behavior in the digital commerce world;
- 15 • enhanced inherent, distributed efficiencies of conventional commerce practices with powerful, reliable electronic security, and with the programmability and electronic automation efficiencies made possible by modern computing;
- 20 • trusted operation of a freely configurable, highly efficient, general purpose digital marketplace in which

- parties "come together" to establish commercial relationships;
- support of "real" commerce in an electronic form (that is, the progressive creation of commercial relationships that form, over time, a network of interrelated agreements representing a value chain business model);
 - enabling content control information to develop through the interaction of (and/or negotiation between) securely created and independently submitted sets of content and/or appliance control information;
 - interconnection of appliances providing a foundation for much greater electronic interaction and the evolution of electronic commerce;
 - a variety of capabilities for implementing an electronic commerce environment;
 - a neutral, general purpose platform for commerce;
 - an architecture that avoids reflecting specific distribution biases, administrative and control perspectives, and content types;
 - a broad-spectrum, fundamentally configurable and portable, electronic transaction control, distributing, usage, auditing, reporting, and payment operating environment;

- systems and methods that uniquely enable electronic commerce participants to protect their interests during the sequence of activities comprising an electronic commerce model;
- 5 • ability of commerce participants to assure protection by specifying rules and controls that monitor and enforce their interests during the processing of remote commerce events;
- 10 • permitting commerce participants to efficiently participate in, and manage, the distributed electronic activities of a digital value chain;
- allowing commerce model participants to, for example, securely and cooperatively govern and automate the distributed electronic activities comprising their collective electronic business models;
- 15 • allowing commerce model participants to securely contribute electronic rules and controls that represent their "electronic" interests;
- 20 • rules and controls that extend a "Virtual PresenceTM" through which the commerce participants govern remote value chain activities according to their respective, mutually agreed to rights;

- a Virtual Presence taking the form of participant specified electronic conditions (rules and controls) that must be satisfied before an electronic event may occur;
- 5 • rules and controls that enforce the party's rights during "downstream" electronic commerce activities;
- control information delivered by, and/or otherwise available for use with, the VDE content containers constituting one or more "proposed" electronic agreements which manage the use and/or consequences
10 of the use of such content and which can enact the terms and conditions of agreements involving multiple parties and their various rights and obligations;
- rules and controls from multiple parties forming aggregate control sets ("Cooperative Virtual
15 Presence™") that ensure that electronic commerce activities will be consistent with the agreements amongst value chain participants;
- control sets defining the conditions which govern
20 interaction with protected digital content (disseminated digital content, appliance control information, etc.);
- conditions used to control not only digital information use itself, but also the consequences of such use to protect the individual interests of commerce participants

- and form cooperative, efficient, and flexible electronic commerce business models;
- true, efficient electronic cooperative governance of value chain activities;
 - 5 • empowering each commerce model participant to securely deliver, and persistently maintain control over, the rules and controls they contributed specifying constraints on, and consequences of, electronic conduct;
 - extending Cooperative Virtual Presence over time and
10 involving the execution of controls, and the use of content, at physically dispersed locations, such as Internet user sites;
 - a chain of handling and control in which dispersed locations are bound together through the use of secure
15 communication techniques and unique, secure digital container technology;
 - ability to preserve the rights of parties through a series of transactions which may occur at different times and different locations;
 - 20 • extending the ability of electronic content providers to control the use of proprietary information;
 - allowing content providers to limit use to authorized activities and amounts;

- 5 • allowing participants (e.g., actors, directors, script and other writers, musicians, studios, publishers, distributors, retailers, advertisers, credit card services, content end-users, and others) involved in a business model to have the ability to embody their range of agreements and requirements, including use limitations, into an "extended" agreement comprising an overall electronic business model;
- 10 • representing such an extended agreement by electronic content control information which can automatically enforce agreed upon rights and obligations;
- a competitive, general purpose electronic commerce architecture supporting the distributed, secure "unmanned" electronic interaction;
- 15 • distributing such capabilities across networks and involving the sequence (or web) of distributed activities underlying electronic value chains;
- cooperative electronic governance of distributed electronic commerce processes that optimizes electronic commerce value propositions;
- 20 • the capability of electronically, remotely representing the interests of commerce participants to support efficient, flexible, commerce model automation;

- 5 • enabling rules and controls that are independently contributed by multiple parties to securely merge together and form the collective rules and controls sets that reflect the electronic commerce agreements between parties;
- using rules and controls sets to collectively, automatically, govern remote electronic conduct;
- securely managing the integration of control information provided by two or more parties;
- 10 • constructing electronic agreements between VDE participants that represents a "negotiation" between the control requirements of two or more parties and enacts the terms and conditions of a resulting agreement;
- ensuring and/or enforcing the rights of each party to an
15 electronic agreement regarding a wide range of electronic activities related to electronic information and/or appliance usage;
- the ability to broadly support electronic commerce by securely managing independently delivered VDE
20 component objects containing control information (normally in the form of method, data, or load module VDE objects);
- using independently delivered control information to negotiate with senior and other pre-existing content

control information to securely form derived control information;

- 5 • ensuring that all requirements specified by derived control information are satisfied before VDE controlled content is accessed or otherwise used;
- ensuring that all load modules and any mediating data which are listed by the derived control information as required are available and perform their required function;
- 10 • use of independently delivered control components to allow electronic commerce participants to freely stipulate their business requirements and trade offs;
- allowing electronic commerce, through the various control requirements stipulated by VDE participants, to
15 evolve into forms of business which are the most efficient, competitive and useful -- much as with traditional, non-electronic commerce;
- providing commerce participants with the ability to
20 freely fashion the chains of handling and control pathways that protect data and processes and the freedom to shape the models within which their Virtual Presence operates -- allowing commerce participants to optimally formulate their electronic commerce value propositions;

- VDEs configured to support the various underlying agreements between parties that define important electronic commerce pathways of handling for electronic content, content and/or appliance control information, content and/or appliance usage information and payment and/or credit;
5
- allowing content creators and other providers to specify the pathways that, partially or fully, must be used to disseminate commercially distributed property content, content control information, payment administrative content, and/or associated usage reporting information;
10
- empowering commerce participants, subject to the rules and controls previously set in a value chain, to freely fashion control models implementing their Virtual Presence by using GUI templates or rights programming languages employing commerce/rights management components;
15
- component based control methods that allow the present inventions to efficiently operate as a highly configurable content control system;
20
- content control models that can be iteratively and asynchronously shaped, modified, and otherwise updated to accommodate the needs of VDE participants;

- iterative and/or concurrent multiple participant processes through the submission and use of secure, control information components (e.g., executable code such as load modules and/or methods, and/or associated data);
- 5 • control information for Virtual Presence employed in protected processing environment nodes located at user sites to ensure that digital events are governed in accordance with the collective rights of commerce model participants;
- 10 • digital events that launch or require other digital events;
- digital events that may include, for example, content use consequences such as collection of audit information, secure communication of such information, payment for content use, or satisfaction of any other electronically stated condition;
- 15 • events that occur within either the secure setting of a local node, or more widely within the secure environment of a distributed system of nodes;
- the association of Virtual Presence rules and controls
- 20 with protected information enclosed within one or more electronic content containers to achieve a high order of configurability for Virtual Presence chains of handling and control;

- distribution using VDE that may package both the electronic content and control information into the same VDE container, and/or may involve the delivery to an end-user site of different pieces of the same VDE managed property from plural separate remote locations and/or in plural separate VDE content containers and/or employing plural different delivery means;
- content control information that is partially or fully delivered separately from its associated content to a user VDE installation in one or more VDE administrative objects;
- delivery of portions of said control information from one or more sources;
- making control information available for use by access from a user's VDE installation secure sub-system to one or more remote VDE secure sub-systems and/or VDE compatible, certified secure remote locations;
- use of delivery means that may include electronic data storage means such as optical disks for delivering one portion of said information and broadcasting and/or telecommunicating means for other portions of said information;

- allowing a content provider to deliver different business rules to a large corporate customer, compared with rules delivered to "retail" customers;
- 5 • supporting separation of content and Virtual Presence controls to allow a provider to associate different control sets with the same content – and not requiring the provider to create one set of content controls that apply to all types of customers;
- 10 • allowing content provider modification over time of rules and controls to reflect sales, new pricing, special discounts, etc. – while limiting this right by rules and controls provided by other parties having more senior rights;
- 15 • employing secure object container technology to efficiently implement Virtual Presence chains of handling and control;
- use of software container technology to significantly facilitate the organized dissemination of digital content, including the specialized form of digital content
20 constituting rights control information;
- employing object software technology and using object technology to form containers for delivery of at least in part encrypted or otherwise secured information;

- using containers that contain electronic content products or other electronic information and some or all of their associated permissions (control) information;
- 5 • distributing container objects along pathways involving content providers and/or content users;
- securely moving containers between nodes of a VDE arrangement, which nodes operate VDE foundation software and execute control methods to enact electronic information usage control and/or administration models;
- 10 • employing delivered containers both for distributing VDE control instructions (information) and/or to encapsulate and electronically distribute content which has been at least partially secured;
- supporting the essential needs of electronic commerce value propositions by uniting fundamental
- 15 configurability with secure Virtual Presence;
- virtual presence across virtual networks in accordance with the underlying agreement amongst commerce model participants to allow each participant to enjoy secure,
- 20 reliable electronic automation of commerce models;
- allowing each rights holder's Virtual Presence at a remote site to possess the sole authority to administer or delegate the participant's electronic rights;

- capabilities that contribute to establishing an environment of trusted cooperative governance;
- practical enhancements relating to the establishment of secure event management and the maintenance of secure audit, encryption, budget, and other relevant information;
- control structures for an overall, distributed, secure rights/event administration environment;
- processes for interaction between independently delivered rules and controls, including electronic negotiation;
- creating distributed rights operating systems;
- integrating control processes into host operating environments;
- secure semiconductors to support protected processing environments;
- a secure, programmable, digital event management component architecture in which components are fully assembleable and reusable;
- differing assemblages of components formed to reflect an exhaustive array of commerce model functional capabilities, overall model implementations, and ad hoc event management scenarios;

- support for the full range of digital content types, delivery modes, and reporting and other administrative activities;
- traveling objects;
- 5 • smart agents;
- "atomic" load module operation to support "sparse space," cost-effective, secure processing semiconductors;
- smart card and other traveling client nodes;
- creating rights management software container technologies, including extraction, embedding, and other secure container content management processes;
- 10 • Chain of Handling and Control generation of secure objects (containers) and associated control information;
- audit reconciliation and usage pattern evaluation processes;
- 15 • specialized cryptographic implementations;
- use of a specialized electronic rights and commerce language, unique applications for fingerprinting and/or watermarking technologies, secure control structures, the formulation of new types of metering technologies, reciprocal event management (employing dispersed user sites) for automating web-like commerce models, and many other designs and capabilities;
- 20

- mechanisms to persistently maintain trusted content usage and reporting control information through both a sufficiently secure chain of handling of content and content control information and through various forms of usage of electronic information;
5
- rights management technology supporting persistent, distributed controls;
- means enabling continuing Virtual Presence through Chains of Handling and Control;
- 10 • persistency of control as a unique and fundamentally important attribute underlying Virtual Presence and Chain of Handling and Control for enabling true commerce behavior in cyberspace including ad hoc relationships and activities, distributed processes, and
15 reliable enforcement of agreements between parties;
- Persistent Virtual Presence controls that continue to be enforced -- to the extent required by the controls themselves -- as protected digital content is, for example, used and reused, copied and further distributed, extracted
20 and embedded, audited and reported;
- persistency responsive to rules and controls associated with electronic events, that causes new secure content containers to be created automatically by systems and methods supplying the procession of secure transport

- vehicles required by Chain of Handling and Control for conveying disseminated content, associated rules and controls, and audit information and payment;
- 5 • container creation to carry extracted content, payment tokens, control information, audit information, and the like;
 - securely generated containers carrying with them rules and controls stipulated by rules and controls associated with one or more triggered electronic events;
 - 10 • capabilities for persistency and independent secure delivery and merging of rules and controls that provide technical means for ensuring that dynamic user behavior can be encouraged, rather than discouraged;
 - 15 • dynamic user behavior encouraged as a critical link in building ad hoc relationships and cost-effectively distributing content, while simultaneously ensuring that rights holders are protected and retain control over their business models;
 - 20 • enabling ad hoc behavior that frees users from constraints on their conduct resulting from inflexible, first generation technologies;
 - support for enterprising behavior that is characteristic of traditional commerce resulting in more efficient and more satisfying electronic commerce experiences;

- general purpose character electronic commerce technologies provided by a combination of important capabilities including component, object oriented, programmable control language; secure specialized container technology; independent delivery of secure control information mechanisms; Chain of Handling and Control persistency of control mechanisms; event driven operating system functions; and the advanced security architecture – allowing multiple simultaneous models to evolve, and practically and efficiently operate;
- general purpose rights and event management architecture that is intrinsically reusable for many simultaneous models -- providing enormous competitive economic advantages over technologies that are essentially single model by design;
- commerce architecture client nodes that are basic pieces of reusable cyberspace infrastructure;
- generalized configurability resulting, in part, from decomposition of generalized requirements for supporting electronic commerce and data security into a broad range of constituent "atomic" and higher level components (such as load modules, data elements, and methods) that may be variously aggregated together to form control methods for commercial electronic agreements and data security arrangements;

- a secure operating environment employing VDE foundation elements along with securely deliverable VDE components that enable electronic commerce models and relationships to develop;
- 5 • the unfolding of distribution models in which content providers, over time, can expressly agree to, or allow, subsequent content providers and/or users to participate in shaping the controls for, and consequences of, use of electronic content and/or appliances;
- 10 • a very broad range of the functional attributes important for supporting simple to very complex electronic commerce and data security activities;
- electronic information and/or appliance usage control (including distribution), security, usage auditing, reporting, other administration, and payment
15 arrangements;
- capabilities that rationalize the support of electronic commerce and electronic transaction management stemming from the reusability of control structures and user interfaces for a wide variety of transaction
20 management related activities;
- content usage control, data security, information auditing, and electronic financial activities that can be

supported with tools that are reusable, convenient, consistent, and familiar;

- 5 • a general purpose Rights Operating System employing a reusable kernel and rights language components that provides the capabilities and integration needed for the advanced commerce operating systems of the future;
- 10 • a general purpose, reusable electronic commerce capabilities that all participants can rely on will become as important as any other capability of operating systems;
- 15 • such a rights operating system providing rights and auditing operating system functions and other operating system functions -- the rights and auditing operating system functions securely handling tasks that relate to virtual distribution environment;
- 20 • secure processing units and/or protected processing environments that provide and/or support many of the security functions of the rights and auditing operating system functions;
- an overall operating system designed from the beginning to include the rights and auditing operating system functions plus the other operating system functions -- or incorporation of the rights and auditing operating system

- functions as an add-on to a preexisting operating system providing the other operating system functions;
- operating system integration and the distributed operating systems; and
 - 5 • a rational approach - a transaction/distribution control standard - allowing all participants in VDE the same foundation set of hardware control and security, authoring, administration, and management tools, for widely varying types of information, business market
 - 10 model and/or personal objectives;

Any or all of these features may be used in combination with the inventions disclosed herein.

Brief Description of the Drawings

15 These and other features and advantages will be better and completely understood by referring to the following detailed description of presently preferred example embodiments in accordance with the drawings, of which:

20 Figures 1A-4 show "prior art" examples of how it is hard to find things you need or want;

Figures 5-12 are simplified examples of what example systems, methods and techniques in accordance with these inventions can do;

Figures 13, 14 and 14A show an example matching and classification utility system architecture;

Figures 15-15G show examples of how a matching and classification utility system can interact with other commerce utility systems;

Figures 16A-16C show examples of distributed matching and classification utility system organizations;

Figure 17 shows example matching and classification utility system functionality definitions;

Figures 18-46(B) show example steps that may be performed by the example matching and classification utility system; and

Figures 47-70 show some example matching and classification utility system applications.

Detailed Description Of Presently Preferred Example Embodiments

Figures 5-12 and the discussion above provide an introduction to the following detailed description of presently preferred embodiments in accordance with these inventions. The "electronic matchmaker" shown in Figures 5-12 is implemented in these more detailed embodiments by a matching and classification utility system 900.

Example Matching And Classification Utility

Figure 13 shows an example matching and classification utility system 900 as including:

- an object classifier 902;
- 5 • a user (people) classifier 904; and
- a matching engine 906.

Object classifier 902 classifies things. User classifier 904 classes people. Matching engine 906 matches things with other things, things with people, and/or people with other people.

10 In more detail, object classifier 902 receives information about objects and uses that information to classify those objects into groups based on the qualities or characteristics of the objects. For example, the object classifier 902 may classify objects of the type described in in "Ginter et al". Such objects may comprise information and/or
15 associated rules for using the information. For example, object classifier 902 may receive as inputs:

- rights management information 909 such as rules and/or associated consequences;
- things 908 controlled or affected by such rights
20 management information including, for example content objects or other information subject to such rules;
- items 910 such as metadata, abstracts or the like that describe the things 908; and/or

- other information of any type.

Object classifier 902 classifies and/or selects things based at least in part on these inputs.

In this example, user classifier 904 is a type of object classifier
5 that is specially adapted to classify people. User classifier 904 can
classify people based, for example, on:

- audit trails 912 indicating how people have used their computers and other electronic appliances;
- profiles 914 developed by asking users questions
10 about their preferences;
- controls 909' that are associated, at least in part, with the user or things the user uses;
- object descriptors 910' that describe objects used by the user; and/or
15 • other information about and/or relating to the user.

User classifier 904 classifies and/or selects people based at least in part on these inputs.

Matching engine 906 receives, as inputs, the classifications and/or selections made by the object classifier 902 and/or the user
20 classifier 904. Matching engine 906 matches things with things, things with people and/or people with people (or any combination of these) based on these selection and/or classification inputs.

Example More Detailed Architecture

Figure 14 shows a more detailed architectural diagram of matching and classification utility 900. In this example, matching and classification utility 900 receives a variety of inputs including, for

5 example, some or all of the following:

- objects 908 and/or information about objects including controls 909 and/or object descriptors 910;
- content 950;
- audit trail information 916;
- 10 • user information such as profiles 914;
- class information 952;
- user information 954;
- other rights management information 956;
- matching criteria 958;
- 15 • selection criteria 960; and/or
- other information.

Matching and classification utility 900 in this example can provide a variety of different outputs including, for example, some or all of the following:

- 20 • matching information 920;
- class hierarchies 962;
- category definitions 922 and class definitions 970;
- classified objects 908C;
- audit records 964 indicating the results of
- 25 classification, matching, and or selecting processes;

- reports 966 indicating the results of classification, matching, and/or selecting processes;
- targeted objects and/or pointers 968;
- controls 909;
- 5 • other rights management information; and
- other classification, matching and/or selection related information.

10 **A Preferred Embodiment Matching and Classification Utility 900 is a VDE-Aware Commerce Utility System**

 In the preferred embodiment, matching and classification utility 900 is constructed as a commerce utility system 90 as described in "Shear et al", and may comprise one or more processes securely distributed over one or more secure electronic appliances within a

15 "Virtual Distribution Environment" as described in "Ginter et al". Furthermore, the present inventions can be used in combination with and/or make use of a wide array of distributed electronic administrative and support services that may be referred to as the "Distributed Commerce Utility." Such a Distributed Commerce

20 Utility may be, among other things, an integrated, modular array of administrative and support services for electronic commerce and electronic rights and transaction management. The Distributed Commerce Utility provides, among other advantages, comprehensive, integrated administrative and support services for secure electronic

25 commerce and other forms of electronic interaction. These

administrative and support services can be used to supply a secure foundation for conducting financial management, rights management, certificate authority, rules clearing, usage clearing, secure directory services, and other transaction related capabilities functioning over a vast electronic network such as the Internet and/or over organization internal Intranets, or even in-home networks of electronic appliances. Such electronic interactions supported by the Distributed Commerce Utility may, for example, entail the broadest range of appliances and distribution media, non-limiting examples of which include networks and other communications channels, consumer appliances, computers, convergent devices such as WebTV, and optical media such as CD-ROM and DVD in all their current and future forms.

These administrative and support services can, for example, be adapted to the specific needs of electronic commerce value chains in a number of vertical markets, including a wide variety of entertainment applications. Electronic commerce participants can, for example, use these administrative and support services to support their interests, and/or they can shape and reuse these services in response to competitive business realities. Non-exhaustive examples of electronic commerce participants include individual creators, film and music studios, distributors, program aggregators, broadcasters, and cable and satellite operators.

The Distributed Commerce Utility can, for example, make optimally efficient use of commerce administration resources, and

can, in at least some embodiments, scale in a practical fashion to optimally accommodate the demands of electronic commerce growth. The Distributed Commerce Utility may, for example, comprise a number of Commerce Utility Systems. These Commerce Utility

5 Systems can provide a web of infrastructure support available to, and reusable by, the entire electronic community and/or many or all of its participants. Different support functions can, for example, be collected together in hierarchical and/or in networked relationships to suit various business models and/or other objectives. Modular support

10 functions can, for example, be combined in different arrays to form different Commerce Utility Systems for different design implementations and purposes. These Commerce Utility Systems can, for example, be distributed across a large number of electronic appliances with varying degrees of distribution.

15 Such a "Distributed Commerce Utility" provides numerous additional capabilities and benefits that can be used in conjunction with the particular embodiments shown in the drawings of this application, non-exhaustive examples of which include:

· Enables practical and efficient electronic commerce and rights

20 management.

· Provides services that securely administer and support electronic interactions and consequences.

- Provides infrastructure for electronic commerce and other forms of human electronic interaction and relationships.
- Optimally applies the efficiencies of modern distributed computing and networking.
- 5 · Provides electronic automation and distributed processing.
- Supports electronic commerce and communications infrastructure that is modular, programmable, distributed and optimally computerized.
- Provides a comprehensive array of capabilities that can be
10 combined to support services that perform various administrative and support roles.
- Maximizes benefits from electronic automation and distributed processing to produce optimal allocation and use of resources across a system or network.
- 15 · Is efficient, flexible, cost effective, configurable, reusable, modifiable, and generalizable.
- Can economically reflect users' business and privacy requirements.
- Can optimally distribute processes -- allowing commerce
20 models to be flexible, scaled to demand and to match user requirements.

- Can efficiently handle a full range of activities and service volumes.
 - Can be fashioned and operated for each business model, as a mixture of distributed and centralized processes.
- 5 · Provides a blend of local, centralized and networked capabilities that can be uniquely shaped and reshaped to meet changing conditions.
- Supports general purpose resources and is reusable for many different models; in place infrastructure can be reused by different
- 10 value chains having different requirements.
- Can support any number of commerce and communications models.
 - Efficiently applies local, centralized and networked resources to match each value chain's requirements.
- 15 · Sharing of common resources spreads out costs and maximizes efficiency.
- Supports mixed, distributed, peer-to-peer and centralized networked capabilities.
 - Can operate locally, remotely and/or centrally.
- 20 · Can operate synchronously, asynchronously, or support both modes of operation.

· Adapts easily and flexibly to the rapidly changing sea of commercial opportunities, relationships and constraints of "Cyberspace."

Any or all of these features may be used in combination with
5 the inventions disclosed herein.

In more detail, as shown in Figure 14A, matching and classification utility 900 may include one or more rights operating system layers 90-1; one or more commerce utility support service layers 90-4; one or more service application connect layers 90-3; and
10 one or more service functions 90-B. One or more protected processing environments 154 may be used to support secure functions 90-D. Matching and classification utility 900 may be controlled, at least in part, by rights management information such as for example:

- VDE-compatible controls 909;
- 15 • rules and/or their consequences; and/or
- other rights management information.

Matching and Classification Utility Can Interact With Other Commerce Utility Systems

Figure 15 shows that matching and classification utility 900
20 can interact and interrelate with other commerce utility systems described in "Shear et al" including for example:

- financial clearinghouses 200,
- usage clearinghouses 300,
- rights and permissions clearinghouses 400,

- certifying authorities 500,
- secure directory services 600,
- transaction authorities 700,
- VDE administrators 800, and/or
- 5 • other commerce utility systems 90.

Figures 15A-15G show example detailed interactions between matching and classification utility 900 and these various other commerce utility systems 90.

Figure 15A shows interactions between matching and
10 classification utility 900 and a financial clearinghouse 200. For example, matching and classification utility 900 may send the financial clearinghouse 200:

- requests for information,
- class information such as classes and/or class
15 assignments,
- bills and charges, and/or
- other information.

Financial clearinghouse 200 may send matching and
classification utility 900:

- 20 • money,
- audit records,
- payment data,
- user data, and/or
- other information.

Figure 15B shows example interactions between matching and classification utility 900 and usage clearinghouse 300. Matching and classification utility 900 may send the usage clearinghouse 300:

- requests for information,
- 5 • class information such as classes and/o class assignments,
- audit information, and/or
- other information.

Matching and classification utility 900 may receive from usage
10 clearinghouse 300:

- requests for class information,
- usage and/or rights management information,
- audit records, and/or
- other information.

15 Figure 15C shows example interaction between matching and classification utility 900 and rights and permissions clearinghouse 400. In this example, rights and permissions clearinghouse 400 sends matching and classification authority 900:

- controls sets and/or object information;
- 20 • requests for class information;
- clearinghouse usage data; and/or
- other information.

In this example, matching and classification utility 900 sends the rights and permissions clearinghouse 400:

- rights management information such as control sets,
 - requests for information,
 - class related information such as classes and/or class assignments, and/or
- 5 • other information.

Figure 15D shows example interaction between matching and classification utility 900 and certifying authority 500. In this example, certifying authority 500 sends matching and classification utility 900:

- 10 • revocation lists,
- certificates,
 - certifying authority usage information,
 - requests for classification information, and/or
 - other information.

15 In this example, the matching and classification utility 900 sends the certifying authority 500:

- revocation list checks,
 - requests for certificates,
 - requests for usage information,
- 20 • classification related information such as classes and/or class assignments, and/or
- other information.

Figure 15E shows an example interaction between the matching and classification utility 900 and a secure directory services 600. In

this example, the matching and classification utility 900 sends the secure directory services 600:

- directory lookup information,
- class related information such as classes and/or class assignments,
- requests for information, and/or
- other information.

In this example, the secure directory services 600 sends the matching and classification utility 900:

- directory services usage information,
- directory information,
- requests for classification information, and/or
- other information.

Figure 15F shows an example interaction between the matching and classification utility 900 and a transaction authority 700. In this example, the matching and classification utility 900 sends the transaction authority 700:

- class related information such as classes and/or class assignments,
- requests for transaction usage information,
- requests for control sets, and/or
- other information.

In this example, the transaction authority 700 sends the matching and classification utility 900:

- transaction usage information,
- transaction control sets,
- requests for classification information, and/or
- other information.

5 Figure 15G shows an example interaction between the matching and classification utility 900 and a VDE administrator 800. In this example, the matching and classification utility 900 sends the VDE administrator 800:

- requests for administration,
- 10 • class related information such as classes and/or class assignments,
- requests for node and/or web information, and/or
- other information.

 In this example, the VDE administrator 600 sends the matching
15 and classification utility 900:

- requests for classification information,
- administrative information,
- node and/or user data, and/or
- other information.

20 **Matching and Classification Utility System Can Be In a Hierarchy of Commerce Utility Systems**

 Figure 16A shows an example of an administrative and support service hierarchy including matching and classification utility system(s) 900. In this example, a number of centralized overall

matching and classification utility systems 900 and/or other
Commerce Utility Systems 90 delegate some or all of their work
responsibilities to other Commerce Utility Systems 90. In the
particular example shown, Commerce Utility Systems 154 may
5 provide services to one or more members of one or more classes, for
example, to members of the class "manufacturing companies in the
Pacific rim." Organizations, such as companies, non-profit groups or
the like may have their own Commerce Utility Systems 156. Certain
electronic commerce or other activities (the entertainment industry,
10 for example) might have their own vertically-specialized Commerce
Utility Systems 158. Certain geographical, territorial or jurisdictional
groups (e.g., Commerce Utility Systems services provided with a
particular nation or state within nation, one example of which might
be all purchasers of particular products within the state of Wisconsin)
15 may have their own territorial/jurisdictional specialized Commerce
Utility Systems 160. Commerce Utility Systems 154, 156, 158, 160
lower in the hierarchy may, in turn, further delegate authorities or
responsibilities to particular consumers, organizations or other
entities.

20 In one example arrangement, the Commerce Utility Systems 90
to which authority has been delegated may perform substantially all
of the actual support work, but may keep the delegating Commerce
Utility Systems 90 informed through reporting or other means. In
another arrangement, the delegating Commerce Utility Systems 90
25 have no involvement whatsoever with day to day activities of the

Commerce Utility Systems to whom they have delegated work. In still another example arrangement, the more specialized Commerce Utility Systems do some of the work and the more overarching Commerce Utility Systems do other parts of the work. The particular
5 division of work and authority used in a particular scenario may largely depend on factors such as efficiency, trustedness, resource availability, the kinds of transactions being managed, and a variety of other factors. Delegation of clearing authority may be partial (e.g., delegate usage aggregation but not financial or rights management
10 responsibilities), and may be consistent with peer-to-peer processing (e.g., by placing some functions within consumers' electronic appliances while keeping some other functions centralized).

**Matching and Classification Utilities Can Provide
Services to Classes of Nodes, Users, Content Services
and/or Transaction Services**
15

Figure 16B shows an example of how Matching and Classification Utilities 900 can provide services to classes of nodes, users, content services and/or transaction services. In this example, matching and classification utility systems 900(1), ... 900(N) provide
20 horizontally specialized matching and/or classification services for different purposes. For example, matching and classification utility 900(1) serves VDE administrative type functions by classifying VDE deployment related information and associated objects. Matching and classification utility 900(2) specializes in higher education
25 classification tasks. Matching and classification utility 900(3)

specializes in business information related tasks, and matching and classification authority 900(N) specializes in trading transactions. Any of these specialties can be combined together, so that a single utility system 900 can perform multiple functions or portions of
5 functions.

Multi-Function Commerce Utility Systems Can be Organized Hierarchically or Peer-to-Peer

Figure 16C shows a still different, more complex Matching and Classification Commerce Utility System 900 environment including
10 elements of both a hierarchical chain of command and a high degree of cooperation in the horizontal direction between different multi-function matching and classification utility systems 900. In this example, there are five different levels of responsibility with a master or overarching matching and classification utility system 900(1) on
15 level 1 having the most authority and with additional matching and classification utility systems on levels 2, 3, 4, and 5 having successively less power, authority, control, scope and/or responsibility. Figure 16C also shows that different matching and classification utility systems 900 on the same level may have different
20 functions, scopes and/or areas of responsibility. For example:

- a Matching and classification utility system 900(2)(1) may be a "type A" Matching and classification utility system,
- Matching and classification utility system 900(2)(2) might be a "type B" Matching and classification utility system, and

- Matching and classification utility system 900(2)(3) might be a "type C" Matching and classification utility system.

On the next level down, Matching and classification utility systems might be type A Matching and classification utility system (such as, 900(3)(1) and 900(3)(2)), they might be type B Matching and classification utility systems (such as, 900(3)(4)), they might be type C Matching and classification utility systems (such as, 900(3)(5), 900(3)(6)), or they might be hybrids -- such as, Matching and classification utility system 900(3)(3) which is a hybrid having type A and type B functions. Figure 16C also shows that additional clearinghouses on levels 4 and 5 might have sub-types as well as types.

A matching and classification utility 900 might break out along content classes (e.g., movies; scientific, technical and medical; and software). Subtype A might include first run movies, oldies, and art films; subtype B might handle journals and textbooks; and type C might be responsible for games, office, educational content. Peer-to-peer communications between clearinghouses could involve differing classes of consumers, differing jurisdictional classes, differing payment methods classes, and/or any other class distinction.

Matching and Classification Utility System Can Be Constructed From Object-Oriented Service Functions

Figure 14A shows Matching and Classification Utility 900 can be constructed from service functions. Figure 17 shows in more

detail how a matching and classification utility system 900 can be constructed based on service functions such as for example:

- automatic class generation,
- automatic matching,
- 5 automatic class assignment,
- class based searching,
- class based directory,
- audit by class,
- market research,
- 10 rights management language processing,
- other service functions.

Example Detailed Steps Carried Out By Matching and Classification Utility System 900

- 15 The next section of the specification describes some example steps performed by the matching and classification utility 900.

Example Steps to Categorize Objects and/or Users and/or Appliances

- Figure 18 shows example steps to categorize objects, and
20 Figure 19 shows example steps to categorize users 95 and/or

appliances 100. The overall categorization steps in these examples are -- at this level -- similar to one another. The processes begin by getting input data (Figure 18, block 1840, Figure 19, block 1840'). Next, a classification and/or categorization method is selected (Figures 18, block 1842; Figure 19, block 1842'). The process then assembles a data matrix and applies the selected classification method to the data matrix (Figure 18, blocks 1844, 1846; Figure 19, blocks 1844', 1846'). In addition or alternatively, other data reduction methods may be used (Figure 18, block 1848; Figure 19, block 1848'). Next, the process assigns objects and/or users and/or appliances to the categories developed by the classification method that has been applied (Figure 18, block 1849; Figure 19, block 1849'). Finally, the process stores the results in electronic and/or non-electronic storage in the "write output data" step (Figure 18, block 1850; Figure 19, block 1850').

The "get input data" step 1840, 1840' may involve obtaining attribute and/or parameter data from various sources including, for example:

- electronic appliance related attribute data;
- user demographic data;
- user psychographic data;
- available rights management rules and/or consequences (e.g., permissions records);

- exercised rights management rules and/or consequences (e.g., permissions records);
- rights management and/or other audit and/or usage records;
- any third party source of any information, including rights management, usage, audit, statistical, personal, organizational, political, economic, social, religious, business, government, medical, research, academic, literary, military, and/or information and/or data in any format known or unknown concerning any and all other topics that may contribute to the definition of at least one class and/or the assignment of at least one object to a class.

Detailed example steps for harvesting this data are detailed below in connection with Figures 24-46B. This resulting attribute data may be accumulated and aggregated together to form a composite record used as the input to the classification process.

Figure 20 shows an example composite record 1852. This composite classification record may contain attributes derived from any or all of a variety of rights management and/or other data "harvesting" processes. For example, composite record 1852 may include demographic and/or psychographic data obtained by querying the user 95. It may contain usage data obtained by monitoring audit information produced by various usage transactions. It may contain information reflecting user choices concerning rights management

information, the rights management information available to particular users and/or objects, and rights management processes actually performed with respect to particular users and/or particular objects. The information may be analyzed first to provide statistical and/or other summary information, or individual, more granular information may be provided. The composite record 1852 may also contain attributes of particular electronic appliance 100 installations. The particular example composite record 1852 shown in Figure 20 is one non-limiting example composite attribute record containing attributes obtained through a number of different "harvesting" processes. The composite record 1852 may be organized in a way to allow easy and efficient selection of desired attributes in the course of a database lookup, for example, and to allow easy and efficient selection and/or coding as input to any aspect of a classification and/or the assignment of one or more objects to at least one or more classes.

The Figure 21 example cluster analysis process is one example of steps that may be performed as part of the "apply classification method(s)" block 1846, 1846' of Figures 18, 19. (A classification method, or any other method described in these processes, may be utilized as part of a "knowbot", "agent", "traveling agent", and/or "smart agent", a non-limiting example of which is described in "Ginter et al", for example, Figure 73.) In this particular example, the process selects variables and cases (blocks 1860, 1862, Figure 21), and then assembles an appropriate data matrix (block 1864). A

conventional cluster analysis is then applied (block 1866, Figure 21). The clusters may be interpreted to determine what they mean (Figure 21, block 1868), or they may be compared with previous results and if sufficiently similar, they may be assumed to reflect the same classes as the earlier classification procedure thus minimizing the need for additional interpretation of the clustering results. Step 1868 may be performed automatically or manually, or a combination of automatic and manual processing may be used. Finally, individual cases may be assigned to individual clusters to complete the classification process (Figure 21, block 1870).

Figures 22, 23 show two examples of classification outputs produced by the Figure 21 process. In the Figure 22 example, information from several individuals has been used to create two example categories that reflect differing use profiles. More classes may have been defined than those example classes shown here. Users assigned to the same class have many more features, behavior, and/or other attributes in common than each of them does with members assigned to other classes.

In example Figure 22, members of class 1 tend to spend more per content item purchased, travel abroad more frequently, are more interested in national and international news, business and travel information, and generally do not participate in "pay per view" events and/or content consumption. Members of class 1 also tend to add new rights and/or modify existing rights management controls for

content, for instance, to add a markup and redistribute the content in one example, may be less likely to express a religious preference and/or affiliation, and tend to use the Internet as an area for "surfing" and exploration.

5 Members of class 2 tend to pay less for content purchased, seldom travel abroad, tend to be interested in sports, religious content and events, and are more often consumers of movies than are members of class 1. Members of class 2 are more likely to "pay per view" than are members of class 1, and are much less likely to add
10 new controls to content and/or modify rights acquired. Members of class 2 are more likely to express a religious preference and among those that do, Protestant denominations are more frequently mentioned. Members of class 2 may use the Internet, but tend to do so as part of their work role and responsibilities rather than as
15 entertainment, hobbies, and other leisure-time pursuits.

 Some methods of classification produce parameter data rather than assignment of objects to more discrete (or fuzzy or other kinds of) classes. Instead, this parameter data may indicate the extent to which an object possesses one or more traits, attributes, or class
20 characteristics. For instance, a person may have been assigned to class 1 (call it "the cosmopolitan class") or class 2 (call it "the parochial class") as shown in Figure 22; however, using other procedures the same example persons may be assigned parameter data

reflecting the extent or degree to which they are "cosmopolitan" or "parochial" or some of each.

In the example process that generates the information shown in Figure 23A, data for several individuals has been arranged in a case (row) by variable (column) matrix and using means known to those skilled in the arts, subjected to principal components analysis with subsequent Varimax axis rotation. Components with eigenvalues >1.0 were retained for subsequent rotation and use. After rotation, each case was assigned a score on each retained (and rotated) component. Each score indicates the extent to which the case has the characteristic represented by the component.

The hypothetical data in Figure 23A shows how strongly each variable (the column of the input matrix) is correlated with the underlying characteristic or component. For example, "region of the US" and "Family income" are highly correlated while "owns a sports utility vehicle" is not.

Using results such as these plus the input data matrix, a score is assigned to each case indicating the extent to which they possess the trait, attribute, characteristic indicated by each factor or component. The hypothetical data in Figure 23B shows how strongly each case -- a person or thing -- is a member of the class, and/or possesses the underlying variable represented by each component. A higher score shows that example case 1 has more of the underlying component 1 than does example case 3, whose score is close to zero. Components

(factors) may be bipolar with a zero point and cases whose scores may be positive, negative or zero. Hypothetical example case 5 has a negative score on this component.

This component score information may be used by the
5 matching and classification utility 900 to define certain other classes, such as "the class consisting of the top 5% of those who are cosmopolitan," that is, the 5% with the highest scores on example component 1. The original scores and/or derivative class assignments may be included on attribute records with attribute and/or class
10 information harvested from other sources and/or through other processes.

DATA HARVESTING

Example Steps For Collecting Appliance Related Data

Figure 24 shows example steps performed by the matching and
15 classification utility 900 to collect appliance attribute data. In this example, an electronic appliance 100 may have certain information associated with it. For example, a VDE administrator 800 may initialize appliance 100 with certain information upon appliance installation. In this example, the matching and classification utility
20 900 can collect this appliance attribute data and use it as part of a matching and/or classification and/or selection process. As shown in Figure 24, the matching and classification utility 900 may initially specify desired appliance attribute fields or other information characteristics the utility is going to collect (Figure 24, block 1502).

The information to be collected depends upon the purpose and use to which the matching and classification utility 900 is to put the information to. The matching and classification utility 900 may use a data dictionary or other mechanism for specifying the desired types of appliance information it is going to collect.

The matching and classification utility 900 next determines whether it already possesses the desired information for this particular appliance 100 (Figure 24, block 1504). For example, the information may have been previously gathered as part of a prior process. If the information is already available, the matching and classification utility 900 sends one or more events to a "create appliance attribute record" method to process the previously gathered data (Figure 24, block 1506). (In all these processes, if the appropriate method is has been sent previously to a VDE installation, only the associated administrative events necessary to activate the method need to be sent in the VDE container.) Alternatively, if the desired data is not already available ("no" exit to decision block 1504, Figure 24), the matching and classification utility 900 performs the other steps shown in Figure 24 to collect the appliance attribute data.

These collecting steps shown in Figure 24 may include sending a VDE container 152 with a "create appliance attribute record" method, and one or more associated administrative events to activate the method, to the VDE administrator 800 (Figure 24, block 1508). The next step (Figure 24, block 1510) may be performed by the VDE

administrator 800 processing the administrative event(s) using the "create appliance attribute record" method to determine whether the administrator already has the desired information for the particular electronic appliance 100. If the operation is successful ("yes" exit to
5 decision block 1512, Figure 24), the VDE administrator 800 may send, to the matching and classification utility 900, a VDE container 152 containing one or more administrative events and the appliance attribute record (Figure 24, block 1514). If the operation is not successful ("no" exit to decision block 1512, Figure 24), the "create
10 appliance attribute record" method operating at VDE administrator 800 may, in this example, collect the data directly from the electronic appliance 100 by sending a VDE container to the appliance, the container containing a "create appliance attribute record" method and one or more associated administrative events (Figure 24, block 1516).
15 The appliance 100 may itself process the administrative event(s) using the "create appliance attribute record" method (Figure 24, block 1518) to produce the required appliance attribute record. Appliance 100 may then send a VDE container 152 containing the appropriate administrative event(s) and the appliance attribute record
20 to the matching and classification utility 900 (Figure 24, block 1520).

In another example, blocks 1508-1514 may be bypassed entirely, and the matching and classification utility 900 may (assuming appropriate authorizations are in place) perform block 1516 to send a container 152 with one or more administrative events

and the "create appliance attribute record" method directly to the electronic appliance 100.

Figures 25(A) and 25(B) together show example steps performed by the "create appliance attribute data" method shown in Figure 24, blocks 1506, 1510 and 1518. As disclosed in "Ginter et al", the actual processing steps are performed by one or more load modules associated with the method. This example method (which, as explained above, may be performed by the matching and classification utility 900, the VDE administrator 800, the electronic appliance 100, any other electronic appliance, or a combination of any or all of these) first locates the site configuration record(s) corresponding to the electronic appliance for which appliance attribute data is to be collected (Figure 24A, block 1522). This site configuration record(s) may, for example, be stored in the electronic appliance secure database. The method next locates the permissions record for the site configuration record(s) (Figure 24A, block 1523). The SPE next determines, based upon the permission record(s), whether the method has permission to access and/or use the site configuration record(s) (Figure 25A, block 1524). If the method does not have the appropriate permission ("no" exit to decision block 1524, Figure 25A), the protected processing environment 154 reports the failure and reason for the failure, and the method writes an associated audit record (Figure 25A, block 1525, 1526) and goes on to process a user configuration record(s). On the other hand, if the method does have permission to use the site configuration record(s) ("yes" exit to

decision block 1524, Figure 25A), the method copies the required fields from the site configuration record(s) to create an appliance attribute record, and may then write an appropriate audit record (Figure 25A, block 1527).

5 After completing processing of site configuration records, the method then locates the user configuration record(s) corresponding to the electronic appliance for which appliance attribute data is to be collected (Figure 25B, block 1528). This user configuration record(s) may, for example, be stored in the electronic appliance secure
10 database. The protected processing environment 154 next locates the permissions record for the user configuration record(s) (Figure 25B, block 1529). The protected processing environment 154 determines next, based upon the permission record(s), whether it has permission to access and/or use the user configuration record(s) (Figure 25B,
15 block 1530). If the method does not have the appropriate permission ("no" exit to decision block 1530, Figure 25B), the protected processing environment 154 reports the failure and reason for the failure, and the method writes an associated audit record (Figure 25B, block 1531, 1532) and exits the process. On the other hand, if the
20 method does have permission to use the user configuration record(s) ("yes" exit to decision block 1530, Figure 25B), the method copies the required fields from the user configuration record(s) to create an appliance attribute record, and may then write an appropriate audit record (Figure 25B, block 1533). The method may then, if desired,
25 create a new permissions record corresponding to the appliance

attribute record (Figure 25B, block 1534). If a new permissions record is desired, the method may include appropriate "shared secrets," expiration interval(s), and/or other data in an associated MDE to, for example, provide a basis for controlling access, use, and
5 modification of the permissions record.

Figures 26A-26C show examples of appliance attribute records created by Figure 25B, block 1532. Figure 26A shows an example appliance attribute record that may include, for example, an appliance identification field 1536(1) and any number of attribute fields
10 1538(1)...1538(n). Figure 26B shows a more specific appliance attribute record example including an appliance ID field 1536(1), an operating system field 1538(A), a country field 1538(B), a state field 1538(C), a VDE administrator organization field 1538(D), a VDE version field 1538(E), and a VDE maintenance level field 1538(F).
15 Figure 26C shows that different encodings may be used for any/all of the various attribute fields 1538.

Example Steps for Collecting Demographic Data

Figures 27A, 27B show example steps for collecting demographic data. In this example, the matching and classification
20 utility 900 initially specifies demographic data fields it is interested in (Figure 27A, block 1540). The matching and classification utility 900 next determines whether the required data is already available to it (e.g., based on previous inquiries responded to by the user 95) (block 1542, Figure 27A). If the required data is already available ("yes"

exit to decision block 1542, Figure 27A), the matching and classification utility 900 may send one or more events to a "create demographic attribute record" method to process the data (block 1544, Figure 27A).

5 On the other hand, if the required data is not available to the matching and classification utility ("no" exit to decision block 1542, Figure 27A), the matching and classification utility may send a container 152 to another commerce utility system 90, the container including one or more administrative events associated with a
10 "demographic data query" method and a "create demographic attribute record" method (Figure 27A, block 1546). The other commerce utility system 90 may then process the one or more events using the "demographic data query" method, and write an associated audit record (Figure 27A, block 1548). It may determine whether the
15 required demographic data is available (Figure 27A, block 1550). If the information is available ("yes" exit to decision block 1550, Figure 27A), the commerce utility system 90 may process one or more events using a "create demographic attribute record" method in order to analyze the available demographic data, and write a corresponding
20 UDE audit record (Figure 27A, block 1552). The other commerce utility system 90 may then send appropriate one or more administrative events and the demographic data attribute record within a container 152 to the matching and classification utility 900 (Figure 27A, block 1554)).

If the required demographic data is not available ("no" exit to decision block 1550, Figure 27A), the commerce utility system 90 may send an administrative event to the matching and classification utility system 900 within a container 152 informing the matching and classification utility that the required data is not available (Figure 27B, block 1556). The matching and classification utility 900 may then send a "demographic data query" method and a "create demographic attribute record" method within a container 152 (along with appropriate administrative events to activate such methods) directly to the user 95 about which demographic information is to be collected (Figure 27B, block 1558). The user's electronic appliance 100 may, in response, process the one or more events using the "demographic data query" method, which may write an associated audit record (Figure 27B, block 1560). If the required data is not collected ("no" exit to decision block 1562, Figure 27B, the user's appliance 100 may send a "failure" message associated with the appropriate administrative event to the matching and classification utility 900, and write an associated audit record (Figure 27B, block 1564, 1566). If the required demographic data is successfully collected ("yes" exit to decision block 1562, Figure 27B), the user's electronic appliance may process one or more events using the "create demographic record" method supplied by step 1558, which may write an associated audit record (Figure 27B, block 1568). The electronic appliance may then send appropriate administrative events and the

demographic attribute record to the matching and classification utility within one or more containers 152 (Figure 27B, block 1570).

Figure 28 shows an example questionnaire "pop-up" screen that may be displayed by the user's appliance 100 as a result of processing
5 events using the "demographic data query" method of block 1548, Figure 27A, and/or block 1560, Figure 27B. In this example, information is collected directly from a user 95 by displaying a questionnaire on a display device that is part of the user's appliance 100. The questionnaire may ask for various demographic information
10 such as:

- name
- address
- city
- state
- 15 • zip code
- gender
- date of birth
- education level
- marital status
- 20 • number of children

- age of first child
- gender of first child
- other information

The user is requested to provide the information by filling in the
5 various fields within the questionnaire. The questionnaire may assure
the user that all information the user provides will be treated as
confidential, by, for example, disclosing the rules that will be
associated with access to and use of the information.

Steps similar to those shown in Figure 25A, 25B may be
10 performed to create a demographic attribute record based on the
results of a demographic data query. Figure 29A-29C show examples
of different user demographic attribute information records resulting
from this process. Figure 29A shows an example demographic
attribute record 1572 including a user ID field 1574 and any number
15 of attribute fields 1576(1), ... 1576(n). Figure 29B shows a more
specific example of a demographic attribute record including, for
example, a user ID number 1574, a gender attribute field 1576(A), an
age field 1576(B), a highest educational level field 1576(C), a
citizenship field 1576(D), a country of residence field 1576(E), a
20 district field 1576(F), a city field 1576(G), and a street address field
1576(H). Figure 29C shows a different detailed encoding example
for demographic attribute record 1572-1.

Example Steps for Collecting Psychographic Data

Figure 20 shows example steps that may be performed to collect user psychographic data. In this particular example, the matching and classification utility 900 initially specifies desired psychographic data it requires in order to perform a particular classification/matching process (Figure 30, block 1580). The matching and classification utility 900 determines if the required data is already available to it (Figure 30, block 1582). If the required data is already available ("yes" exit to decision block 1582, Figure 30), the matching and classification utility 900 sends one or more events to a "create psychographic attribute record" method in order to analyze the available data and provide appropriate psychographic attributes (Figure 30, block 1584). If, on the other hand, the required data is not available to the matching and classification utility 900 ("no" exit to decision block 1582, Figure 30), appropriate steps are performed to collect the required data. In this example, the matching and classification utility 900 may send a "psychographic data query" method and a "create psychographic attribute record" method within one or more containers 152 (along with appropriate administrative events to activate such methods), to appropriate repositories that may contain the required data (Figure 30, block 1586). If the required data is available from the repositories ("yes" exit to decision block 1588, Figure 30), then an electronic appliance at the repository (in this example) processes one or more events using the "create psychographic attribute record" method supplied by block 1586 in

order to generate an appropriate attribute record(s) containing the attribute information the matching and classification utility 900 is interested in (Figure 30, block 1590). This information, and associated one or more events, may be sent to the matching and classification utility 900 within one or more containers 152 (Figure 30, block 1592).

If the required data is not available from the repository ("no" exit to decision block 1588, Figure 30), then the repository may send a "failure" message associated with one or more administrative events to the matching and classification utility 900 within a container 152 (Figure 30, block 1594). The matching and classification utility 900 may, in response, send one or more administrative events, a "collect psychographic data" and "create psychographic attribute record" method directly to the user's electronic appliance 100 within one or more containers 152 (Figure 30, block 1596). The user's electronic appliance 100 may, in turn, process the events using the "collect psychographic data" and "create psychographic attribute record" methods (Figure 30, block 1598, 1600), and send the resulting attribute data record(s) to the matching and classification utility (Figure 30, block 1592).

Figure 31 shows an example psychographic questionnaire "pop-up" screen that may be displayed to the user 95 upon performance of Figure 30, block 1598. This questionnaire may

collect various psychographic information from the user, including for example:

- mood information
- emotion information
- 5 • habit information
- behavioral information
- cognitive information
- medical information
- physical information
- 10 • patient information
- counseling information
- aptitude information
- testing information
- other information
- 15 • combinations of types of information.

The questionnaire may inform the user that all information collected will be treated as "confidential," and may also, if desired, indicate that the user will be compensated for providing the information.

Figures 32A-32C show some example user psychographic attribute information records 1602 that may be created by Figure 30, block 1584, 1590 and/or 1600. Figure 32A shows that a psychographic attribute record 1602 may include a user ID field 1604 and any number of attribute fields 1606(1), ... 1606(n). Figure 32B shows a more detailed user psychographic attribute record 1602 example including a user ID field 1604, a field 1606a indicating whether the user is introverted or extroverted, a field 1606b indicating whether the user is a sensing or intuitive person, a field 1606c indicating whether the user is primarily a thinking person or a feeling person, a field 1606(d) indicating whether the user is primarily a judging person or a perceiving person, and a field 1606(e) indicating an overall psychographic / behavioral profile such as, for example, the iVALS standard provided by SRI. Figure 32C shows a different kind of encoding (in this case, binary) for the various attributes 1606.

Example Method for Determining Attributes Based on Available Rules and Consequences

Figure 33 shows an example method for determining attributes based on available rules and consequences. The matching and classification utility 900 may first send one or more administrative events and a "send permission records" method request to an electronic appliance 100 within one or more containers 152 (Figure 33, block 1610). In response, the appliance may process the events using the method, which may write an associated audit record (Figure 33, block 1612). If this step is performed successfully ("yes" exit to

Figure 33, decision block 1614), the appliance sends appropriate administrative events and the requested permission records to the matching and classification utility 900 within one or more containers 152, and the method writes an associated audit record indicating it has
5 performed this transaction (Figure 33, block 1616). The matching and classification utility may process events using a corresponding "create attribute record from permission records" method to obtain attributes from these provided permission records (Figure 33, block 1618). If the step of block 1612 failed (as indicated by the "no" exit
10 to decision block 1614, Figure 33), the method may send a "failure" message to the matching and classification utility 900, and write an associated audit record (Figure 33, block 1620).

Figure 34 shows a variation on the Figure 33 example in which the appliance 100 rather than the matching and classification utility
15 900 creates the rules attribute record based on a "create rules attribute record from permissions records" method supplied by the matching and classification utility, and then sends the rules attribute record to the matching and classification utility (see Figure 34, blocks 1622, 1624).

20 **Example Method to Construct Attribute Records from Permissions Records**

Figures 35A, 35B show example steps for constructing attribute records from permissions records. The steps shown in Figure 35A, 35B may, for example, be performed as part of the method shown in
25 block 1618 of Figure 33.

In this example method 1618, the matching and classification utility 900 may first check relevant permissions to ensure that it has the authority to perform the desired transactions (Figure 35A, block 1630). For example, the matching and classification utility 900 may
5 examine a permissions record about the permissions records it has collected, this permissions record it is examining indicating what entities have authority to perform operations with respect to the permissions record to be analyzed. Presuming the matching and classification utility 900 has the appropriate permission, it opens a
10 permissions to be analyzed (Figure 35A, block 1632), and performs a sequence of steps 1634-1650 to extract relevant information from the permissions record. For example, information from the permissions record header can be copied into the attribute record (Figure 35A, block 1634), and then the method may locate the rights record header
15 (block 1636, Figure 35A). Information from the rights record header may be copied into the attribute record (block 1638, Figure 35A), along with the identifier for the corresponding right(s) (blocks 1640, 1642, Figure 35A). The process may then recursively locate and harvest data from each method header contained within the rights
20 record (blocks 1644, 1646, 1648, Figure 35B). The process may recursively repeat steps 1638-1648 for each rights record within the permissions record (as tested for by decision block 1650, Figure 35B). Finally, the entire process of steps 1632-1652 may be performed recursively for multiple permissions records to harvest the

appropriate rules and consequences information from each of a number of permissions records (see decision block 1652, Figure 35B).

Figure 36 shows example steps to perform the "check permissions" operation shown in Figure 35A, block 1630. In this example, the process locates the permissions record from which information is desired to be harvested (Figure 36, block 1660), and then determines whether there is a permissions record for that permissions record (Figure 36, decision block 1662). If there is no permissions record that controls that permissions record (and assuming that authorization or additional permission is required to access the permissions record from which information is to be harvested) (Figure 36, "no" exit to decision block 1662), the process reports failure, writes an audit record, and ends (Figure 36, blocks 1664, 1666, 1668). On the other hand, if there is a permissions record that controls access to the permissions record from which information is to be harvested ("yes" exit to decision block 1662, Figure 36), the process determines whether that permissions record for the permissions record enables usage by the matching and classification utility 900 (Figure 36, decision block 1670). If the matching and classification utility 900 does not have permission ("no" exit to decision block 1670, Figure 36), the process reports failure, writes an audit record to that effect, and ends (blocks 1672, 1674, 1676, Figure 36)). On the other hand, if the matching and classification utility 900 is granted permission ("yes" exit to decision block 1670, Figure 36), the process accesses and uses the permissions record for the

permissions record from which information is to be harvested (Figure 36, block 1678).

Figures 37A-37C show examples of attribute records containing information harvested from permissions records. Attribute record 1680-1 shown in Figure 37A includes a user identification field 1682, an object identification field 1684, and any number of attribute fields 1686(1), ..., 1686(n). The attribute record 1680-2 shown in Figure 37B includes, as a more detailed example, a user ID number field 1682, an object ID field 1684, a right ID field 1686a, a method identifier field 1686b, another right ID field 1686c, and corresponding method type fields 1686(d), a further right ID field 1686e and two corresponding method attribute fields 1686f, 1686g, a further right ID field 1686h and corresponding method attribute fields 1686i, 1686j.

Figure 37C shows a different example in coding for the Figure 37B example attribute record.

Example Steps for Assembling Rules and Consequences

Figure 38 shows example steps for assembling rules and consequences. In this example, the matching and classification utility 900 may send one or more administrative events and a "get user rights table" method within a container 152 to an electronic appliance (Figure 38, block 1690). The electronic appliance 100 processes the one or more events using the "get URT" method, which may writes an

associated audit record (Figure 38, block 1692). The method then determines whether the associated URT records are available (Figure 38, decision block 1694). If the records are not available ("no" exit to decision block 1694, Figure 38), the method sends a failure notice to
5 the matching and classification utility 900, and writes an associated audit record (block 1696, Figure 38). If, on the other hand, the URT records are available ("yes" exit to decision block 1694, Figure 38), the method packages the URT records and associated one or more administrative events into a container 152, and sends the container to
10 the matching and classification utility 900 (Figure 38, block 1698). The matching and classification utility 900 may then process the administrative events using a "create attribute record from URT" method in order to extract or harvest the information from the URT(s) (Figure 38, block 1700).

15 Figure 39 shows another example sequence of steps for assembling rules and consequences. In this example, the matching and classification utility 900 sends one or more administrative events and a "create attribute record from URT" method to the electronic appliance 100 that stores or has access to the user rights table
20 information (Figure 39, block 1702). The appliance then processes the events using the method sent to it, and the method writes associated audit information as it processes (Figure 39, block 1704). If the URT records are available and the step completes successfully ("yes" exit to decision block 1706, Figure 39), the method sends the
25 resulting URT attribute record(s) and one or more administrative

events to the matching and classification utility within a container 152, and writes corresponding audit information to an audit trail (Figure 39, block 1710). On the other hand, if an error condition arises either because the URT records are not available or because the method for some other reason cannot complete successfully, the
5 method sends a failure notice within a container 152, and writes an associated audit record ("no" exit to decision block 1706, Figure 39, block 1708).

Figures 40A, 40B show example steps performed by blocks
10 1700, 1704 to "create attribute record from user rights table." The method begins by checking associated permissions for the user rights table records (Figure 40A, block 1720). Assuming that appropriate user and/or group permission is available, the method next locates the user rights table (Figure 40A, block 1722), and then begins
15 recursively analyzing the user rights table information to harvest desired attribute information from it (Figure 40A, blocks 1724 and following). In this particular example, the method locates the user rights table record (block 1724, Figure 40A, and then locates the first rights record header within the first user choice record within the
20 URT record (blocks 1726, 1728, Figure 40A). The method copies rights record header information to the attribute record (block 1730), and then locates the right identifier and copies that to the attribute record (blocks 1732, 1734). The method then recursively locates each method header within the user rights table right record, and
25 copies corresponding attribute information to the attribute record

(blocks 1736, 1738, 1740, Figure 40B). Steps 1728-1740 are performed recursively for each rights record within the user choice record (see Figure 40B), decision block 1742), and the above steps are performed recursively for each user choice record within the user rights table (see decision block 1744, Figure 40B). Additionally, steps 1724-1744 are performed recursively for each user rights table record within the user rights table (see Figure 40B, decision block 1746). As a last example step, the method creates a permissions record that controls access and use of the attribute record it has created (Figure 40B, block 1748).

Figure 41 shows example steps performed by the check permissions block 1720 shown in Figure 40A. For example, the sequence of steps may begin by locating a corresponding permissions record (Figure 41, block 1750) and then determining whether there is a permission record corresponding to the corresponding user rights table entry (Figure 41, decision block 1752). If there is no such entry ("no" exit to decision block 1752), the method may report failure, write an audit record, and end (blocks 1754, 1756, 1758, Figure 41). If there is a corresponding permissions record ("yes" exit to decision block 1752, Figure 41), then the permissions record may be examined whether it enables usage for the matching and classification utility 900 (decision block 1760, Figure 41). If the permissions record does not enable usage by the matching and classification utility 900 ("no" exit to decision block 1760, Figure 41), the method may report a failure, write an audit record, and end (blocks 1762, 1764, 1766,

Figure 41). On the other hand, if the matching and classification utility 900 does have the required permissions to enable usage ("yes" exit to decision block 1760, Figure 41), the method may access the permissions record (if any) for the user rights table for use in
5 controlling access to the user rights table itself (block 1768, Figure 41).

Figures 42A-42C show example rights attributes records 1770 that may be obtained from the processes above. The Figure 42A example rights attribute record 1770-1 includes a user or group ID
10 field 1772, an object ID field 1774, and any number of attribute fields 1776(1), ... , 1776(n). The more detailed example rights attribute record 1770-2 shown in Figure 42B includes a user ID number field 1772, an object ID field 1774, a right ID field 1776a and
15 1776c and corresponding method attribute field 1776d, a right ID field 1776e and corresponding method attribute fields 1776f, 1776g, and another right ID field 1776h and corresponding method attribute field 1776i.

Figure 42C shows how the rights attribute record 1770 can be
20 encoded numerically as opposed to using characters, as one example.

Example Steps for Assembling Usage Audit Records

Figure 43 shows example steps for assembling usage audit records for purposes of matching and/or classification. In this example, the matching and classification utility 900 may send one or

more administrative events and a "get audit records" method to a VDE appliance 100 within a container 152 (Figure 43, block 1780). The appliance 100 may process the one or more events using the "get audit records" method, which may write an associated audit record (block 1782, Figure 43). If the audit records are not available ("no" exit to decision block 1784, Figure 43), the method may send a failure notice within a container to the matching and classification utility 900, and may then write an associated audit record (Figure 43, block 1786). On the other hand, if the audit records are available ("yes" exit to decision block 1784), the method may send one or more administrative events and the audit records within a container 152 to the matching and classification utility 900, and write an associated audit record (block 1788, Figure 43). The matching and classification utility 900 may then process the one or more administrative events using a "create attribute record from audit record" method in order to extract or harvest the information from the audit record it will use to perform matching and/or classification (block 1790, Figure 43).

Figure 44 shows another sequence of example steps that may be used to assemble usage audit records for purposes of matching and/or classification. In the Figure 44 example, the matching and classification utility 900 sends one or more administrative events and a "create attribute record from audit record" method to an electronic appliance 100 within one or more containers 152 (Figure 44, block 1792). The appliance 100 may then process the one or more administrative events using the "create attribute record from audit

record" method, which may write an associated audit record (block 1794, Figure 44). The method may determine, in this process, whether audit records are available (Figure 44, decision block 1796). If no audit records are available ("no" exit to decision block 1796),

5 the method may send a failure notice to the matching and classification utility 900 (Figure 44, block 1798). On the other hand, if audit records are available, the method may create the corresponding usage attribute records and associated administrative event(s), package them into a container 152, send the container to the

10 matching and classification utility 900, and write corresponding audit records (Figure 44, block 1799).

Figures 45A, 45B show example steps for performing the method (shown in Figure 44, block 1794, for example) of creating attribute record(s) from audit records. In this example, the method

15 first locates the audit records in a secure database or other storage facility (Figure 45(A), block 1800). The method next selects an appropriate UDE audit record to analyze (Figure 45(A), block 1802), and determines whether a permission record is available that applies to this particular audit record (Figure 45(A), decision block 1804). If

20 a permissions record is required and is not available, the process reports failure, writes an associated audit record, and ends (Figure 45 blocks 1806, 1808, 1810). If, on the other hand, a required permissions record is available ("yes" exit to decision block 1804, Figure 45), the process determines whether the permissions record

25 grants the device or process permission to use the audit record(s) for

this particular purpose (decision block 1812, Figure 45). If such permission is not available ("no" exit to decision block 1812, Figure 45A), the process reports failure, writes an associated audit record, and terminates (Figure 45A, blocks 1814, 1816, 1818).

5 If any applicable permissions record is available and grants permission to the matching and classification utility 900 ("yes" exit to decision block 1812), the process determines multiple audit records need to be analyzed together as an overall event (Figure 45A, decision block 1820). For example, an "atomic transaction" in which
10 multiple steps are performed to achieve an overall result may have multiple audit records (e.g., from multiple appliances 100) that may need to be analyzed together in order to make sense out of the overall transaction. As another example, an object may have subparts (e.g., sub-objects) on which operations can be performed – but it may be
15 important for matching and/or classification purposes to analyze the results of such multiple operations together in order to determine appropriate attribute(s) for matching and/or classification. If it is necessary to aggregate multiple audit records together for analysis (decision blocks 1820, 1822, Figure 45A), then the process proceeds
20 to analyze those audit records together and create corresponding summary transaction information (Figure 45A, block 1824).

The process next determines whether it needs to produce aggregated audit statistics in order to perform the associated matching and/or classification operation (Figures 45A, 45B, decision block

1826). For example, multiple operations may be performed on a certain object. It may be important to know statistics about such operations (e.g., the number of times the object was opened on a certain day, the number of users who opened the object in a certain
5 time period, etc.). If such aggregated statistics are required ("yes" exit to decision block 1826, Figure 45B), the process proceeds to create such aggregated statistics (block 1828, Figure 45B).

The process next copies selected audit record information to an audit attribute record (Figure 45B, block 1830). The process then
10 determines whether it needs to process more audit records (decision block 1832, Figure 45B). If more audit records are required to be processed ("yes" exit to decision block 1832, Figure 45B), control returns to Figure 45A, block 1802 to select the next audit record. Otherwise ("no" exit to decision block 1832, Figure 45B), the process
15 creates a permissions record associated with the newly created attribute record(s) (Figure 45B, block 1834), and completes.

Figures 46A, 46B show example usage attributes/statistic records that the Figure 45A-B process may create. The Figure 46A attribute record 1830-1 may include, for example, a user ID 1832, an
20 object ID 1834, and any number of attribute fields 1836(1), ... , 1836(n). The more detailed Figure 46B example attribute record 1830-2 includes a user ID number 1832, an object ID 1834, a right ID 1836a and associated method characteristic 1836b, another right ID 1836c and associated method 1836d and associated statistic 1836e, a

further right ID 1836f and associated method attribute 1836g, another right ID 1836h and associated methods 1836i, 1836j, and associated additional attributes 1836k-1836o. The characteristics shown in fields 1836k-1836o could, for example, be derived from an aggregate
5 of any number of individual audit records recording individual transactions associated with the object identified in field 1834.

EXAMPLES

The following are some non-limiting examples of how Matching and Classification Utility 900 may be useful in certain
10 applications.

Example: Matching and Classification Utility 900 Can Support Narrowcasting or "Push" Distribution Models Based On Classes

15 Interactions with content, transactions, and other events on the World Wide Web are mainly driven today by following chains of hypertext links, using various search engines, and/or indexes, to say nothing of just plain luck and persistence, to find interesting and/or useful content and/or services. Time consuming and generally
20 inefficient, these search activities share in common the feature that each consumer must intentionally "pull" desired content from a Web site to their computer after successfully identifying specific content or services of interest at that time. The present inventions also support "pull" models—a topic to be addressed shortly. However, the present

inventions also support narrowcasting or "push" models of content distribution as well.

In one example, the matching and classification utility 900 can facilitate much more automated and therefore more efficient and effective content creation, access and/or distribution services that "push" information and/or services to users. Example Figure 47 shows an example "information push" model 2000 in which an arbitrary number of users 2001(1)-2001(n) each have a VDE node (e.g., a protected processing environment 154) installed on their appliances. These example appliances may be of any kind, including computers, so-called Web television or Web-TV, DVD appliances with some form of backchannel, a settop box with a "back channel", and so on.

Perhaps with the permission of the user or other authority, such as an administrator within an organization, the VDE node collects various usage information or "info exhaust" according to the rules and usage consequences provided by one or more value chain participants. At times specified by default and/or by the associated rules and consequences, audit records are sent, in this example, in VDE containers 2006(1)-2006(n) to a usage clearinghouse 300, which in turn, may send all or a portion of these audit records in a VDE container 2008 to the matching and classification utility 900. The audit records may contain rights management information, including, but not limited to the amount of usage, the amount paid, if any, the

payment method used, if any, VDE control sets, and/or data that identify various attributes of the node, user, and/or known and/or used object(s). The audit records may also contain information about objects known to the VDE node (objects with PERC records - see
5 Figures 35A, 35B and associated discussions) and/or objects that have been used (objects with URT entries - see Figures 40A-40B and associated discussions) on the node.

The matching and classification utility 900 may also receive from one or more providers 2010 content objects 2003 themselves,
10 for example, information in text format and/or metadata 2005 associated with content objects. Using at least one classification method, the matching and classification utility 900 may create at least one object class hierarchy, object class, object classification scheme, object category and/or object category scheme using at least some
15 rights management information and assign at least one object to at least one category and/or class.

The matching and classification utility 900 takes the usage information and other rights management information received from the VDE nodes and/or other information sources and may create at
20 least one category and may assign at least one node and/or user to a category and/or class. In Figure 47, the matching and classification utility 900 sends a VDE container 2002 to content provider 2010 with information showing the classes of content used by one or more nodes and/or users along with a request that the provider 2010 send similar

content back to one or more users 2001. At least one content provider
2010 then sends at least one VDE container 2004 to user A with
content and/or information about available content that may be of
interest to user A given the history of content usage as reflected in
5 VDE audit records and/or other rights management information. In
this "push" example, classes of content or information about available
content may be pushed automatically from (a class of) content
providers to one or more members of class of users and/or nodes.
Consequently, users do not have to search as intensely, if at all, for
10 content of interest to them.

In this example, user A receives content that may be most like
content the user has already used, perhaps like content used most
frequently in the recent past. The present inventions also support the
matching and classification utility 900 and/or content provider
15 sending content that is in a class or classes more distant from topics
of prior and current interest to a particular user and/or group of users.
Certain classification methods familiar to those skilled in the arts may
provide quantitative indicators of distance that, in turn, may be used
as at least one criterion for selection.

20 In another example, matching content to users and/or nodes
may be based in part on class assignments that are in turn based in
part on information concerning user preferences solicited by the
matching and classification utility 900 or other value chain
participant, such as a market research firm, advertising agency,

provider, distributor, VDE administrator 800, or other Commerce Utility System.

Although the matching and classification utility 900 and/or content provider may send "more of the same," in another example
5 the present inventions support providers at least occasionally sending content more distantly related to the user's apparent interests to determine if the user's circle of interest might be a little larger than that indicated by past usage and other, related rights management information alone.

10 In another example, providers may from time to time send content unrelated to the user's apparent interests that may nevertheless reflect the interests of persons and/or groups sharing at least one attribute with the user. For instance, the matching and classification utility 900 may, by sending a VDE container with
15 appropriate user and content class information, suggest to a provider that user A receive content similar to content used by another member or members in the same group or class as user A. In one example, the matching and classification utility 900 may suggest sending business information related to a particular vertical market segment because
20 others in the same class as user A have paid attention to that market.

In support of various content narrowcasting or "push" models, the matching and classification utility 900 may provide content class related information to a "subject switch" or "subject mapper," which in turn, matches participants desiring information in one or more

specified classes with one or more sources of content in the requested class or classes.

The non-limiting subject switching example 2050, Figure 47A, shows a number of customers 2053(1)-2053(n) each with an
5 appliance 2052(1) -2052(n) such as a personal computer. Other arrangements may include appliances such as a WebTV interface and/or an intelligent "settop box" connected to an interface device that uses one or more (digital) TVs for display. Still other arrangements may include an NC computer without a local hard disk
10 logically connected to at least one server, a personal digital assistant with a network connection, and/or any other appliances with suitable processing, storage, and communications capabilities.

Referring again to Figure 47A, each customer appliance 2052 may have a VDE secure node installation 2054 incorporating a
15 protected processing environment 154, as described in "Ginter et al", and messaging services software 2058 that manages communications with other appliances. (In an alternative example, some appliances may lack secure nodes or sufficiently secure nodes, and receive appropriate one or more protected processing environment 154 based
20 services from one or more servers and/or peers.) These appliances may be located in the same physical and/or logical environment, such as on the same local area network, and/or may be distributed across wide area networks such as multi-location corporate Intranets and/or the Internet itself. Among other tasks, messaging services 2058

"listens" for messages destined for that particular appliance or for broadcast messages intended for at least one appliance in the set of appliances that receive the broadcast. In certain instances no appliance may actually be "listening." In other examples, the

5 messaging services 2058 may incorporate delivery assurance capabilities that assure delivery through use of explicit or implicit acknowledgments of receipt combined with the ability to retransmit information that has not been acknowledged. Messaging services

2058 may be designed such that an operator may select from one or

10 more delivery assurance levels, for example "no receipt acknowledgment," "retry n times, then notify operator if not received," "retry until a certain date/time, then notify operator if not received," "retry n times and/or until a certain date/time, no operator notification necessary," et cetera.

15 Messaging services 2058 may use the secure node 2054 to package one or more messages in a VDE secure container that may also include one or more sets of rules and usage consequences that may be associated with one or more messages in the container as described in "Ginter et al". In this example, messaging services 2058

20 then sends the secure container to one or more destinations using, for instance, TCP/IP and/or some other network protocol(s). Also, messaging services 2058 may broadcast a VDE container to one or more other customers 2053.

In this example, a customer 2053 uses application 2060 to persistently request or "subscribe" to one or more particular classes of content. For example, a highly detailed class might include "business information concerning the US market share of PC vendors, information in text format, costing less than a dollar per item, and for which the subscriber receives the right to excerpt at least one whole paragraph, provided that the excerpted amount constitutes less than 25% of the entire item based on word count." This same and/or another application may also be used to interact with instances of content in the desired class, for example, by displaying information on a computer screen and/or another output device in accordance with the rules and usage consequences associated with that item. If a user no longer has an interest in one or more classes, they may also use the same (or similar) application 2060 to "unsubscribe" from a particular subject, or specify further narrowing or broadening criteria to adjust the flow of content from one or more classes.

Items in the desired class or classes may be available from more than one content source 2074(1)-2074(n). To enhance the efficiency of locating content of interest to the subscriber or other participant, the matching and classification 900 may have created such a class definition and assigned one or more content items to that class. In one example, the matching and classification 900 may have sent one or more methods, and administrative events necessary to invoke the method(s), in a VDE secure container to one or more content sources 2074 where the classification methods are executed.

Such methods may, for example, assign content items to one or more classes. One or more object and/or item identifiers may have been transmitted to the matching and classification utility 900 along with class assignments for each item. If the matching and classification utility 900 has not previously created the desired class and assigned items to it, in response to a request from the subject switch 2051, the matching and classification utility 900 may do so using any appropriate combination of one or more such classification methods and procedures. The matching and classification utility 900 may create at least one object class hierarchy, object class, object classification scheme, object category and/or object category scheme using at least some rights management information and assign at least one object, item, and/or subscriber to at least one category and/or class.

Subsequent to receipt of the request and/or "subscribe" message from the customer 2053, the subject switch 2051 may query the matching and classification 900 for content sources 2074 that have items in the desired class or classes. The matching and classification utility 900 may respond with information indicating known sources of information in the desired class(es), if any. The subject switch 2051 may then send a VDE container to the appropriate content source(s) 2074 indicating that certain customers 2053 are interested in items in the desired class and that the content source 2074 should send items in this class to this customer 2053

and/or groups of customers, and/or include such content in broadcasts which may be received by such subscribers.

The content sources 2074 may have already received class definitions and class assignment information from the matching and classification utility 900 and/or may have received from the matching and classification utility 900 or another party to the transaction one or more classification methods and associated events to invoke one or more of these methods to perform classification and/or class assignment processes.

10 In one arrangement, the content source 2074 may send the desired items directly to the subscribing customers 2053 by using the messaging services 2058 and subject switch 2051 to publish each item as it becomes available for distribution. In another example, the content source 2074 may broadcast the information such that
15 subscribers' messaging services 2058 will have the opportunity to access the such items from a broadcast. The content source 2074 may call on messaging services 2058 to use the VDE secure node to package the item in a VDE container along with associated rules and usage consequences and then send that container such that one or
20 more listening messaging services 2058 on other appliances 2052(1)-2052(n) will receive it. Based on subject information contained in the message header and/or in unencrypted (but optionally protected for integrity) areas of the VDE container, the listening messaging services 2058 may identify the message as belonging to a subject

class it is listening for, then use the VDE node to open the container and view or otherwise use the item in accordance with that item's associated rules and usage consequences.

In another arrangement, the subject switch 2051 may be located
5 on each customer appliance 2052(1)-2052(n). Using messaging services 2058, each subject switch 2051 may communicate with the matching and classification utility 900 to locate sources of content matching the subscribed classes. In this example, the subject switch 2051 on the local appliance then uses the messaging services 2058 to
10 communicate with one or more content sources 2074 indicating classes of content to which it wishes to subscribe. Using the messaging services 2058, one or more content sources 2074 may directly send and/or broadcast items in the desired classes to subscribing customers 2053 in VDE secure containers along with
15 associated rules and consequences. In another arrangement, the content source 2074 may send one set of rules and usage consequences that apply to members of one or more item classes, thus potentially improving the efficiency of distribution and of rights management. In another example, the rules and content items may be
20 sent in separate VDE containers. In this example, the messaging services 2058 and subject switch 2051 listen for messages that are addressed to those customers who subscribe to a particular content item class and makes those items available to customers using an application 2060.

In another arrangement, messaging services 2058 and/or subject switch 2051 may be installed and run on network routers, network switches, one non-limiting example being ATM switches, and other packet and/or cell switches.

5 Example: Digital Broadcasting Based On Matching And Classification

“Shear et al” discloses a Digital Broadcasting Network (“DBN”) that may function as a cooperative of Web sites and, for example, service providers, with a central and perhaps regional and
10 logical (e.g., market based) headquarters groups, or it may function as a for profit, shareholder corporation in a business model reminiscent of television broadcast companies (e.g., NBC), or it may function as a cooperative or virtual corporation that has some mix or combination of mixes of the above attributes and employ distributed peer to peer,
15 hierarchical, and centralized administrative business relationships and activities.

In one example, plural corporations may join together to provide the advantages of size and coordination with individual participants providing some degree of specialty expertise and the
20 body of entities coordinating together in some fashion in a “higher” level cooperative or corporation.

Figure 48 shows one non-limiting example 2100 of a DBN that includes one or more DBN Web servers 2104(1)-2104(n) and one or more Web users each with VDE nodes. Users are attracted to a

specific DBN server (or servers) because it provides access to specialized content and/or services 2108. Based at least in part on rights management information 2110 collected from DBN servers, for example, controls associated with the most frequently requested

5 information, the matching and classification utility 900 creates categories of content (and/or services) and assigns DBN servers to one or more classes according to their specialization(s). The matching and classification utility 900 may may create at least one class hierarchy, class, classification scheme, category and/or category

10 scheme using at least some rights management information and assign at least DBN server and/or at least some information to at least one category and/or class.

For example, one DBN server may specialize in consumer sports information while another may specialize in legal information.

15 DBN servers may specialize in plural content (and/or service) areas. This class and class assignment information is provided to DBN servers, to content (and/or service) providers, or both.

The matching and classification utility 900 in one example sends VDE containers 2112 to content sources 2102 indicating

20 specific classes of content that should be sent to one or more DBN servers 2104. Using this information, content providers 2102(1)-2102(n) then send content in these categories in VDE containers 2106 that match the categories of most frequently hit and/or consumed content on a DBN server 2104(1)-2104(n). (In another example,

other information may be used as the basis of classification, matching, and selection.) For instance, the matching and classification utility 900 sends a VDE container 2112(2) to content source 2102(1) with instructions to send content in categories 1,11, and 15 to DBN server 1 (2104(1)). This content may, in turn, be sent to one or more consumers in VDE containers 2108(1), 2108(3).

In one aspect, this example process is analogous to hard goods manufacturers and distributors keeping Wal-Mart shelves stocked with those items in greatest demand based on point of sales and inventory data. One difference, of course, is that in this example, the DBN server is stocked with intangibles in the same or similar class as the intangibles sold rather than providing replacements for hard goods that have been sold off the shelf. In another example, a DBN server may send its classification data to content providers along with a request that they send more of the same. The request may be sent independently of the class information.

In another example, the matching and classification utility 900 may receive content and/or rights management information from providers and go on to create classes of content and/or content providers in which the classes may be partly defined using rights management data. Content on one class may, among other things, be distinguished from content in another class by price, payment methods, usage opportunities (e.g., available for printing, available for viewing pay-per-use), usage consequences, and/or specific

permissions. The matching and classification utility 900 may subsequently send a communication, perhaps in a VDE container, to providers indicating that they send content in one or more specified classes to at least one DBN server.

5 Non-limiting example Figure 48 shows that the DBN 2100 may consist of video 2202 and/or audio 2203 content providers who send certain categories of video and/or audio content 2206 to DBN servers 2204(1)-2204(n) based on the categories of content each server may specialize in, which, in turn, may be determined at least in part on
10 frequency of usage and/or other rights management information sent in VDE containers 2213 to the matching and classification utility 900, or to a usage clearinghouse 300 and then to a matching and classification utility 900. (In another example, other information may be used as the basis of classification, matching, and selection.) The
15 matching and classification utility 900 sends VDE containers 2212 to content sources indicating that they should send content in specific categories 2206 to specific DBN servers 2204. In turn, each DBN server 2204(1)-2204(n) delivers video 2208 and/or audio 2209 in VDE containers to parties interested in such content. In another
20 example, a VDE container may hold both video and audio and/or any other content type.

**Example: Matching and Classification Utility 900
Can Also Support "Pull" Distribution Models Based
On Classes**

Notwithstanding the noted trend toward "push" content
5 delivery models, the present inventions also enhance the efficiency,
focus, specificity, and convenience of content "pull" models. In one
example 2300 (Figure 49), the matching and classification utility 900
sends in VDE containers 2306(1)-2306(n) at least one administrative
event and/or associated method that performs classification and/or
10 class assignments to a VDE-aware appliance. The administrative
events and method(s) are processed under the control of the VDE
node. In one example, the results of processing the classification
method may indicate at least one class of content and/or services of
interest to a user and/or node. The classification method may also
15 create at least one class hierarchy, class, classification scheme,
category and/or category scheme using at least some rights
management information and assign at least one service and/or at
least some content to at least one category and/or class.

Subsequently, a VDE container 2308 may be sent to a provider
20 2302 with information indicating at least one class of content,
services, transactions, rules and/or usage consequences, such as the
ability to modify, excerpt and/or reformat, and/or events and a request
that that the provider send content and/or pointers to services that
meets the stated criteria and/or descriptive information about such
25 content, services, transactions, and/or events to the requesting user

and/or node. The request may, for example, be initiated explicitly by the user and/or node or may be initiated by the node according to one or more administrative events and associated methods and/or control sets. In turn, the content provider 2302 sends a VDE container 2304
5 to the requesting user 2306(1) with content that matches the desired selection criteria and/or profile.

The user may elect to use, consume, purchase, and/or rent one or more content objects (or use one or more services). As this one example shows, the user pulls in content and/or interacts with services
10 by matching at least one class indicating user preferences with at least one class of content objects and/or services and/or transaction types.

Example: The Enterprise Distributed Matching And Classification Utility

Businesses and other organizations may be concerned with
15 privacy and confidentiality regarding information and/or services used within the company. This concern may be manifest regardless of whether the information and/or services originated inside and/or outside the organization. Thus some organizations may have strong incentives to take advantage of the present inventions by operating a
20 distributed matching and classification utility 900 to provide matching and classification services within the enterprise while at the same time maintaining a higher degree of confidentiality and privacy by selecting and/or limiting the nature, range, and detail of information sent outside the organization.

Figure 50 shows an example 2400 of an entity 2406 that has one or more VDE enabled appliances and users 2420(1)-2420(5) on a corporate Intranet 2418. These appliances may be, for example, computers, workstations, mainframes, or more specialized devices, such as supercomputers and/or graphics workstations for animation and special effects. The company may also operate internally one or more Commerce Utility Systems, perhaps including a financial clearinghouse 200, a usage clearinghouse 300, and a matching and classification utility 900. The company may also operate at least one content server 2414. These commerce utility systems and servers are also connected to the company Intranet 2418. The company 2406 also maintains one or more connects to the Internet 2402. (In another example the company may maintain connections to at least one private network operated by themselves and/or another party in addition to, or instead of one or more connections to the public Internet.) The content server(s) may provide access to internal, proprietary company information and/or to external, often commercial information. The internal content server may act as a gateway to external providers 2404(A)-2404(C) and/or may host commercial content locally on a content server 2408.

In one example, VDE audit records and/or other rights management information are sent in VDE containers 2412 from one or more VDE nodes 2420 to the enterprise usage clearinghouse 300 which may forward at least some of this usage information in VDE containers 2410 to the enterprise matching and classification utility

900. The enterprise matching and classification utility 900 may also collect from internal information sources 2414 information in addition to audit and rights management information, such as information in a human resources, accounting, and/or budgeting
5 database containing data about company employees. These data may indicate, in one example, titles and responsibilities within the company, budgets allocated for external information and/or services, authority to spend, and budget remaining. The budget and financial information may have come in part from the financial clearinghouse
10 200. The matching and classification utility 900 may also create at least one class hierarchy, class, classification scheme, category and/or category scheme using at least some rights management information and assign at least service and/or at least some content to at least one category and/or class.

15 In one example, using at least some VDE rights management data, for example, whether certain information can be viewed by anyone, by any employee, or only by employees in certain job classes, such as "manager," the enterprise matching and classification utility 900 creates one or more categories and assigns one or more
20 employees and/or VDE nodes to one or more topic categories. These categories may, for example, indicate content and/or service topics, subjects, and/or content areas of potential interest to each employee and/or groups of employees sharing at least one attribute in common, for example, similar interests and/or responsibilities.

In turn, the enterprise matching and classification utility 900 sends to at least one external content and/or service provider 2404 on Internet 2402 one or more VDE containers 2424 with information that indicates categories of interest. The content providers 2404 may themselves be specialized; in one example, a content provider may specialize in general business and financial news while another may specialize in scientific, medical, and/or technical information. In another example, a single content and/or service provider may provide an extremely broad range of content and/or services.

10 The external provider may send at least one VDE container 2422(1) with content and/or rules and consequences and/or metadata about content and/or services to a content server internal to the enterprise. In another example, such VDE container(s) 2422(2) may be sent directly to an employee and/or one or more groups of employees. The information sent by the external provider is tailored to, or in some way responsive to the content and/or service categories requested by the enterprise matching and classification utility 900.

20 In another example, the enterprise matching and classification utility 900 itself may be a distributed commerce utility implemented on more than one computer and/or other appliance within the enterprise. These several matching and classification utility 900s may serve different geographic areas and/or may themselves specialize in particular content and/or service areas.

In another example, the enterprise matching and classification utility 900 send class and/or class assignment information to a matching and classification utility 900 in another organization that, in turn, may be part of a common value chain.

5 **Example: Chain of Handling and Control Entails Class-based Rules and Usage Consequences**

VDE-based value chain management or "chain of handling and control" disclosed in "Ginter et al" enables, amongst other things, plural parties to independently contribute rules and usage
10 consequences under the authority and/or control of more senior or prior participants in the value or distribution chain. Class-based rules may play a role in the efficiency and effectiveness of creating, operating, and/or extending value chain processes.

Figure 51A shows an example 2500 of a publisher ABC 2502
15 using a VDE packaging application 2510 to put into a VDE secure container 2512 sets of rules and usage consequences that may vary according to class. In this non-limiting example, the class is "content type." The publisher may have rights in a wide variety of content and content types. Consequently, the publisher may create rules for text
20 objects that may differ from rules for audio objects.

The publisher 2502 sends the class-based rules and usage consequences to a first creator 2504 who also has installed VDE on her or his appliance 2516 and who has also been given one or more certificates and/or other digital credentials by the publisher (and/or

trusted third party) indicating that he is indeed a creator authorized by the publisher 2502. The publisher has included rules that allow only authorized value chain participants to package content using publisher provided rules and/or to modify, enhance, extent, and/or change some
5 or all of the publisher's rules.

The first creator 2504 then uses a VDE packaging application 2510 to package an image he has created in a VDE container 2514 according to the rules provided by the publisher and with the addition of the creator's own rules. In one example, the first creator
10 contributes rules that implement a one-time 50 cent charge to the consumer for opening and viewing the creator's image. The creator may also contribute rules reflecting his wish to receive audit records with information concerning the consumer and/or context in which the image was used. These creator rules and usage consequences are
15 contributed generally independently of the rules and usage consequences contributed by the publisher. Note that the VDE container 2514 now holds at least the publisher's 2502 rules for each object class, the first creator's image and his associated rules and usage consequences.

20 A second creator 2506 receives the VDE container from the first creator and using a VDE packaging application 2516 adds a text file to the container 2520 along with her rules and usage consequences. As before, she also has a certificate and/or other digital credential(s) identifying her as authorized by publisher ABC to

add and/or modify content and rules and usage consequences. As in the case of the first creator 2504, she adds her text and rules and usage consequences generally independently of controls contributed by prior participants. She may, in one example, prevent printing of
5 the text and charge \$1.00 the first time a consumer opens and views the text.

The VDE container 2508 now holds text and rules and usage consequences contributed by creator 2 (2506), an image and rules and usage consequences contributed by creator 1 (2504), and the class
10 based rules (and perhaps other rules as well) contributed by example publisher ABC 2502.

Creator 2 (2506) sends the VDE container 2508 to publisher ABC 2502 who then sends the container 2522 directly and/or indirectly to consumers. When the consumer uses the content, the
15 rules and usage consequences of all three value chain participants (and of other possible participants as well, distributors and repackagers, for example) are applied.

Example 2600, Figure 51B shows that the publisher 2602 may have sent a VDE container 2612 with various rules and usage
20 consequences to a matching and classification authority 900 who may classify the rules and send the rules and their class assignments to a rights and permissions clearinghouse 400. The matching and classification utility 900 may also create at least one class hierarchy, class, classification scheme, category and/or category scheme using at

least some rights management information and assign at least one rule to at least one category and/or class.

An authorized first creator 2604 may send a VDE container 2617 to the rights and permissions clearinghouse 400 asking for rules in the class "rules for authorized creators, for image objects, from publisher ABC." The rights and permissions clearinghouse 400 returns a VDE container 2614 with rules in the requested class. The first creator 2604 uses a packaging application 2616 to package his image using these rules plus rules and usage consequences reflecting his rights and wishes and sends the VDE container 2614 to the second creator 2606.

The second creator 2606 also sends a VDE container 2619 to the rights and permissions clearinghouse 400 asking for rules and consequences in the class "rules for authorized creators, for text objects, from publisher ABC." The rights and permissions clearinghouse 400 returns a VDE container 2621 with rules and consequences in the desired class. The second creator 2606 uses a packaging application 2618 that determines that she is a creator authorized by publisher ABC 2602 and goes ahead and adds her text object and her rules and consequences to the VDE container 2608, which is then sent to the publisher ABC 2602 for further augmentation, vending, and/or distribution to other value chain participants.

Example: Secure Directory Services May Provide Class And Class Assignment Information

Whole industries have arisen to target communications to individuals, organizations, groups, and/or other classes sharing at least one common attribute, and/or to provide directories from which others can locate individuals, organizations, groups, and/or other classes. Examples of these industries include direct marketing, advertising, yellow and white pages directories, directories of directories, and various electronic and paper membership lists and professional directories.

In addition to identifying information such as names, e-mail addresses, physical mailing addresses, phone numbers, fax numbers, and/or similar attributes, the secure directory services 600 may also provide information about class membership(s) for individuals, devices, services, groups, and/or organizations. The non-limiting example 2700 shown in Figure 52 includes a secure directory service 600 that has received class and class assignment information for one or more individuals 2716(1)-2716(n). The class assignment information is shown in the bottom four rows of the directory record 2718(1) for one individual.

In this example, a content provider 2702 sends a VDE container 2704 to a secure directory services 600 asking whether the service can provide a list of individuals in class "AF." The requested class could be any class defined by one or more attributes and may be based on usage profiles that include rights management information,

non-exhaustive examples of which include price, payment methods accepted, permitted operations, meters, and privacy controls.

The secure directory services 600 returns to the content provider in a VDE container 2706 an indication that there are
5 presently 57 individuals known to that service in class "AF." In turn, the content provider 2702 sends a VDE container 2708 with at least one piece of content and/or rules and usage consequences back to the secure directory services 600 along with instructions requesting that
10 the secure directory services 600 forward the content and/or control sets to each of the 57 members of class "AF" who might be interested in this piece of content. The secure directory services 600, in turn, forwards the content and/or controls (in VDE containers 2714(1)-2714(n)) to members of class "AF," who may elect to interact with the content in accordance with their associated rules and consequences.

15 In another example, the secure directory service 600 may send identifying information 2710 directly to the content provider 2702 who may then send content 2712 in one or more classes directly to one or more members 2716(1)-2716(n) of the class. The secure directory services 600 may, for example, include permissions for the
20 class information that have expiration dates and/or limits on the number of times the information can be used.

**Example: Matching And Classification Utility 900
Supports Class-Based Micro-Merchandising And
Micro-Segmented Sales Processes**

The present inventions may be used in support of services as
5 well as content distribution based business. Example 2800 (Figure
53) shows a travel company 2801 sending a VDE container 2810 to a
matching and classification utility 900 requesting information on
those individuals who may be interested in certain combinations of
leisure-time activities. These classes might have been defined at least
10 in part on the basis of usage and other rights management information
2816, for example, the kind of leisure-time information recently
looked at, for how long, and/or its cost, and/or the kind of Web sites
recently frequented, sent from consumer VDE nodes 2802(1)-2802(n)
to the matching and classification utility 900, and/or to a usage
15 clearinghouse 300 who, in turn, sends at least some of the usage
information (or a summary form of such information) to the matching
and classification authority 900. Classes may also be defined using
information gathered directly from the consumer 2818, perhaps under
the control of VDE. The matching and classification utility 900 may
20 also create at least one class hierarchy, class, classification scheme,
category and/or category scheme using at least some rights
management information and assign at least one consumer, service,
and/or at least some information to at least one category and/or class.

Example Figure 53 shows that a consumer 2802(1) has recently
25 indicated a preference and/or interest in skiing, music, and flying to

Colorado. Another consumer 2802(n) has indicated a preference for and/or interest in surfing Hawaii. These preferences may be determined at least in part on the basis of rights management information. In response queries sent in one or more VDE containers

5 2810 from the travel company asking for interest and preference information, the matching and classification utility 900 returns one or more VDE containers 2812 with identifying and class information. The travel company may send information about already existing vacation packages and/or packages specially created to meet the

10 specific interests of one or more individuals, for example, information about skiing in Colorado, and rock concerts 2604 to consumer 2802(1) and information 2614 about surfing Hawaii to consumer 2802(n). The recipients may send VDE containers 2806 to the travel company 2801 indicating agreement to buy the package offered or

15 may request additional information or may negotiate terms and conditions such as price, departure date, insurance, and the like. These negotiations may be conducted using the inventions described in "Ginter et al", Figures 75A-76B using VDE negotiations.

Both services and/or hard goods may be offered to particular

20 persons, nodes, groups, and/or entities based on the class membership of the potential purchaser and the class membership of the goods and/or services to be purchased. Thus in another example, the travel company could have included the purchase and/or rental of the skis or of the surf board.

**Example: Matching And Classification Utility 900
Supports Trading in Hard Goods**

Business to business trading in goods and/or services may be substantially facilitated through services provided by the matching and classification utility 900. Information on certain classes of goods and services may be delivered to certain people, groups, or entities based on the class membership of the recipient. In one example, these various class memberships may be determined using control set and audit information regarding trading preferences and/or transaction patterns. In another example class membership may be determined by actions and/or information provided by at least one value chain participant.

Example 2900 (Figure 54) shows a buyer A 2904 sending a VDE container 2908 to a trading company 2902 with a request asking if trading company will sell company A one or more desired items. Trading company 2902 may then send a VDE container 2910 to a matching and classification utility 900 with a query asking who can supply the desired items under terms and conditions that are also included in the container. Since these terms and conditions may be the subject of negotiations, they may be in a format conducive to VDE-based negotiations as described in "Ginter et al" Figures 75A-76B.

The matching and classification utility 900 may send inquiries 2910 to one or more suppliers 2906(A)-2906(N) and/or may have already received information and/or associated control sets from

suppliers in VDE containers 2912. Based on the request from trading company 2902 and supplier 2906 information obtained 2912, the matching and classification authority 900 returns a VDE container 2916 indicating that in this one example, suppliers A 2906(A) and Z 5 2906 (N) can provide goods in the class(es) defined by trading company's 2902 request(s) 2910. In turn, trading company 2902 sends at least one VDE container 2918 to buyer A 2904 indicating that they will sell buyer A the previously requested items under the enclosed terms and conditions. In another example, there may be 10 some VDE-based (see "Ginter et al", Figures 75A-76B) negotiations between the various parties in this value chain, including between trading company 2902 and buyer A 2904.

In another example, buyer A 2904 may consult the matching and classification authority 900 directly and may then purchase 15 directly from one or more suppliers 2906.

Example: Matching And Classification Utility 900 Supports Securities Trading/Brokering

In addition to hard goods, the matching and classification authority 900 may also support securities trading. Example 3000, 20 Figure 55, shows the matching and classification authority 900 sending to a VDE-aware appliance with one or more stock trading related applications 3004 a VDE container 3010 with an administrative event and method (as described in "Ginter et al") for classifying equities related information, including, as non-limiting 25 examples, current and historical price, volume, and index

information, financial performance data for publicly held companies, forecasts, risk management information, options and futures, and the like. The classification method may also utilize rights and permissions, including access control information, permitted
5 operations, and/or expiration times and/or dates for rights management information. The classification method may also create at least one class hierarchy, class, classification scheme, category and/or category scheme using at least some rights management information and assign at least one element to at least one category
10 and/or class.

In turn, using the VDE aware appliance 3004, the stock trader 3006 sends a smart object 3012 to at least one information source 3002 asking for information in at least one class identified by the classification method. In one example, the class may be information
15 concerning "publicly traded companies with annual revenue greater than \$500M in the healthcare sector in which the CEO has been in place less than 5 and greater than 1 year and with access restricted to customers (rather than available to anyone) with access and use expiring in 90 days." The information provider(s) 3002 returns a
20 VDE container 3014 with information meeting and/or more closely meeting the stated class criteria. Based upon this and other information, the trader 3006 may go ahead and enter an order for at least one trade in at least one stock 3008. In another example, the trader may create or obtain methods that trade automatically in certain
25 classes of securities.

**Example: Matching And Classification Utility 900
Supports Trading in Currency and Debt Instruments**

Among the classes of great value to traders are the classes of items whose trading maximize profits and/or minimize losses.

- 5 Example 3100, Figure 56, shows a trader in currency and/or debt instruments 3102 sending a VDE container with market and other financial and economic information and VDE control set information 3108 to a matching and classification authority 900 with a query 3114 asking the matching and classification authority 900 to identify the
- 10 class of currency trades and/or debt instrument trades that maximizes profit and/or minimizes losses. The matching and classification authority 900 applies one or more methods to the data and returns at least one class definition 3112, the assignment of possible trades to that class 3110, and relevant control set information, such as controls
- 15 indicating who may see the information, and those that prevent unauthorized modification of the information. The matching and classification authority 900 may also return methods for executing the trade. The matching and classification utility 900 may also create at least one class hierarchy, class, classification scheme, category and/or
- 20 category scheme using at least some rights management information and assign at least some trading information to at least one category and/or class.

- The example trader 3102 examines the recommendation and sends VDE containers 3118 (A, B) with trade methods and control
- 25 sets to a foreign exchange market 3104 and/or to a debt instrument

market 3106 where the trades are consummated. The markets send back VDE containers 3116(A, B) with audit information indicating the results of the trading order. In another example, the matching and classification authority 900 may be instructed to send trading orders
5 directly to the market(s) for execution. In another example the trader may send a VDE container to at least one source of relevant information asking that source to send certain information to the matching and classification authority 900. In another example, having established the desired trade(s) using the matching and
10 classification authority 900, the trader may place the trade by phone and/or computer and/or other communications device without using VDE.

**Example: Matching And Classification Utility 900 Supports Consumers Locating Services That Are
15 Members Of A Specified Class**

The services of the matching and classification authority 900 may also benefit consumers by locating certain classes of services. Example 3200, Figure 57, shows a consumer sending a VDE container 3206 to a matching and classification authority 900 asking,
20 "which banks are in class A?," where class A are "those banks that offer the highest savings interest, no ATM fees, online/Web banking using VDE, insured accounts, free checking with balances larger than \$2,500, "image" statements (where check images rather than the actual checks are returned), and complete privacy protection (except

where legally required to disclose) for VDE based banking transactions.

The example matching and classification authority 900 sends a query in a VDE container 3208 to one (or more) information sources 5 3202 and receives one or more VDE containers 3210 with the requested information. The matching and classification authority 900 then determines which bank or banks meet the stated criteria of the consumer 3204 and then sends a VDE container 3212 with the answer to the consumer, in this example, banks A, B, and C. The consumer 10 3204 may then go ahead and execute a financial transaction, for example, transferring funds from one bank to a bank identified by the matching and classification utility 900 as offering higher interest rates, while being assured of maximal privacy for this (and perhaps other) transactions.

15 In another example, after determining which banks are in the desired class, the matching and classification authority 900 may send a VDE container to one or more banks saying that the consumer wishes to know about their services and requesting the bank to contact the consumer directly. The bank may send controls ensuring 20 the privacy of future interactions with the customer. For example, controls that apply to audit records such that only the bank and the consumer will have permission to access these records.

**Example: Matching And Classification Authority 900
Supports Class-Based Software Distribution**

VDE and the inventions disclosed in "Ginter et al" at last provide a way of ensuring that the efforts expended on creating software will be rewarded since the software can now be persistently protected, usage information can be collected, and payment ensured. These inventions also support micropayments and microtransactions, thus creating a world in which the price of software objects—any kind of objects actually—may become very small. Pay per use, rental, rent to own, and other pay as you go pricing models together with VDE may create a new explosion of creativity in software design and creation, since use prices will be low and providers can be assured of receiving payment.

The present inventions provide opportunities for software providers to more efficiently market their wares. Example 3300, Figure 58, shows a number of users with VDE installed on their appliances 3304(A-F). These people are using software (and other content). VDE meters usage of various objects and sends audit records in VDE containers 3306 (A-F) to a usage clearinghouse 300, which then sends audit records 3308 to the matching and classification authority 900. A software distributor 3302 sends a VDE container 3310 to the matching and classification authority 900 with a query asking who is in the class, "buys Java applets, with pay per use pricing, and for which the cost per use is between \$.0001 and \$.001?"

The matching and classification authority 900 returns a VDE container 3312 with a list of names and (network) addresses of those matching, or most nearly matching the desired characteristic(s). The software distributor 3302 then sends at least one VDE container 3314
5 with at least one software object, and/or a pointer to a software object, in this case a Java applet, and perhaps other relevant information, such as VDE control sets and/or various metadata describing some aspect of the object, for example, what it does, what it costs, etc. The user may then elect to use the object or not. In another example,
10 instead of individuals or VDE nodes, the users might be groups of nodes, users, organizations, parts of an organization, and others that can be identified as belonging to at least one class. In this case, the software may be offered to some or all members of class, group and/or organization.

15 **Example: Matching & Classification Utilities Provide Services To Authenticated Classes of Nodes, Users, Content Services and/or Transaction Services**

Among the ways in VDE nodes, users, content services, and/or transaction services can be authenticated is through the use of
20 certificates and/or other digital credentials issued by an appropriate trusted third party, a certifying authority 500, for instance, that warrants and/or attests to some fact or facts, which may include membership in one or more classes, including the identity class. Figure 59 shows a non-limiting example 3400 in which a number of
25 matching and classification authority 900(1-N)s, each of which may

provide its services to different classes, where class membership is authenticated using certificates and/or other digital credentials. In other examples, additional authentication mechanisms may be used in combination with, or instead of certificates, such as information
5 known only to the user, VDE node, and/or appliance, including passwords, cryptographic keys, information stored in hardware, and/or software.

In example 3400, Figure 59, commerce participants including, the matching and classification authority 900, may make rules and
10 consequences conditional on class definitions and/or the assignment of members to a class. Class membership may be authenticated by a certificate and/or other digital credential issued by one or more commerce participants in addition to, and/or instead of a trusted third party such as a certifying authority 500. For example, a certificate
15 and/or other digital credential may attest to user identity, that is, that a user is the user he or she claims to be. Nodes, devices, networks, servers, clients, and services, are other non-limiting examples of other commerce elements that may be authenticated with certificates and/or other digital credentials. Any commerce participant may issue a
20 certificate, but other participants are not required to accept a given certificate as an authenticator.

Figure 59 shows multiple matching and classification authorities 900(1)-900(N), each of which may provide services to members of a particular class, in these non-limiting examples, to

nodes in a particular deployment (matching and classification authority 900(1)), in a particular vertical segment and/or institution of society, such as Higher Education (matching and classification authority 900(2)), one or more value chains, such as business information content providers (matching and classification authority 900(3)), and/or a particular transaction and/or service arena, such as hard goods trading (matching and classification authority 900(n)). Other commerce utility systems, a certifying authority 500 shown in Figure 59, for instance, may also provide services to a class. In each of these instances, the services of the matching and classification authority 900 may depend upon finding certain authenticating certificate(s) and/or other digital credentials on the appropriate VDE nodes.

For example, matching and classification utility 900(1) provides services to nodes 3410(1-n) in the deployment 3402 administered by VDE administrator 800. Each node may have a certificate 3412 issued by certifying authority 500(1) that provides services to this deployment.

In another example, certifying authority 500(2) provides certificates and/or other digital credentials to participants in a higher education value chain 3404 consisting of an arbitrary number of colleges and universities 3416(1)-3416(n), providers 3418(1) and students 3418(n), and a matching and classification utility 900(2) that provides classification, matching, and selection services to higher

education 3404. In one example, the matching and classification utility 900(2) only provides services to value chain participants who have a certificate 3420 issued by certifying authority 500(2).

Matching and classification utility 900(3) services can be
5 provided only to members of one or more classes based on certificates issued by a certifying authority 500(3). In one example, the class is participants in a business information value chain 3406, comprising an arbitrary number of content providers 3424(1)-3424(n), an arbitrary number of users and/or consumers of business information
10 3422(1)-3422(n), and a certifying authority 500(3) that issues certificates and/or other digital credentials to members of the value chain 3406.

In addition to membership in certain deployment, institutional, and/or content usage classes, the matching and classification authority
15 900(4) may provide services to members of a certain transactional value chain, in one example, traditional transactions 3408. In this example, a certifying authority 500(4) issues certificates 3432 to one or more companies 3428(1)-3428(n) and one or more trading companies 3430(1)-3430(n). In another example, other participants
20 may receive certificates and/or other digital credentials, including banks and financial institutions, government authorities, for example, tax and/or customs authorities, consumers, suppliers, and/or transportation companies. The matching and classification utility 900(4) provides services only to those entities and/or individuals in

possession of the appropriate certificate 3432 indicating that the holder of the certificate is an authenticated participant in one or another trading value chains.

In other examples, a commerce utility system may provide
5 services to more than one class where class membership is indicated
by at least one certificate and/or other digital credential issued by a
certifying authority 500 and/or value chain participant. In one
example, matching and classification authority 900 might provide
services to the class "Higher Education" and to the class "K-12
10 Education."

Possession of a certificate and/or other digital credential may
be among the information used to classify a node, user, appliance,
device, entity, and/or other commerce participant, and rules and
consequences can be made conditional on membership in one or more
15 authenticated classes and/or on the degree of confidence the rule
provider has in the trustedness of the certificate and/or other digital
credential issuer. In one example, a discount to higher education may
be larger if the root for chain of trust for a given certificate is a well-
known, highly respected and trusted third party, such as an
20 authoritative accrediting organization, and smaller if the root belongs
to the MIS department of a small college. In this example, the
provider is willing to grant a higher discount when there is higher
certainty that the recipient is in fact a member of a specific class or
classes.

Example: Matching And Classification Authority 900 Supports Control Sets Based In Part On Employee Classes, Content Classes, And/Or Certificates And/or Other Digital Credentials

5 Chain of handling and control enables, amongst other things, multiple organizations to work together in secure, trusted, efficient, cooperative commerce processes. One way in which the present inventions extend these ideas is through control sets with rules and usage consequences that may be based in part on classes and the
10 assignment of persons, entities, devices, content, services, or other process elements to classes of one kind or another by the matching and classification authority 900.

One example technique to classify employees is at least in part according to their roles and responsibilities within an organization.
15 The matching and classification utility 900 supports classification, matching, creation and/or modification of VDE control set(s) based at least in part the class assignment of individual and/or groups of employees. In part by virtue of their employee classification, at least one employee may receive certain rights management information,
20 for example, permission to access certain classes of information or permission to perform one or more permitted operations, transactions and/or events.

Example 3500, Figures 60A-60C shows a nurse 3504(1), physician 3504(2) , and billing clerk 3504(3) all work directly for an
25 example hospital. The present inventions are in no way limited to

hospitals, but apply to any organization, group, entity, and/or institution with at least some defined roles and responsibilities and/or other class definitions that apply to employees, members, and/or others associated, affiliated, and/or employed by the organization, group, entity and/or institution. Rights management information may be part of the claim definition, for example, permissions to view, modify, excerpt, and so on.

Control sets may provide permissions conditional on employee class, for example, certain classes of employees may modify certain information and/or classes of information in a database while others may not. Class membership may be indicated by digital credentials, non-limiting examples of which include digital certificates and digital membership cards. Controls may be conditional on other information as well, for example, some computers and/or display devices may not show certain classes of data or updates to certain data elements may not be performed from certain computers or display devices.

Another example role is a representative 3504(4) of an insurance company 3508, who may have access to certain classes of hospital information by virtue of her or his class membership(s), some of which may derive from her or his role in the insurance company 3508 and/or from the insurance company's relationship with the hospital and/or with some of the hospital's patients and/or staff. The present inventions are not limited in application to an insurance company, but may be applied to any individual, group, organization,

entity, and/or institution with whom the example hospital and/or other entity has some form of relationship.

An example insurance company 3508 have received a certificate in a VDE container 3534 issued by certifying authority 500(1) attesting to the identity of the insurance company. In another example, this certificate and/or one or more additional certificates may attest to the fact that the insurance company has the appropriate charter, licenses, and other grants of authority to be in the health insurance business. The certifying authority 500(1) may also send a certificate in a VDE container 3532 attesting to hospital's identity. In another example, this certificate and/or one or more additional certificates may attest to the fact that the hospital has the appropriate charter, licenses, and other grants of authority to provide hospital and related services.

The insurance company 3508 may have sent one or more control sets to the hospital in a VDE container 3542. These controls may be based in part on one or more certificates 3530 and/or on the classification output of an example matching and classification utility 900(2) operating within and/or on behalf of the insurance company 3508. The controls in container 3542 may indicate which individuals are actually employees of the insurance company, employee membership in one or more classes, permissions associated with that individual and/or class, and/or permissions associated with specific devices, communications channels (devices, ports, etc.), and/or

processes. In this one example, the hospital matching and classification utility 900(1) may create controls using the same and/or additional classes and controls received from the insurance company 3508.

5 The insurance company 3508 may also provide one or more certificates to the hospital attesting to the fact that one or more information sources within the insurance company are to be taken by the hospital as trusted sources. Lastly, in this regard, the insurance company may issue one or more certificates on behalf of each
10 employee attesting that each is in fact an employee of the company and may have certain authorizations.

In example 3500, Figures 60A-60C, a matching and classification utility 900(1) has identified various classes of hospital employees using information from at least one hospital information
15 system 3502 and/or VDE node. The matching and classification utility 900(1) may also make use of certificates issued by a certifying authority 500(1) outside (a trusted third party) and/or a certifying authority 500(2) inside the hospital. Using data dictionaries 3522, patient records 3520, various employee information 3524, automated
20 procedures, and/or other means, the matching and classification utility 900(1) creates classes 3526 of patient record information and associates one or more control sets 3528 with each class of information and/or with a patient record as a whole. These control sets may specify who has permission to use and/or modify the record

and/or an element(s) of the record that has been assigned to one or more classes on which the control set(s) may in part depend. In one example, the class based controls 3528 may be combined with other hospital and/or other party controls, controls from the insurance
5 company 3508, to create new controls 3510(1)-3510(n) associated with patient records 3512(1)-3512(n).

The example nurse 3504(1) and physician 3504(2), for example, may be able to view, modify, print, and/or copy patient's name, address, and other similar descriptive information, next of kin,
10 insurance, and medical information in accordance with controls 3510(1) and 3510(2), respectively . In another example, some members of the class "nurse" and/or the class "physician" may have different permissions by virtue of membership in one or more additional classes. A physician who is in the class "hospital
15 administration" may have different permissions, for example, to billing records.

A billing clerk 3504(3) in the hospital may not have permission in control set 3510(3) to view medical information and/or next of kin, and in this example may be restricted to name and other patient
20 descriptive information, insurance information, and billing information from the patient record. A representative 3504(n) of the insurance company may have permission by virtue of control set 3510(n) to view, but no permission to modify, print, or copy patient record 3512(n). In each of these examples, the VDE control sets are

at least partially conditional on the presence and/or absence of certain certificates indicating membership in one or more classes.

The present inventions may be applied to any information, person, group, device, network, service, database that pertains to any
5 commerce activity whatsoever, and regardless of whether the parties to the commerce activity are individuals, groups, entities, organizations, institutions, nations, and/or societies.

**Example: Matching And Classification Authority 900 Supports Classes And Matching Based In Part On
10 Workflow And Work Process Automation**

Not only do the present inventions enhance commerce processes that principally entail information, but the present inventions enhance workflow and work process automation as well. Example 3600, Figure 61, shows PCs 3608(a-c) functioning as station
15 controllers connected to various manufacturing devices 3610 (a-c). These station controllers that exchange data and instructions with the equipment they control and/or manage. The station controllers are VDE-enabled. In another example, the manufacturing equipment may also have VDE nodes installed.

20 An example work in progress (WIP) and/or manufacturing control application 3606 keeps track of the overall manufacturing processes and exchanges information with other applications not shown, such as materials management, materials ordering, order

databases, logistics, inventory, accounts payable, accounts receivable, general ledger, human resources, time cards, and the like.

An example employee 3602 of the company sends a query 3612 in a VDE container 3604 to an enterprise matching and classification utility 900 within the company asking, "which VDE-controlled equipment will be available 3rd shift today, for 2 hours, capable of performing operations xyz with a nominal error rate of less than .0001 per cent?" The enterprise matching and classification utility 900 may request data 3616 from the WIP/manufacturing process control application 3606 and/or may already have access to the required data, indicating equipment availability, security level, capabilities, and statistical error rates. The WIP/manufacturing process control application 3606 may return a VDE container 3618 with the requested information. Based upon the query and available information, the matching and classification utility 900 responds by sending a VDE container 3620 to the employee 3602 with the answer, "equipment B and equipment C." In turn, the employee 3602 sends another VDE container 3622 to the WIP/manufacturing process control application 3606 with VDE a control set(s) indicating B and C should be scheduled for 2 hours on 3rd shift to do xyz operations. As part of this particular chain of handling and control, the WIP/manufacturing process control application 3606 sends VDE container 3624 to the VDE-enabled station controllers for equipment B or C with control sets that schedule work and specify the manufacturing processes and/or "recipes" for those specific

equipment 3610(b) or 3610(c). In turn, the respective station
controllers carry out their instructions and report progress and
completion in VDE containers 3626 sent back to the
WIP/manufacturing process control application 3606, which may in
5 one example, provide results to other applications and/or to the
employee who originally requested the work to be scheduled and
performed.

**Example: Matching And Classification Authority 900
Supports Classes And Matching Based In Part On
10 Government/Societal Commerce Administration**

Among the rightsholders in commerce processes of all kinds
are societies and governments. Governments may foster rules
indicating that certain classes of individuals may have not have access
to certain classes of content. Some classes of information may be
15 treated as members of classes that define permissions, such as
"confidential," "secret," "top secret," and so on. Other non-limiting
example governmental rights may address permissions for import,
use, and/or export of certain classes of hard goods, services, currency
and financial instruments, and content. Travelers entering the United
20 States, for example, are usually asked about currency (and currency
equivalents) being brought into the country by the traveler. Children,
for example, may be prohibited as a matter of law by governments
from viewing content in the class "sexually explicit."

Another example of government rights is that different tax rules
25 may be applied to different classes of electronic commerce

transactions using VDE. Example 3700, Figure 62A-62B, shows a certifying authority 500 operated by and/or on behalf of a government issuing a certificate and/or other digital credential indicating jurisdiction, namely, country. The certificate is sent in a VDE container 3710(a) to a VDE administrator 800. The government certifying authority 500 also sends certificates in VDE containers 3710(b)-3710(n) to the government matching and classification authority 900 attesting to the "country," in one example, the United States, and another certificate 3716 attesting to the fact that the matching and classification authority 900 is indeed an authorized service of the United States government.

In one example, the government matching and classification authority 900 has created tax class definitions 3712 and tax control sets 3714 that apply those definitions in various classes of circumstances, including the presence of certain control-related information, such as an appropriate country certificate from an authorized issuer of such jurisdictional certificates. The tax class definitions 3712, tax control sets 3714, and government authority certificates 3716' are sent in at least one VDE container to a rights and permissions clearinghouse 400, who, in one example, redistributes the tax class definitions 3712(1), tax class control sets 3714(1), and/or government authorization certificate 3716(1) to content providers 3702, service providers 3704, and other value chain participants. The certifying authority 500 also sends country certificates to one or more VDE administrators 800 who, in turn, send country certificates 3710'

to VDE nodes 3706(A)-3706(n) in their deployment. When content provider 3702 distributes content of any kind, the appropriate tax control sets 3714(A) are also included in the VDE container. A tax control set is applied whenever content is used in accordance with a tax class and provided that the appropriate jurisdictional certificate 5 3710' is present on the VDE node 3706(a). For instance, a VDE node may have a tax control set to be applied to sales of a class of content, specifically, to the class of "software." Whenever a software vend occurs, the appropriate tax is applied according to these rules.

10 In another example, the various country and government authority certificates may be sent directly from the certifying authority 500 to one or more VDE nodes 3706. The VDE controls that implement tax policy for one or more classes may also be sent directly to VDE nodes 3706 and/or to VDE administrators 800.

15 **Example: Classification May Be Used In
Automatically Selecting The Proper Display Context
Based On Classes Of Information**

Content objects may be displayed using one or another formats according to class membership of that object. In example 3800, 20 shown in Figure 63A, a matching and classification utility 900 provides content class information 3810 to information providers 3802. A consumer 3807(1) previously has sent a VDE container to a provider of sports information 3802(1) indicating interest in "class b" stories, and perhaps other classes as well. The sports information 25 provider 3802(1) sends back a VDE container 3808(1) with one or

more stories in "class b," perhaps "all stories about baseball, New York, Yankees, history, heroes with permission to print" an example of which is 3814(1), along with, in this example, one or more VDE control sets. The VDE container 3808(1) is received by a customer
5 3807(1) who then displays the content 3814(1) using one or another page formatting technologies based on macros, scripts, administrative events, methods, and/or other techniques. Also included in the VDE container is an image 3812(1) that was selected by the information provider as especially appropriate to the class of story being sent. In
10 this example, perhaps the image 3812(1) is a faint image of Joe DiMaggio. This image also meets the criteria of "permission to print."

Example 3800, Figure 63A, also shows another instance in which a different consumer 3807(n) previously has informed a nature
15 information provider 3802(n) of interest in class A stories. Here the information provider sends a VDE container 3808(n) that holds a class of stories different from the class of interest in the previous example. This VDE container 3808 holds a "class A" story, an example of which is 3814(n), that is displayed with a different image
20 3812(n), one that is appropriate to the story class, in this case, an image of a dog.

The class assigned to each story may be carried in the container as metadata for one or more story objects in another example. An example Web browser may request of the information provider an

image appropriate to that class, which if available, would be sent in another VDE container.

Class may affect display rules in other example ways as well. For instance, several team sports news stories may be displayed in a Web browser window in which a scene from a football or basketball game is faintly discernible in the background. Which image is displayed may be determined by the user's preferences given the classes of stories being presented on the page. The user, may have looked most at stories about the New England Patriots and a Patriots-related image may be displayed as background even stories about teams in addition to (or even instead of) the Patriots were being displayed.

In (another) example 3850, shown in Figure 63B, a matching and classification utility 900 provides class information to a provider 3852(1). Previously, one user 3857(1) has indicated to the provider 3852(1) that she prefers information in topic class A more than information in topic class C and information that costs less than \$.50 per article while the other user 3857(n) has the opposite preferences and is not price sensitive. A matching and classification utility 900 may provide classification information, class assignments for objects, administrative events, and/or methods for these and related purposes. Regardless, the information provider 3852(1) sends the identical VDE container 3858 to each of the users 3857. However, their browser and page formatting software 3856 produces different pages in

accordance with each user's topic class preferences. In the example first case, the user 3857(1) sees three columns of topic A and one column of topic C while the second example user 3857(n) sees three columns of topic C and one column of topic A. As this example
5 illustrates, the class preferences of users may affect the way in which the user interacts with content in various classes.

In another example, the matching and classification utility 900 may have sent one or more administrative events and/or methods 3859 to at least one user 3857 where the method performs the topic
10 classification on documents and/or establishes topic classes and/or topic classes of greatest interest to the user.

Example: Information May Be Classified With Respect To Difficulty -- And This May Pre-Determine An Appropriate Interface

15 The class of content and/or the class of user may determine at least one display characteristic. One interesting example way of classifying content is with respect to its difficulty. One example measure of difficulty is reading level, which may reflect such aspects as vocabulary and/or complexity. It is well known that children (and
20 adults) of the same approximate age read at different levels. In the example 3900, shown in Figure 64, a provider sends a VDE container 3902(1) with text at a 4th grade reading level and controls indicating that when used by a person reading at that level, the charge is 50 cents. However, if a person reads at less than the 4th grade level, the

charge is only 40 cents. "Reading level" may be indicated by a certificate and/or other digital credential.

A matching and classification utility 900 may send administrative events and/or classification methods 3910 to
5 information providers, one or more other value chain participants, or to the students appliances directly. These methods may, for example, classify documents according to the degree of difficulty and create or modify controls for the whole document and/or subparts of the document, controls that may indicate the different prices for users at
10 different reading levels. The matching and classification utility 900 may also send administrative events and methods to users that know how to make the document appear in the example browser at a lower reading level.

The example VDE container 3902(1) is sent from the provider
15 to a child 3906(1) in the 4th grade who is reading that at that level. When the child opens the container to view (or otherwise use) the text, she or he is charged 40 cents (which might be paid by a third party such as a school and/or parent. The child sees the text as written 3904(1)

20 Example 3900, Figure 64, also shows the exact same document being read by a student 3906(3) in the class of 2nd grade readers. Now the browser displays the document 3904(3) modified by methods that may make the syntax less complex and may substitute simpler words and/or phrases for harder ones. A similar example

document and controls in a VDE container 3902(n) involving a 12th 3906(2) and 9th grader 3906(n) is also shown.

In other examples, the prices may be higher when users are reading text below their capabilities, they may be offered discounts
5 for reading at a higher level, and/or they may be charged more for reading on different levels since modifying the text is a value added process, and providers of that value may wish to be compensated for their efforts.

10 **Example: Classification May Describe Degree Of Focus Of The Content Unit Or Portion On A Topic, Or Characteristics Related To Conventional Formatting, Such As File Type**

Sometimes the most interesting and/or useful content is at the intersection of various topics. Also, user often want content in a form
15 or format that will be most useful, and most practical, to them. In the example 4000, shown in Figure 65, a matching and classification utility 900 receives from user 4002 a VDE container 4004 holding a request for documents in the class, "on economics and politics, costing less than \$5.00, and in MS Word format." The matching and
20 classification utility 900 responds in this example by providing in a VDE container 4006 at least one Uniform Resource Locator (URL) that points to the location of the document(s) on the World Wide Web.

The user 4002 in this example sends a message in a VDE
25 container 4008 asking for the document identified in the URL. A

provider sends back a VDE container 4012 with the desired document 4010 that has been classified by the matching and classification utility 900. In this example, parameter data is provided in the form of scores indicating the relative emphasis on various topic classes, including
5 Economics (score=15), Politics (score=7), and Religion (score=2). Also indicated is the format of the content, which in this example is the desired MS Word. Also conveyed in the VDE container 4012 are a control set indicating, among other things, that the price is \$2.98 and no modifications are allowed.

10 In other examples, the classes might have been much more narrow, for example, "Clinton," "Greenspan", Federal Reserve Policy, Interest Rates. Also, the customer might have requested only those documents for which controls could be obtained that permitted modifications and/or excerpting and/or derivative works. In another
15 example, the matching and classification utility 900 may send one or more administrative events and/or classification and/or matching methods to the customer so that these methods could be applied by the customer. Alternatively, the customer may have send one or more methods as part of a smart object to one or more information
20 providers in search of information meeting the desired criteria.

**Example: The Atomic Aspects Can Support
Automated Extraction Of Portions Of A Content Unit
For Aggregation With Topically Consistent Portions
And/Or Units From Other Sources**

5 Not only may people desire specific information, but that
information may come from different parts of the same object or parts
of two or more objects. The matching and classification utility 900
can support the use of smart, classification based extraction and
aggregation methods. as shown in example 4100, Figure 66, where
10 two documents 4102(1,2) have been classified by the matching and
classification utility 900 into "chunks" or subobjects reflecting topic
classes and VDE controls have been provided for each chunk. The
"chunking", classification, and control set creation may be performed
and stored in a database and/or may be performed "on the fly" or as
15 needed.

To satisfy a request for information concerning travel to and in
the United Kingdom plus background information, an information
provider extracts parts of each document in the desired classes and
creates a new, recombinant document comprised of the subobjects
20 and packages the new document with appropriate controls in a VDE
container 4102(n). VDE controls for the subobjects may also be
carried along and may be modified by the provider and/or other
participants in a chain of handling and control.

The request for information may have been generated using any
25 query and/or search method, including semantic, Boolean, heuristic,

concept-based, and other approaches, and may have been generated explicitly and intentionally by a user and/or other value chain participant, or may have resulted more automatically from the analysis by a matching and classification utility 900 of usage, audit, and/or other rights management information and/or of "info exhaust," and/or of preference, demographic, and/or psychographic data and/or classes of data.

In another example, the matching and classification utility 900 may have sent administrative events and/or classification, search, and/or subobject combining methods 4106 to a provider and/or to a user for execution under the control of a local VDE node.

Example: Matching And Classification Utility 900 Supports Classification For Subsets Of Content Within A Content Unit (Nested Virtual Classifications)

Not only may the matching and classification utility 900 assist in locating whole objects, it may also assist in identifying and/or classifying any number of subobjects for a given whole. New control sets may be associated with each of these subobjects. These new control sets may differ from the control set that applies to the object as a whole. This capability allows matching and classification utility 900 and others value chain participants to locate desired classes of content that may be part of a larger object and possibly to retrieve, pay for, manage, use, or combine these parts in addition to, and/or instead of the whole object.

In example 4200, Figure 67, a VDE container 4202 created by the matching and classification utility 900 holds a text document that in this non-limiting example is the US "State of the Union Address." The matching and classification utility 900 has first classified the entire document in the class "politics." The matching and classification utility 900 has also identified various subparts or subobjects and has classified each them into different classes or categories. In this example, the different classes represent different topic categories.

10 A user and/or other value chain participant may request only subobjects that have been categorized in one or more desired class(es). The desired subobjects may be packaged in a VDE container 4204 along with appropriate VDE controls for both the overall, new composite object and/or for each of the desired subobjects. (The VDE controls can also be sent separately from the content subobjects.) These controls may pertain to the new whole object created from subparts selected on the basis of their membership in one or more specified class(es) and/or to the whole, new object comprised of these selected subobjects. In another example, the subobjects may be drawn from different documents sharing the same overall topic, for example, from State of the Union addresses given in different years.

In one example, any value chain participant may send distribute one or more subparts of the original object.

In another example, the matching and classification utility 900 may send one or more administrative events and/or methods 4206 to value chain participants who may execute the methods to perform the operations to identify subobjects and/or to subset the whole object in
5 to parts based on class assignments.

Search engines can also use the subobject classifications to provide more precise results. For example, a search engine may have retrieved the State of the Union Address because the search criteria were "US politics speeches," but the whole or part of the object may
10 also have been retrieved searching for "US politics speeches welfare" or "speeches US president defense."

**Example: Matching And Classification Utility 900
Supports Classes Of Classes Based On Object
Identifier Standards And/Or Other Object Metadata**

15 Among the numerous advantages of the present inventions is the ability to create classes of classes based in part on rights management information. The feature may enhance search efficiency by enabling search engines to locate members of classes provided by any of numerous schemes for object naming and object metadata that
20 have been proposed. For example, the IETF Uniform Resource Locator (URL), the International Standard Book Number (ISBN), International Standard Serial Number (ISSN), MARC library catalog records, and the recent proposed "Dublin Core"(Weibel, Stuart, Jean Godby, Eric Miller, and Ron Daniel, "OCLC/NCSA Metadata
25 Workshop Report", URL <http://www.oclc.org:5047>

/oclc/research/conferences/metadata/ dublin_core_report.html) are non-limiting examples of prior classifications that can themselves be classified using the present inventions.

- Example 4300, Figure 68A-68B, shows several objects
- 5 4304(1)-4304(n) each of which may have associated with it various metadata 4302(1)-4302(n) that locates the object in one or more classes, non-limiting examples of which may include network address (URL), price, control set information, permission strings, subject category, title, and publisher.
- 10 In example step "1," object metadata 4302 is sent to a matching and classification utility 900 which (example step "2") may create new "classes of classes" 4306. These new classes 4306 are then made available on a Web page 4308 (example step "3") to interested parties who may then search for objects according to their membership in
- 15 one (or more) of these new classes of classes. In example step "4" an interested party 4320 sends a VDE container with a request to retrieve the Web page 4308 with the classes of metadata information. The Web server (in example step "5") returns a copy of the page 4312 to the interested user 4320, who (in example step "6") sends a VDE
- 20 container with a query to the matching and classification utility 900 asking, in this example, for objects in new class 3 that cost less than \$1.98, and that grant a "modify" permission. In example step "7," the matching and classification utility 900 returns a VDE container 4316 with list of objects that match the criteria. The matching and

classification utility 900 may, in turn, provide URLs or other location information for at least one member of the desired class(es) in the list in container 4316.

Example: Matching and Classification Utility 900

5 Supports Electronic Gambling

Electronic gambling may be among the services that will drive Internet growth in coming years. Such services raise many questions for both providers and for users or players of the service. For example, providers want to be able create attractive, compelling
10 entertainment experiences and in doing so, capture an important share of their intended markets. Users of these services will of course want to locate the most stimulating, entertaining, and perhaps most of all, rewarding gambling experiences.

Gambling providers may, in one example, differing classes of
15 games, rules, payoffs, odds, and/or interfaces. The present inventions can assist players in identifying the nature of various classes and locating specific instances of one or more classes. Within a particular class of games, for example, players may be particularly interested in the odds at the game of blackjack. In one example, a player may
20 prefer playing with a single digital deck of 52 cards and a particular number of (emulated) shuffles rather than with say four decks and more shuffles, the affect of the latter being to create a more random distribution. Smaller decks and fewer shuffles may make it easier to count cards and/or to otherwise increase the odds in favor of the
25 player, or at least in favor of the experienced, knowledgeable player.

In example 4400, shown in Figure 69, an arbitrary number of gamblers 4402(1)-4402(n) whose usage information flows in VDE containers 4404(1)-4404(n) to a usage clearinghouse 300. The usage clearinghouse 300 sends in VDE containers 4406 at least some of this
5 usage information to a matching and classification utility 900. In another example, the usage information may be sent directly from at least one user to the matching and classification utility 900. In this example, an arbitrary number of gambling providers 4406(1)-4406(n) may also send in VDE containers 4408(1)-4408(n) descriptive and/or
10 usage information to the matching and classification utility 900. Based on available information from relevant sources, the matching and classification utility 900 may create one or more classes and assign one or more providers, services, and/or users to a class. These class definitions may at least in part be based on privacy-related
15 control information.

In this one example, a gambler 4402(1) sends a VDE container 4410 with a query concerning best odds for blackjack to a matching and classification utility 900, who, in turn, sends back a VDE container 4412 with content indicating that gambling provider 2 gives
20 the best odds in blackjack, "best" here meaning those most favorable to the player. In another example, the gambler may then contact gambling provider 2 to play, and the play may consist of a series of communications in VDE containers between the gambling provider and the gambler.

**Example: Matching and classification utility 900
Supports Electronic Ticket Sales and Distribution**

The performing arts, exhibitions, theaters, and conferences are some non-limiting examples of events that may require tickets for admission. Electronic ticket agencies on the Internet and other electronic arenas provide a connection between the consumer and producers of the event. Consumers may want to know such information as the nature of the event, what classes of tickets exist for a given event and/or class of events, the price for different classes of tickets to an event, the availability of different classes of tickets to different classes of events, and similar information.

In the example 4500, shown in Figure 70, an arbitrary number of users 4504(1)-4504(n) whose usage information is sent in VDE containers 4508 to a usage clearinghouse 300 who, in turn, may send at least some of this usage information in at least one VDE container 4526 to a matching and classification utility 900. The usage information may reflect past ticket purchases, prices, seating preferences, preferred payment methods, preferred theaters and other venues, and other user preference and historical information.

Various ticket agencies 4506(1)-4506(n) may send information about specific events 4512 (1)-4512(n) and/or information about agency services 4514(1)-4514(n) to the matching and classification utility 900. In another example, an event promoter may send event information directly to the matching and classification utility 900.

In one example, a user wishes to find four seats for a particular concert or class of concerts and/or other events whose cost is not more than \$25.00. The user sends a VDE container with a request for information on who can supply the desired tickets to the desired
5 events at the requested price. In turn, the matching and classification utility 900 returns a VDE container indicating that tick agency 2 can provide the tickets.

In this example, user 2 sends a VDE container with a purchase request to ticket agency 2. The purchase request may specify not only
10 the specific event, desired pricing, and class of tickets, seat location, for example, but payment method as well, MasterCard for example. The ticket agency, in turn, may return a VDE container with confirmation of the ticket purchase at a given price, location, date, event, and/or using a particular payment method.

15 In another example, the tickets may be digital and may have associated with them one or more "seals", digital signatures, and/or certificates indicating the authenticity and/or integrity of the digital tickets.

* * * *

20 While the inventions have been described in connection with what is presently considered to be the most practical and preferred embodiments, the inventions are not to be limited to the disclosed embodiments but, on the contrary, is intended to cover various

modifications and equivalent arrangements included within the spirit and scope of the appended claims.

WE CLAIM:

- 1 1. A method including:
 - 2 (a) determining at least one class, class hierarchy, classification
 - 3 scheme, category or category scheme;
 - 4 (b) assigning cases, persons, and/or things to said determined
 - 5 class, class hierarchy, classification scheme, category or category
 - 6 scheme; and
 - 7 (c) selecting and/or matching cases, persons, and/or things
 - 8 based at least in part on said class, class hierarchy, classification
 - 9 scheme, category or category scheme and/or said assignment,
 - 10 wherein at least one of said steps (a)-(c) includes the step of
 - 11 using at least some rights management information.

- 1 2. A method as in claim 1 wherein said using step includes
- 2 using at least one control set.

- 1 3. A method as in claim 1 wherein said using step includes
- 2 using at least some information for controlling use of digital
- 3 information.

- 1 4. A method as in claim 1 wherein said using step includes
- 2 using at least some information for controlling at least one
- 3 transaction.

- 1 5. A method as in claim 1 wherein said using step includes
- 2 using at least some information for controlling at least one event.

1 6. A method as in claim 1 wherein said using step includes
2 using at least some information for controlling at least one
3 consequence of digital information use.

1 7. A method as in claim 1 wherein said using step includes
2 using at least some information for controlling at least one
3 consequence of at least one event.

1 8. A method as in claim 1 wherein said using step includes
2 the step of using at least some information for controlling at least one
3 consequence of at least one transaction.

1 9. A method as in claim 1 wherein said using step includes
2 using at least some information outputted by a rights management
3 process.

1 10. A method as in claim 1 further including the step of
2 outputting at least some rights management information.

1 11. A method as in claim 1 wherein at least one of steps (a)-
2 (c) includes using at least one secure container.

1 12. A method as in claim 1 wherein at least one of steps (a)-
2 (c) includes using at least one protected processing environment.

1 13. A method as in claim 1 further including the step of
2 using at least one of the techniques set forth at pages 60-82 of this
3 specification.

1 14. A method as in claim 1 wherein said using step includes
2 using at least one or more rules and/or their consequences.

1 15. A method as in claim 1 wherein at least one of steps (a)
2 and (b) includes at least one of the following steps:

3 (a) using at least one statistical technique identifying at least
4 one cluster of cases sharing similar profiles and/or features;

5 (b) using numerical taxonomy;

6 (c) using at least one of cluster analysis, factor analysis,
7 components analysis, and other similar data reduction/classification
8 technique;

9 (d) using at least one pattern classification technique, including
10 components analysis and neural approaches;

11 (e) using at least one statistical technique that identifies at least
12 one underlying dimension of qualities, traits, features, and/or
13 characteristics, and assigning parameter data indicating the extent to
14 which a given case has, possesses, and/or may be characterized by the
15 underlying dimension, factor, class, and/or result in the definition of
16 at least one class and/or the assignment of at least one case to at least
17 one class;

18 (f) using at least one statistical method employing fuzzy logic
19 and/or fuzzy measurement and/or whose assignment to at least one
20 class entails probabilities different from 1 or zero;

21 (g) using a Bayesian statistical classification techniques that uses
22 an estimate of prior probabilities in determining class definitions
23 and/or the assignment of at least one case to at least one class;

24 (h) using at least one statistical and/or graphical classification
25 and/or data reduction method that uses rotation of reference axes,
26 regardless of whether orthogonal or oblique rotations are used;
27 (i) using at least one statistical method for two and three way
28 multidimensional scaling; and
29 (j) using at least one knowledge based approach to
30 classification.

1 16. A system including:
2 an automatic class generator that generates at least one class,
3 class hierarchy, classification scheme, category or category scheme;
4 an automatic class assigner that assigns cases, persons and/or
5 things to said determined class, class hierarchy, classification scheme,
6 category or category scheme; and
7 at least one further component for automatically searching,
8 selecting and/or matching cases, persons, and/or things based at least
9 in part on said class, class hierarchy, classification scheme, category
10 or category scheme and/or said assignment,
11 wherein said system uses at least some rights management
12 information.

1 17. A system including:
2 first means for determining at least one class, class hierarchy,
3 classification scheme, category or category scheme;
4 second means for assigning cases, persons, and/or things to
5 said determined class, class hierarchy, classification scheme, category
6 or category scheme; and
7 third means for selecting and/or matching cases, persons,
8 and/or things based at least in part on said class, class hierarchy,
9 classification scheme, category or category scheme and/or said
10 assignment,
11 wherein at least one of said first, second and third means uses
12 at least some rights management information.

1 18. A Commerce Utility System providing a secure
2 execution space, the Commerce Utility System performing at least
3 one component based service function including at least one secure
4 component for execution within the secure execution space, the
5 Commerce Utility System including a communications facility
6 permitting communication of secure control information with at least
7 one electronic community participant,
8 wherein said component based service function uses at least
9 one class based at least in part on rights management information.

1 19. A Commerce Utility System as in claim 18 wherein the
2 component based service function assigns at least one member to at

3 least one class based at least in part on some rights management
4 information.

1 20. A Commerce Utility System as in claim 18 wherein the
2 component based service function matches persons and/or things
3 based at least in part on at least some rights management information.

1 21. A Commerce Utility System as in claim 18 wherein the
2 component based service function selects persons and/or things based
3 at least in part on at least some rights management information.

1 22. A Commerce Utility System as in claim 18 wherein the
2 component based service function narrowcasts information to
3 recipients based at least in part on at least some rights management
4 information.

1 23. A system or method including:
2 a computer network and
3 a control arrangement within the network that determines
4 and/or uses at least one of the following through use of rights
5 management information:

- 6 (a) class hierarchy,
7 (b) class structure,
8 (c) classification scheme,
9 (d) category, and
10 (e) category scheme.

1 24. A class-based system including at least one computer
2 that processes digital information, said system including at least one
3 element that uses at least some rights management information.

1 25. A method of operating a class-based system including at
2 least one computer that processes digital information, said method
3 including the step of using at least some rights management
4 information.

1 26. A system for assigning at least one thing or person to at
2 least one class including at least one computer that processes digital
3 information, said system including at least one element that uses at
4 least some rights management data in making said assignment.

1 27. A system for making and/or using at least one class-
2 based assignment including at least one computer that processes
3 digital information, said system including at least one element that
4 uses at least some rights management information.

1 28. A system for clearing at least one transaction including at
2 least one computer that processes digital information, said system
3 including at least one element that uses at least one class defined,
4 assigned, selected, and/or matched based at least in part on rights
5 management information.

1 29. A method for authorizing at least one computer and/or
2 computer user including the step of using at least one class defined,

3 assigned, selected, and/or matched based at least in part on rights
4 management information.

1 30. A method for authorizing at least one electronic
2 transaction including the step of using at least one class defined,
3 assigned, selected, and/or matched based at least in part on rights
4 management information.

1 31. A method for initiating and/or performing at least one at
2 least in part secure electronic transaction including the step of using
3 class related information defined, assigned, selected, and/or matched
4 based at least in part on rights management information.

1 32. An information processing method including the steps
2 of:
3 securely charging a fee; and
4 conditioning said charging step at least in part on at least one
5 class defined, assigned, selected, and/or matched based at least in part
6 on rights management information.

1 33. A method for securely exchanging digital information
2 including the step of at least in part defining, assigning, selecting,
3 and/or matching at least one class based at least in part on rights
4 management information.

1 34. A method for performing at least one rights operating
2 system based transaction including the step of defining, assigning,

3 selecting, and/or matching at least one class based at least in part on
4 rights management information.

1 35. A method for performing at least one protected
2 processing environment operation including the step of defining,
3 assigning, selecting, and/or matching at least one class based at least
4 in part on rights management information.

1 36. A method of pushing information including the steps of
2 classifying recipients and/or information to be sent to said recipients
3 based at least in part on rights management information, and selecting
4 said information to distribute to said recipients based at least in part
5 on said classifying.

1 37. A method of pushing information including the steps of
2 classifying recipients and/or information to be sent to said recipients
3 based at least in part on rights management information, and
4 matching at least a portion of said information with at least one class
5 of said recipients based at least in part on said classifying.

1 38. A method of pushing information as in claim 37 further
2 including the step of creating a classification scheme and/or hierarchy
3 using at least some rights information.

1 39. A method of pushing information as in claim 37 further
2 including the step of assigning at least some information and/or at
3 least one recipient to a class or category, said assignment based at
4 least in part on rights management information.

1 40. A subject switch for matching subscribers and/or
2 recipients desiring information in one or more classes with one or
3 more sources of information, wherein the subject switch matches at
4 least one subscriber and/or participant with at least one information
5 source on a mapping based at least in part on rights management
6 information.

1 41. A subject switch as in claim 40 wherein said information
2 source:

3 selects at least some information, said selection based on at
4 least one class, and wherein said assignment of said at least some
5 information to said at least one class is based at least in part on rights
6 management information; and

7 sends at least some said selected information to said subscriber
8 in accordance with said subscriber's subscribing to said class of
9 information.

1 42. A subject switch as in claim 40 wherein at least one of
2 said subject switch, said subscriber and/or participant and said
3 information source includes at least one computer providing a
4 protected processing environment.

1 43. A subject switch as in claim 40 wherein at least one
2 subscriber and/or participant uses rights management information at
3 least in part to persistently subscribe to at least some information
4 provided by at least one information source.

1 44. A subject switch as in claim 40 wherein the subject
2 switch includes means for using at least one class definition for said
3 mapping.

1 45. A subject switch as in claim 40 wherein the subject
2 switch includes means for responding to a subscriber and/or
3 participant request by providing information indicating information
4 sources in at least one specified or desired class.

1 46. A subject switch as in claim 40 further including a
2 messaging service for use by at least two of said subject switch, said
3 subscriber and/or participant and said information source and/or
4 participant to communicate electronically.

1 47. A subject switch as in claim 46 wherein said electronic
2 communications uses at least one secure container.

1 48. A subject switch as in claim 40 wherein at least one of
2 said subject switch, subscriber, or information source uses at least one
3 control set associated with at least some information received by at
4 least one subscriber.

1 49. A digital narrowcasting arrangement comprising:
2 a computer; and
3 at least one classifying element used to select content to
4 narrowcast to recipients based at least in part on rights management
5 information.

1 50. A digital narrowcasting arrangement as in claim 49
2 wherein the classifying element classifies at least one of (a) a
3 recipient, and (b) content, based at least in part on rights management
4 information.

1 51. A digital narrowcasting arrangement as in claim 49
2 wherein said classifying element defines at least one class using at
3 least some rights management information.

1 52. A digital narrowcasting arrangement as in claim 49
2 wherein the classifying element assigns at least some content to at
3 least one class, said assignment based on at least some rights
4 management information.

1 53. A digital narrowcasting arrangement as in claim 49
2 wherein the classifying element defines at least one class based at
3 least in part on content selections previously made by the recipients
4 and/or profiles generated based at least in part on recipient input.

1 54. A digital narrowcasting arrangement as in claim 49
2 wherein the classifying element sends a content request including
3 classification data and destination information to at least one
4 provider.

1 55. An information distribution system including: a
2 computer network; and a selection arrangement that selects
3 information for use by individual recipients using classes based at
4 least in part on rights management information.

1 56. An information distribution system as in claim 55
2 wherein the system further includes a classifying element that
3 determines at least one class of content and/or service of interest to at
4 least one recipient.

1 57. An information distribution system as in claim 56
2 wherein said classifying element defines at least one class using at
3 least some rights management information.

1 58. An information distribution system as in claim 56
2 wherein said classifying element assigns at least some content to at
3 least one class, said assignment based on at least some rights
4 management information.

1 59. An information distribution system as in claim 55
2 wherein the system includes means for allowing the user to choose to
3 receive the selected information.

1 60. An enterprise information system including a computer
2 system for classifying employees, said system including at least one
3 rights management component that distributes information to the
4 employees based at least in part on employee classification.

1 61. An enterprise information system as in claim 60 wherein
2 the computer matches the information to employees based at least in
3 part on the employee classification.

1 62. An enterprise information system as in claim 60 wherein
2 the employee classification is used to gather information for

3 employees without revealing substantial information concerning
4 individual employees.

1 63. A method for conducting a chain of handling and/or
2 control including the steps of allowing plural parties to contribute
3 rules and/or consequences, and performing at least one classification
4 based at least in part on said rules and/or consequences.

1 64. A method as in claim 63 wherein at least some of said
2 contributed rules and/or consequences are class based.

1 65. A method as in claim 63 wherein at least one of said
2 parties modifies at least one of said rules and/or consequences based
3 at least in part on class.

1 66. A method as in claim 63 including the step of generating
2 class assignments based at least in part on said rules and/or
3 consequences, and sending said class assignments to at least one
4 clearinghouse.

1 67. A method as in claim 63 including the step of classifying
2 said rules and/or consequences to provide at least one class, and
3 fulfilling at least one request by selecting based on said class.

1 68. A directory services system for classifying confidential
2 information, the system including:
3 a communications component that receives directory requests;
4 and
5 a response component that uses said classification to respond to

6 directory requests while preserving confidentiality of said
7 confidential information.

1 69. A directory services system as in claim 68 wherein said
2 response component uses at least one classification process to classify
3 items in a directory, and uses results of the classification process, at
4 least in part, to respond to directory requests.

1 70. A directory services system as in claim 68 wherein said
2 response component sends information to destinations revealed by the
3 results of the classification process without revealing at least some
4 information concerning said destinations to the information source.

1 71. A microsegmented merchandising technique including
2 the steps of performing classification based at least in part on usage
3 data and/or lifestyle profiles, and distributing offers for products
4 and/or services based at least in part on the classification.

1 72. A microsegmented merchandising technique as in claim
2 71 wherein the performing step includes defining at least one class
3 hierarchy based at least in part on rights management information.

1 73. A microsegmented merchandising technique as in claim
2 71 further including the step of combining plural offers for different
3 products and/or services based at least in part on said classification.

1 74. A trading network including:
2 a communications element for communicating digital signals;
3 and

4 means for matching value chain participants through a
5 classification based at least in part on rights management
6 information.

1 75. A trading network as in claim 74 further including means
2 for defining at least one class hierarchy based at least in part on rights
3 management information.

1 76. A trading network as in claim 74 further including means
2 for determining class membership based at least in part on action
3 and/or information provided by at least one value chain participant.

1 77. A trading network as in claim 74 wherein said matching
2 means includes means for at least in part performing at least one
3 electronic negotiation.

1 78. A securities trading method including the step of
2 performing a classification process at least in part using at least one
3 rights management element, and using the classification process to
4 select securities for trade.

1 79. A securities trading method as in claim 78 wherein said
2 classification process includes defining at least one class hierarchy
3 based at least in part on rights management information.

1 80. A currency/debt trading system including:
2 a currency or debt trading computer; and
3 an arrangement coupled to said computer that performs at least

4 one classification process based at least in part on rights management
5 information.

1 81. A currency/debt trading system as in claim 80 wherein
2 said arrangement includes means for defining at least one class
3 hierarchy based at least in part on rights management information.

1 82. A currency/debt trading system as in claim 80 wherein
2 the arrangement uses classification to maximize return or minimize
3 loss.

1 83. A financial institution selection system including a
2 computer that classifies financial institutions based at least in part on
3 rights management information.

1 84. A software distribution method including the steps of
2 generating class information based at least in part on rights
3 management information, and selecting software to be distributed
4 and/or recipients who are to receive distributed software based at least
5 in part on class information.

1 85. A software distribution method as in claim 84 wherein
2 said generating step includes defining a class hierarchy using at least
3 some rights management information.

1 86. A software distribution method as in claim 84 wherein
2 the selecting step includes selecting software to be distributed by
3 classifying the software based at least in part on rights management
4 information associated with the software.

1 87. A software distribution method as in claim 80 wherein
2 the selecting step includes selecting recipients to receive software
3 based at least in part on usage information provided by a rights
4 management process.

1 88. A classification technique including the step of
2 authenticating class membership based at least in part on digital
3 credentials and/or certificates.

1 89. A classification technique as in claim 88 wherein said
2 digital credentials are digital certificates.

1 90. A classification technique as in claim 88 wherein said
2 digital credentials are digital membership cards.

1 91. A classification technique as in claim 88 further
2 including the step of deciding class membership based at least in part
3 on rights management information.

1 92. A classification technique as in claim 88 further
2 including the step of classifying at least one of users, nodes, devices,
3 networks, servers, clients and services based at least in part on rights
4 management information.

1 93. A classification technique as in claim 88 further
2 including the step of conditioning at least one rights management
3 process at least in part on authenticated class membership.

1 94. A computer system including:
2 a first arrangement that generates class-based controls to
3 participants based at least in part on class and/or class-based
4 assignments; and
5 a second arrangement that allows participants to interact with
6 information and/or one another at least in part using said class-based
7 controls.

1 95. A computer system as in claim 94 further including
2 means for using said class-based controls to limit participants' access
3 to information and/or services based on participants' classes.

1 96. A health care computer system including an arrangement
2 for issuing health care workers, administrators and insurers class-
3 based digital credentials and/or certificates, wherein the digital
4 information sent to said health care workers and administrators
5 includes class-based controls that condition use and/or access to
6 information based at least in part on said class-based digital
7 credentials and/or certificates.

1 97. A health care computer system as in claim 96 further
2 including means for allowing said health care workers, administrators
3 and insurers sharing a common object subject to class-based controls
4 to have access to different portions of the object based at least in part
5 on said class-based controls.

1 98. A work process automation system including a matching
2 and/or classification computer that matches tasks to resources based
3 at least in part on assigning classifying the tasks and/or the resources
4 to at least one class.

1 99. A work process automation system as in claim 98
2 wherein said matching and/or classification computer includes means
3 for defining at least one class hierarchy based at least in part on rights
4 management information.

1 100. A work process automation system as in claim 98
2 wherein said matching and/or classification computer includes means
3 for matching based at least in part on rights management information.

1 101. An automatic governmental and/or societal rights
2 supporting system including a matching and/or classification
3 computing element that assigns and/or classifies entities to at least
4 one class based at least in part on rights management information.

1 102. An automatic governmental and/or societal rights
2 supporting system as in claim 101 wherein the matching and/or
3 classification computing element includes means for defining a class
4 hierarchy based at least in part on rights management information.

1 103. An automatic governmental and/or societal rights
2 supporting system as in claim 101 wherein the matching and/or
3 classification computing element includes means for classifying
4 entities based on at least one of the following:

5 tax status;
6 right to receive certain information;
7 right to engage in certain transactions; and
8 jurisdiction.

1 104. An automatic taxing authority computer including
2 means for issuing tax class control sets based at least in part on tax-
3 based class definitions, and means for using said tax control sets at
4 least in part to collect and/or enforce taxation.

1 105. A method for adaptively presenting information
2 differently to different participants, including associating said
3 participants with classes, and controlling presentation based at least in
4 part on class-based control sets included within the information.

1 106. A method as in claim 105 further including using said
2 class-based control sets to match participants with different portions
3 of said information.

1 107. A method as in claim 105 further including using said
2 class-based control sets to change the form in which information is
3 presented based at least in part on said classes.

1 108. A method as in claim 105 further including the step of
2 operating said class-based control sets based at least in part on
3 metadata associated with different portions of said information.

1 109. A method as in claim 105 further including selecting
2 said class-based control sets between different images for

3 presentation based at least in part on one or more classes associated
4 with a participant.

1 110. A method as in claim 105 further including using said
2 class-based control sets to emphasize certain portions of said
3 information over other portions in said presentation based at least in
4 part on one or more classes associated with a participant.

1 111. A method as in claim 105 further including using at
2 least one computer having a protected processing environment.

1 112. A method for adaptively presenting information
2 differently to different participants including:
3 classifying the different participants based on capability; and
4 using class-based control sets associated with said information
5 to change the difficulty of the presentation based at least in part on
6 said classification.

1 113. A method as in claim 112 wherein the different
2 recipients are classified based on grade level.

1 114. A method as in claim 112 including the step of
2 changing the vocabulary and/or syntactical complexity of the
3 presentation based at least in part on said classification.

1 115. A method as in claim 112 further including the step of
2 using said class-based control sets to ensure that in at least some
3 cases, recipients in different classes pay different levels of
4 compensation for said presentation.

1 116. A method for adaptively presenting information
2 differently to different participants including:
3 classifying different participants based on capability, and
4 using class-based control sets associated with said information
5 to change the language of the presentation based at least in part on
6 said classification.

1 117. An information searching mechanism including a
2 matching computer element that classifies information based at least
3 in part on rights management information, said computing element
4 including means responsive to user requests to search for information
5 based at least in part on said classification.

1 118. An information searching mechanism as in claim 117
2 wherein said matching computer element further includes means for
3 assigning information to classes based at least in part on rights
4 management information.

1 119. An information searching mechanism as in claim 117
2 wherein said matching computer element includes means for scoring
3 information based at least in part on user indicated parameters.

1 120. An information searching mechanism as in claim 117
2 wherein said matching computer element includes means for
3 responding to at least some user requests by providing Universal
4 Resource Locator designations of where information can be found.

1 121. An information handling method including the step of
2 using class-based controls to control support extraction and/or
3 aggregation of information.

1 122. An information handling method as in claim 121 further
2 including using a computing element to extract information from
3 plural objects based at least in part on class-based criteria.

1 123. An information handling method as in claim 121 further
2 including using a computing element to aggregate information based
3 at least in part on class-based criteria.

1 124. An information handling method as in claim 121 further
2 including using said class-based controls to represent nested or multi-
3 level classifications.

1 125. An information classification method including the step
2 of generating at least one class hierarchy from other plural
3 classification hierarchies based at least in part on rights management
4 information and/or class-based rights management information based
5 at least in part on classification metadata.

1 126. An information classification method as in claim 125
2 further including basing said other plural classification hierarchies at
3 least in part on object metadata.

1 127. An information classification method as in claim 125
2 further including specifying said classification object metadata

3 specified classifications based on at least one of location, name,
4 prices, permissions, ISSN, title, author, publisher and/or date.

1 128. An information classification method as in claim 125
2 further including generating said class-based rights management
3 information by classifying classes.

1 129. An electronic gambling system including a computer
2 that matches gamblers with plural gambling providers based at least
3 in part through classifying the gambling providers using rights
4 management information.

1 130. An electronic gambling system as in claim 129 wherein
2 the computer includes means for classifying the gamblers based at
3 least in part on rights management information.

1 131. An electronic gambling system as in claim 129 wherein
2 the computer includes at least one protected processing environment.

1 132. An electronic gambling system as in claim 129 wherein
2 the computer uses at least one control set to classify, select and/or
3 match at least one of said gambling providers, and/or gamblers.

1 133. An electronic ticketing system including a computer
2 that matches recipients with tickets to events through classifying said
3 recipients, said system including a computer that matches tickets
4 and/or said events based at least in part on rights management
5 information.

1 134. An electronic ticketing system as in claim 133 wherein
2 a recipient provides a request containing event and rights
3 management criteria, and the computer matches the recipient with a
4 provider based at least in part on said classifying process.

1 135. An electronic ticketing system as in claim 133 wherein
2 the rights management information includes method of payment
3 information.

1/96

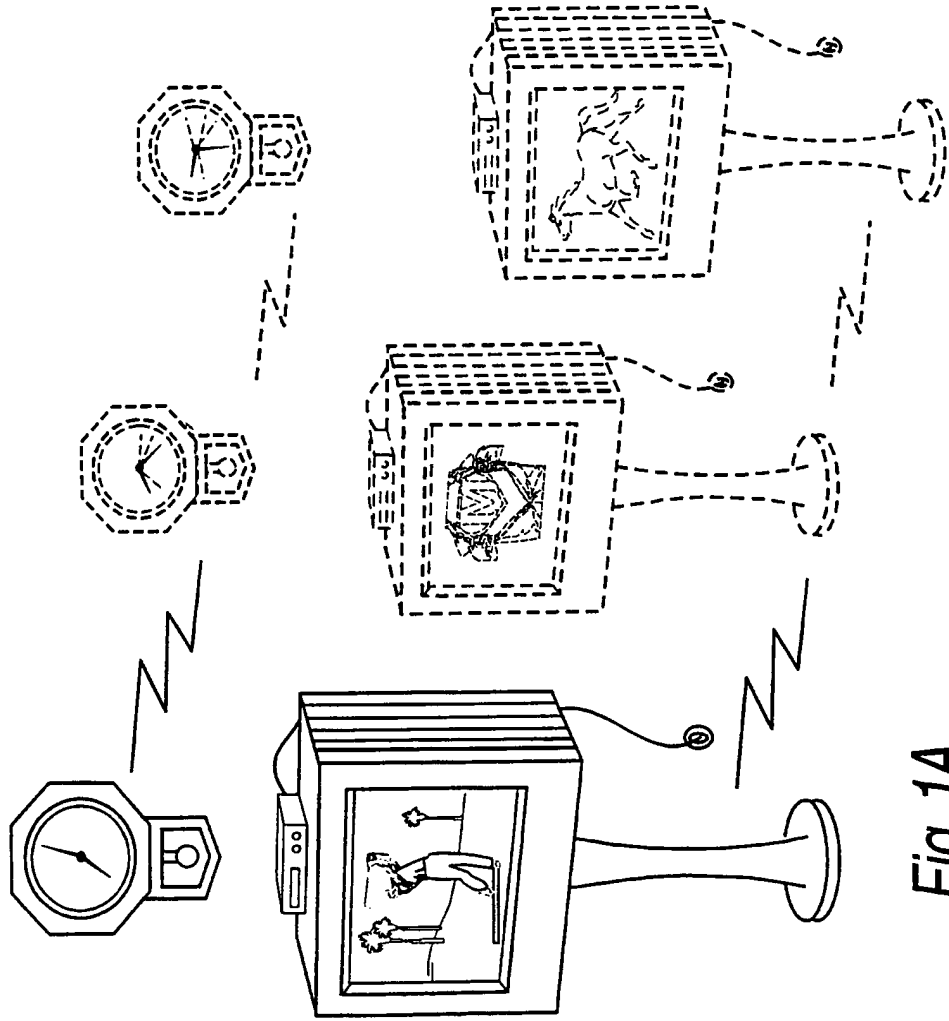


Fig. 1A
(Prior Art)



2/96

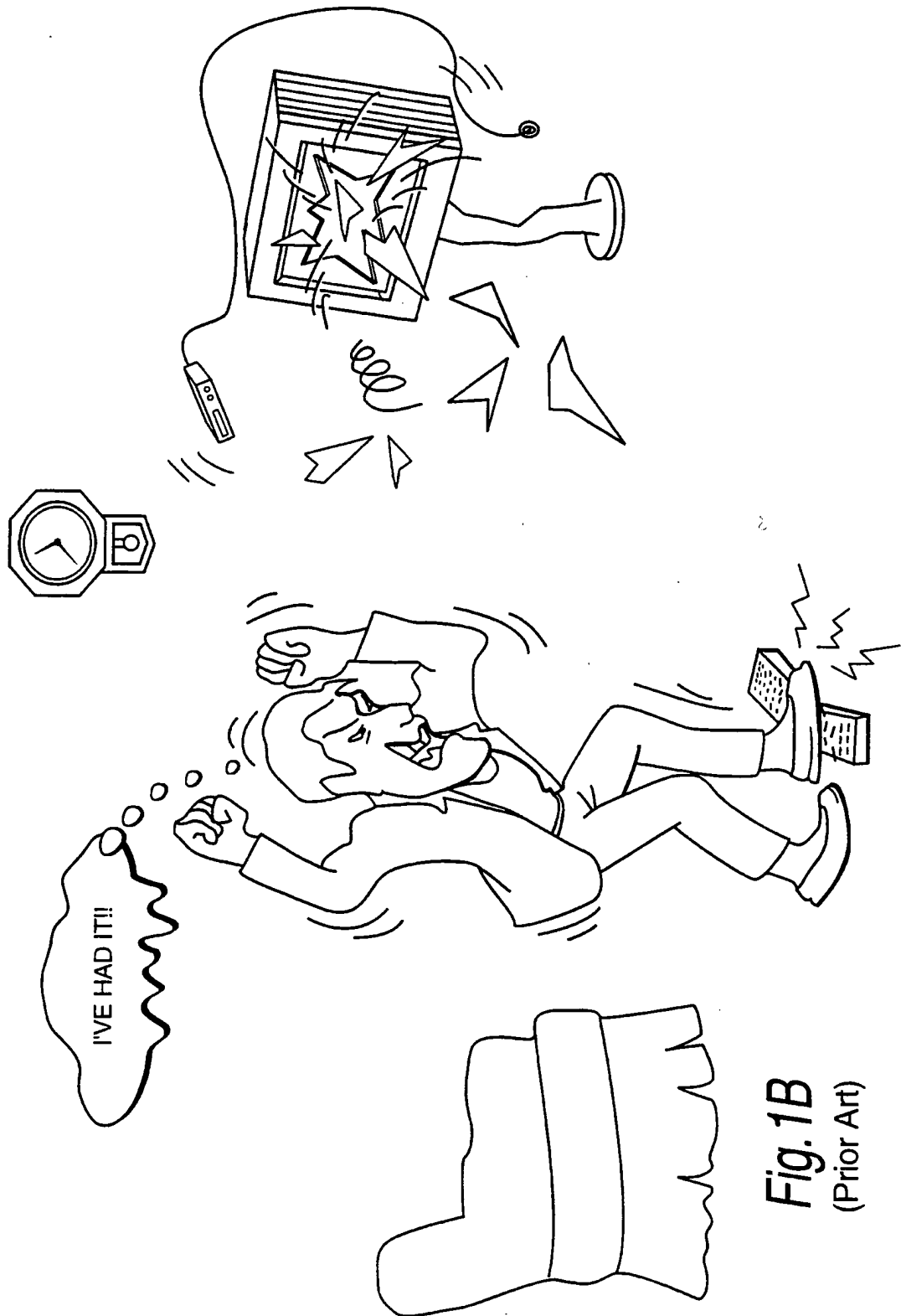


Fig. 1B
(Prior Art)

3/96



Fig. 2
(Prior Art)

4/96

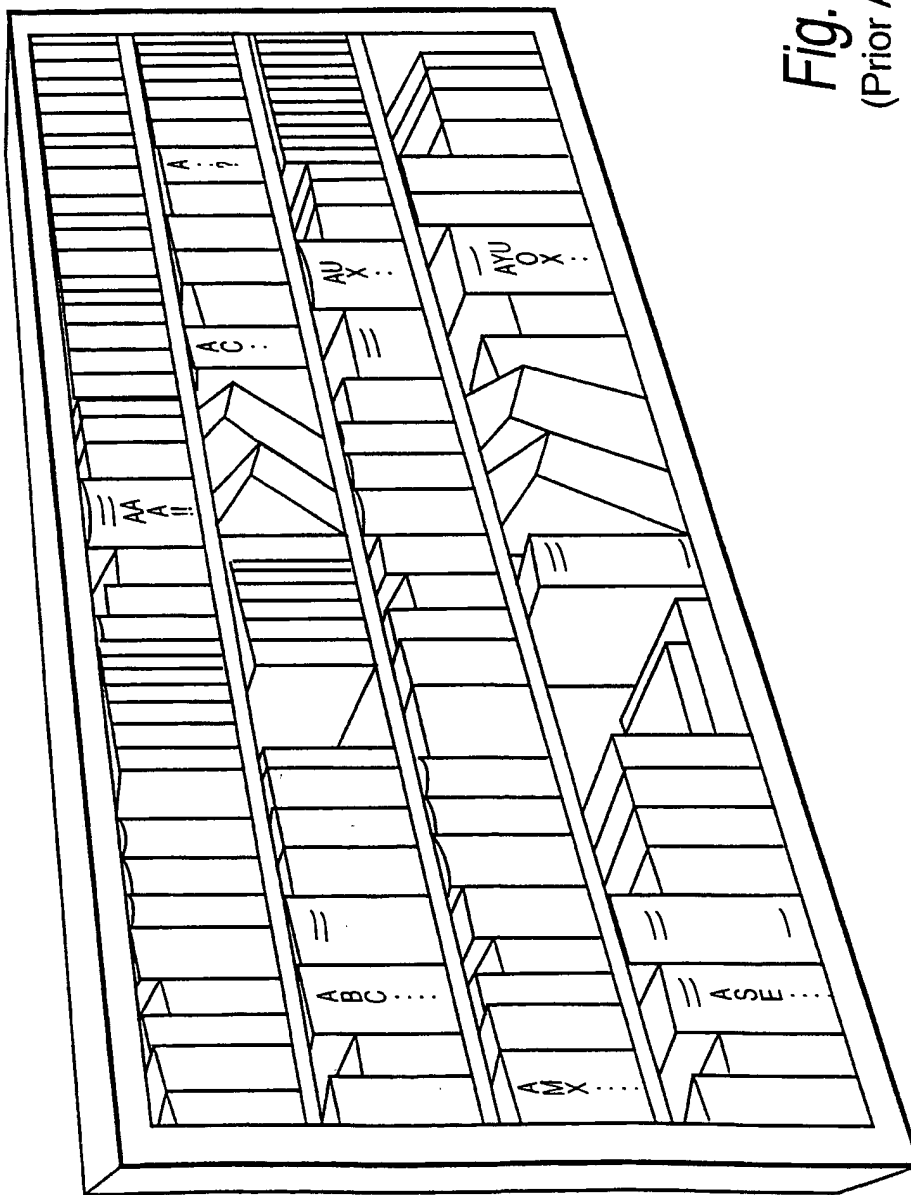
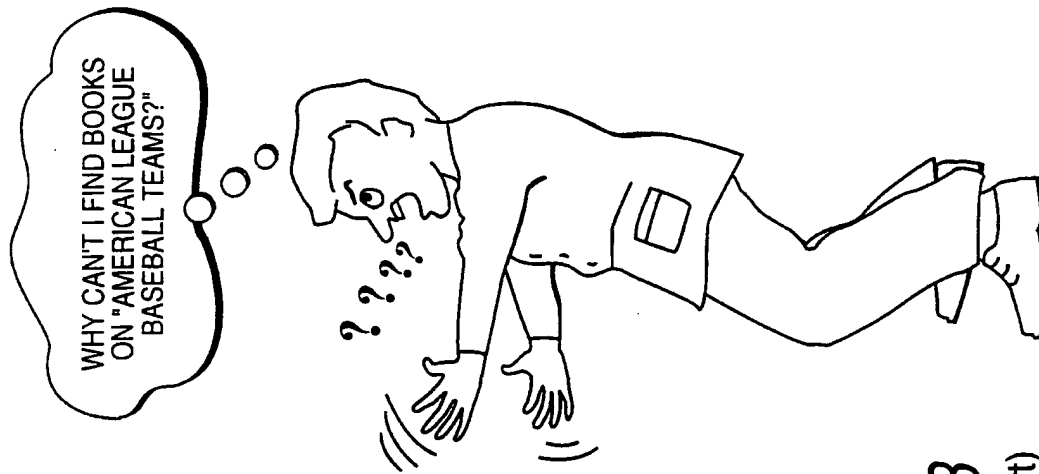


Fig. 3
(Prior Art)

5/96

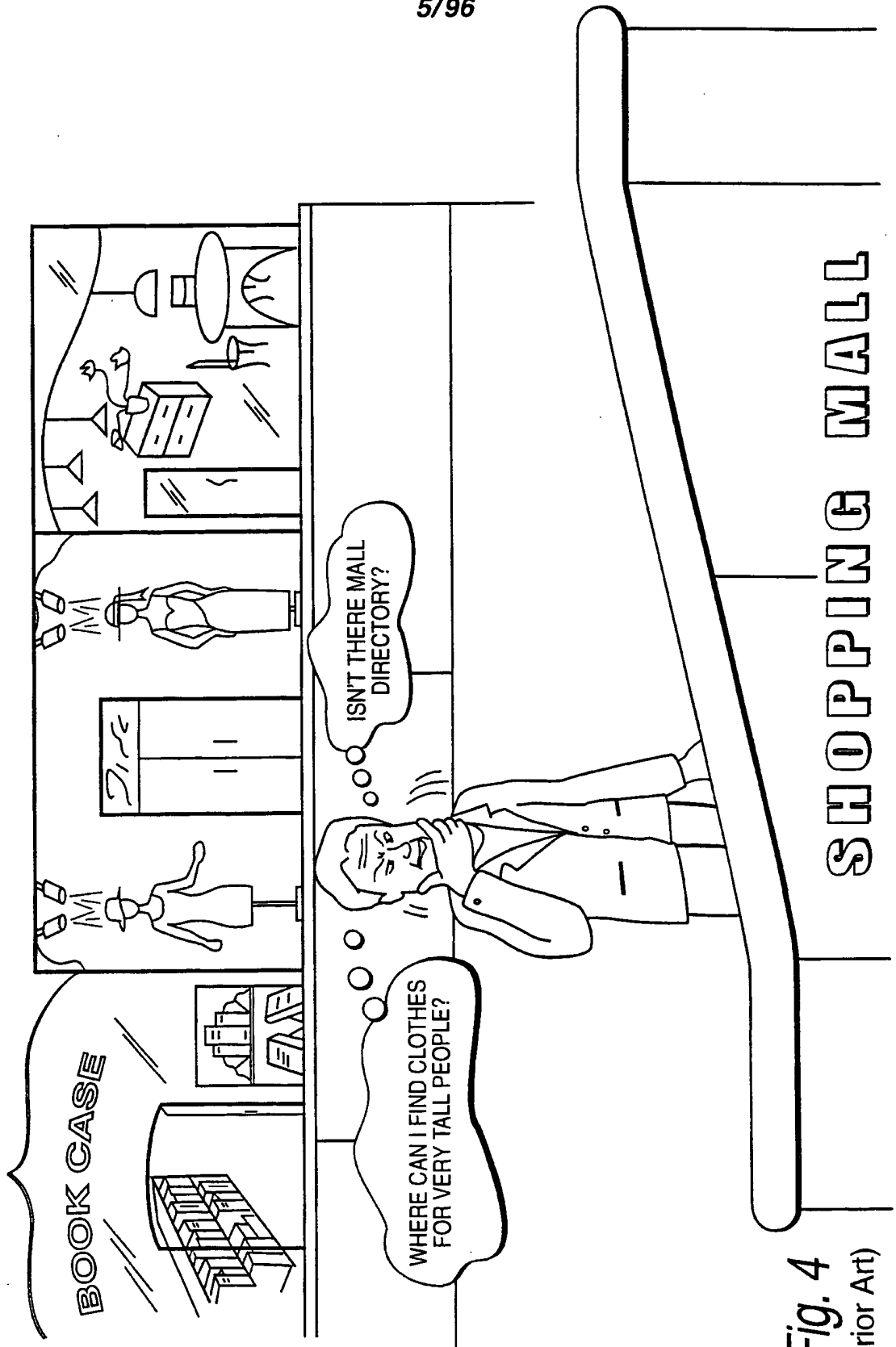


Fig. 4
(Prior Art)

6/96

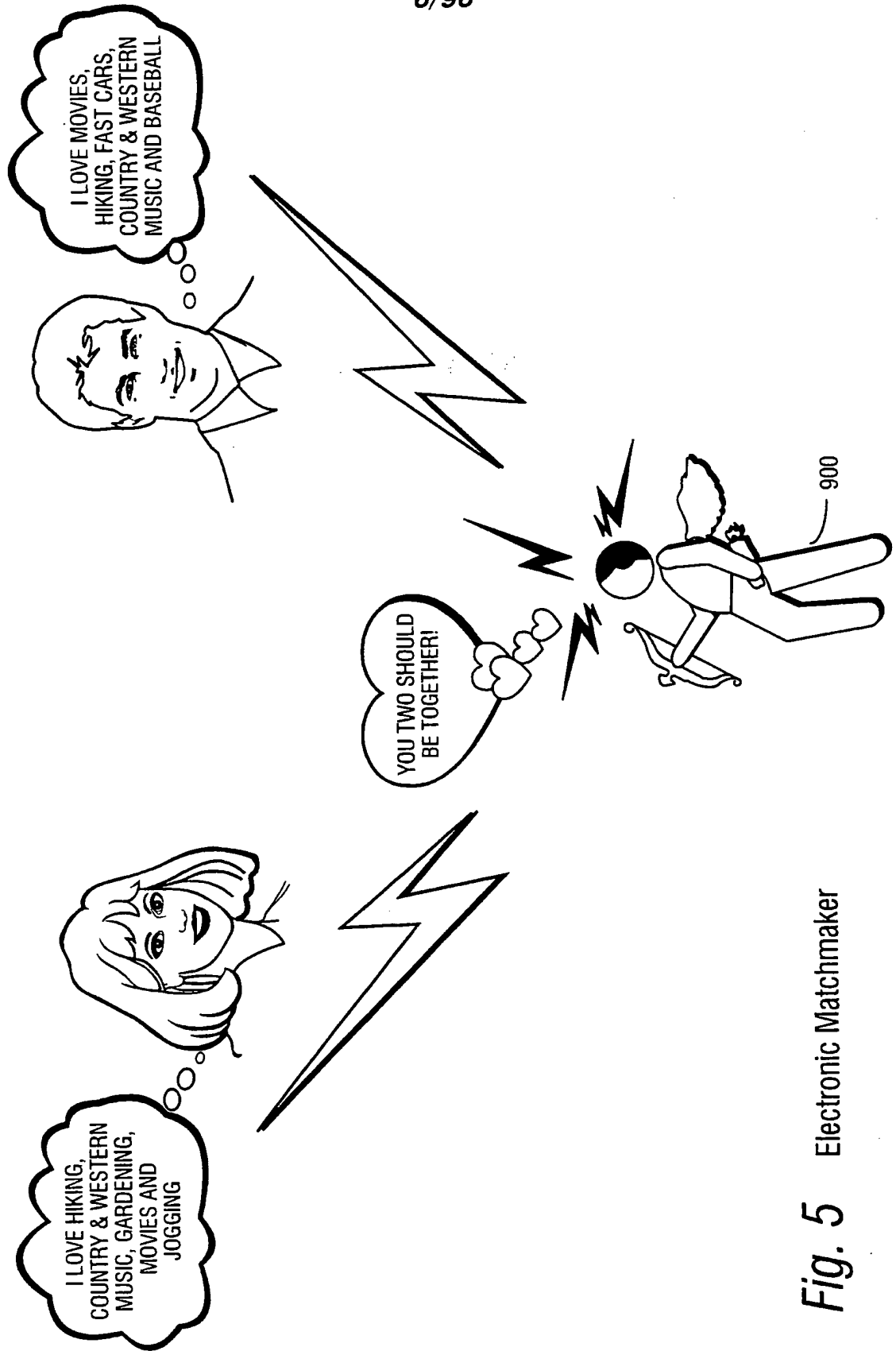
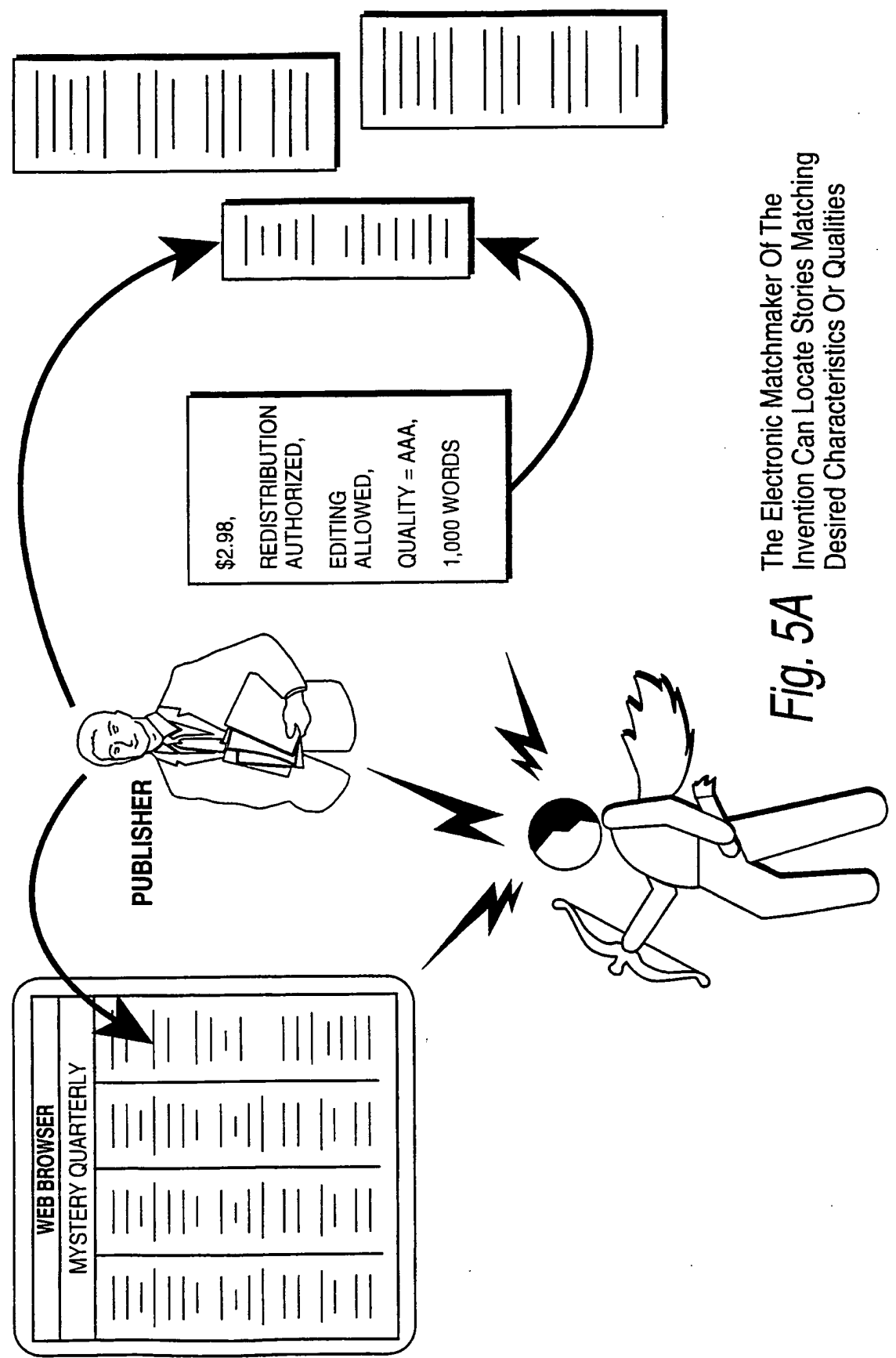


Fig. 5 Electronic Matchmaker



The Electronic Matchmaker Of The Invention Can Locate Stories Matching Desired Characteristics Or Qualities

Fig. 5A

8/96

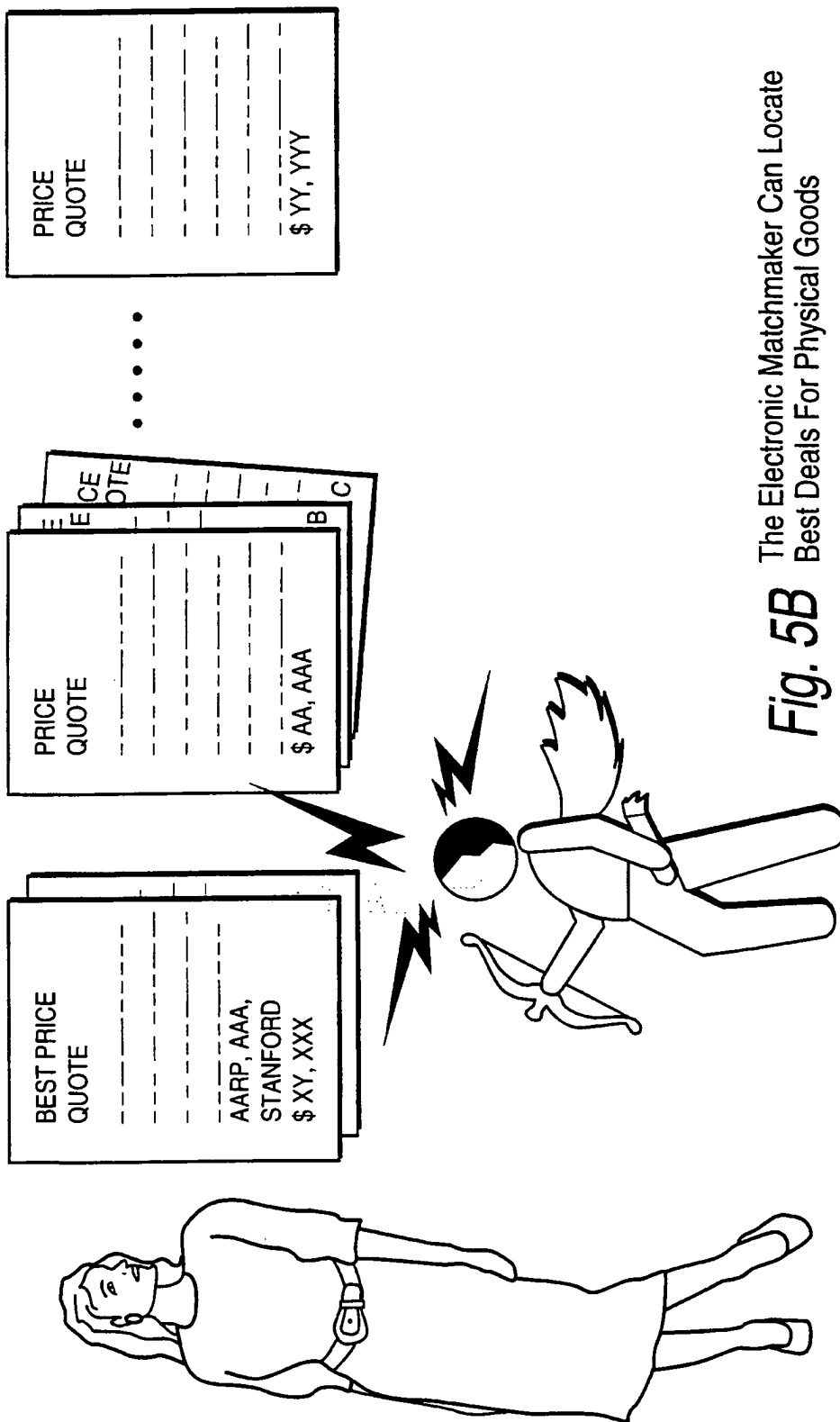


Fig. 5B The Electronic Matchmaker Can Locate Best Deals For Physical Goods

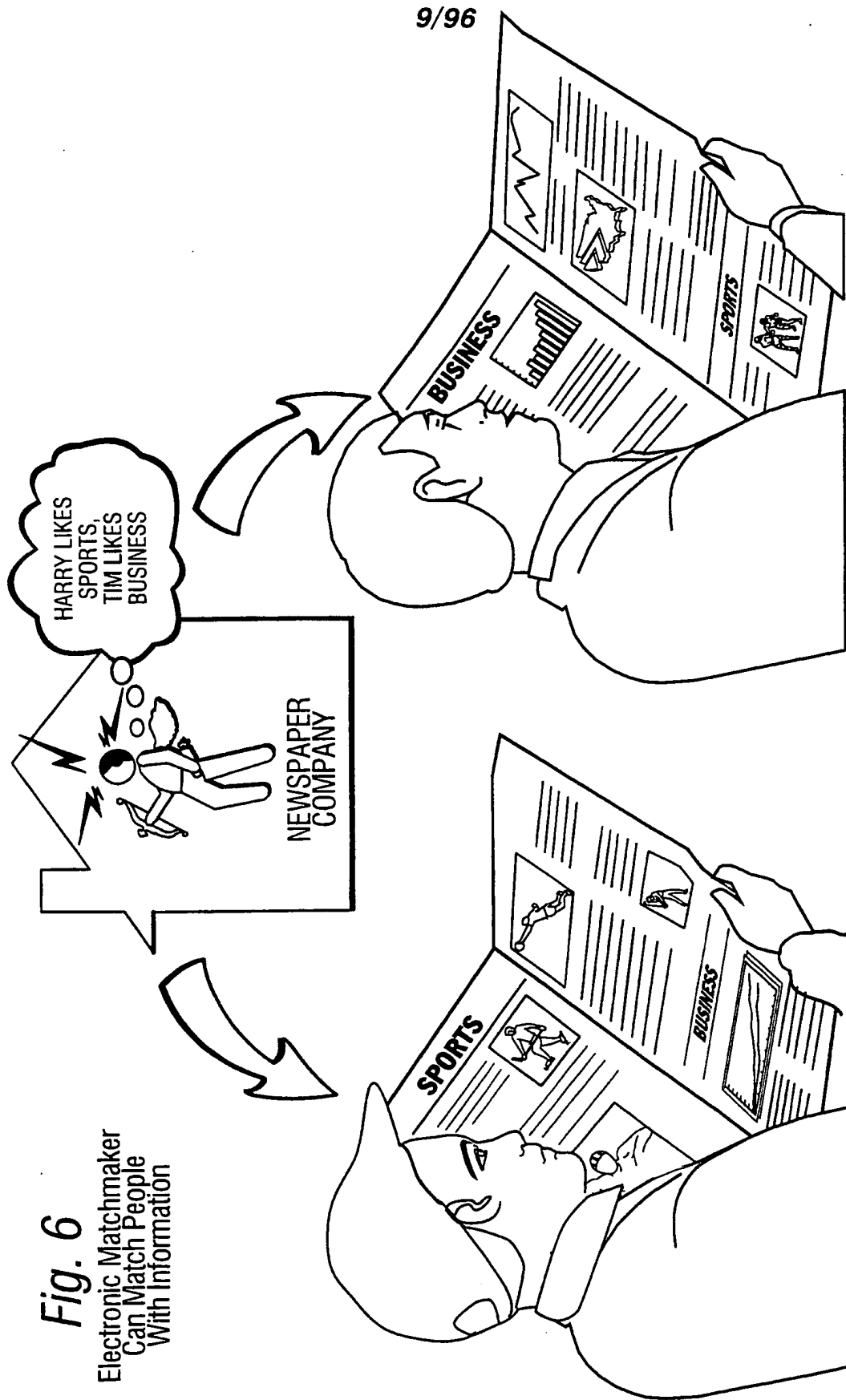


Fig. 6
Electronic Matchmaker
Can Match People
With Information

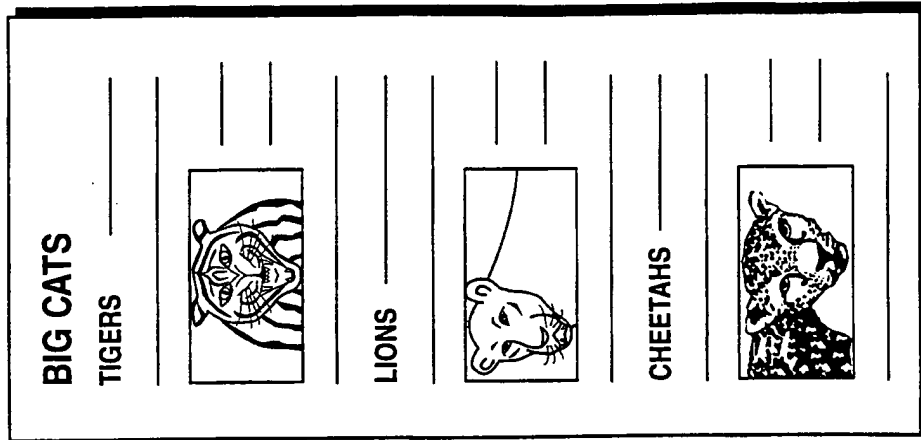


Fig. 7 Electronic Matchmaker Can Match Different Kinds Of Content

11/96

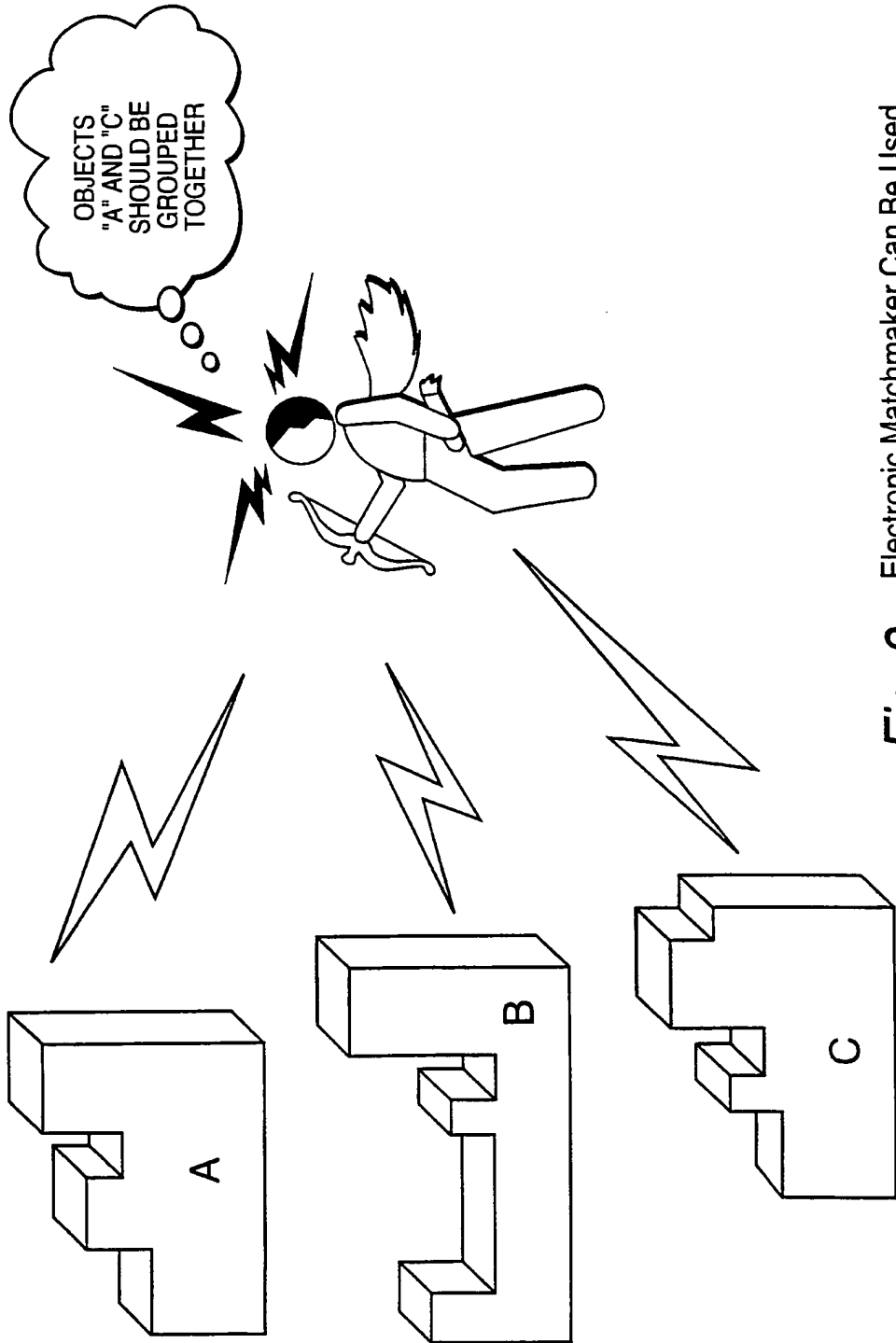


Fig. 8 Electronic Matchmaker Can Be Used For Matching Any Kinds of Things

12/96

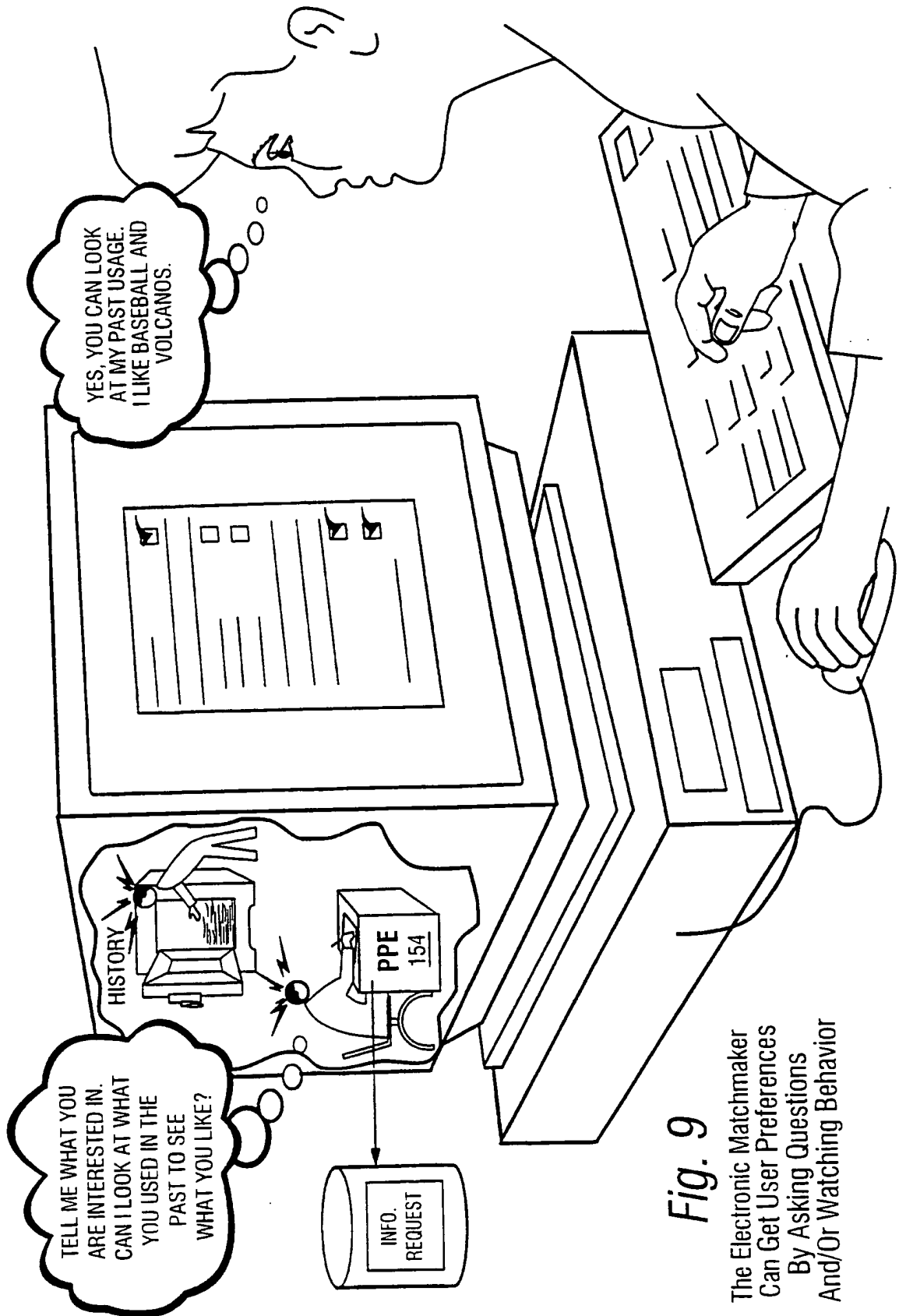


Fig. 9
The Electronic Matchmaker
Can Get User Preferences
By Asking Questions
And/Or Watching Behavior

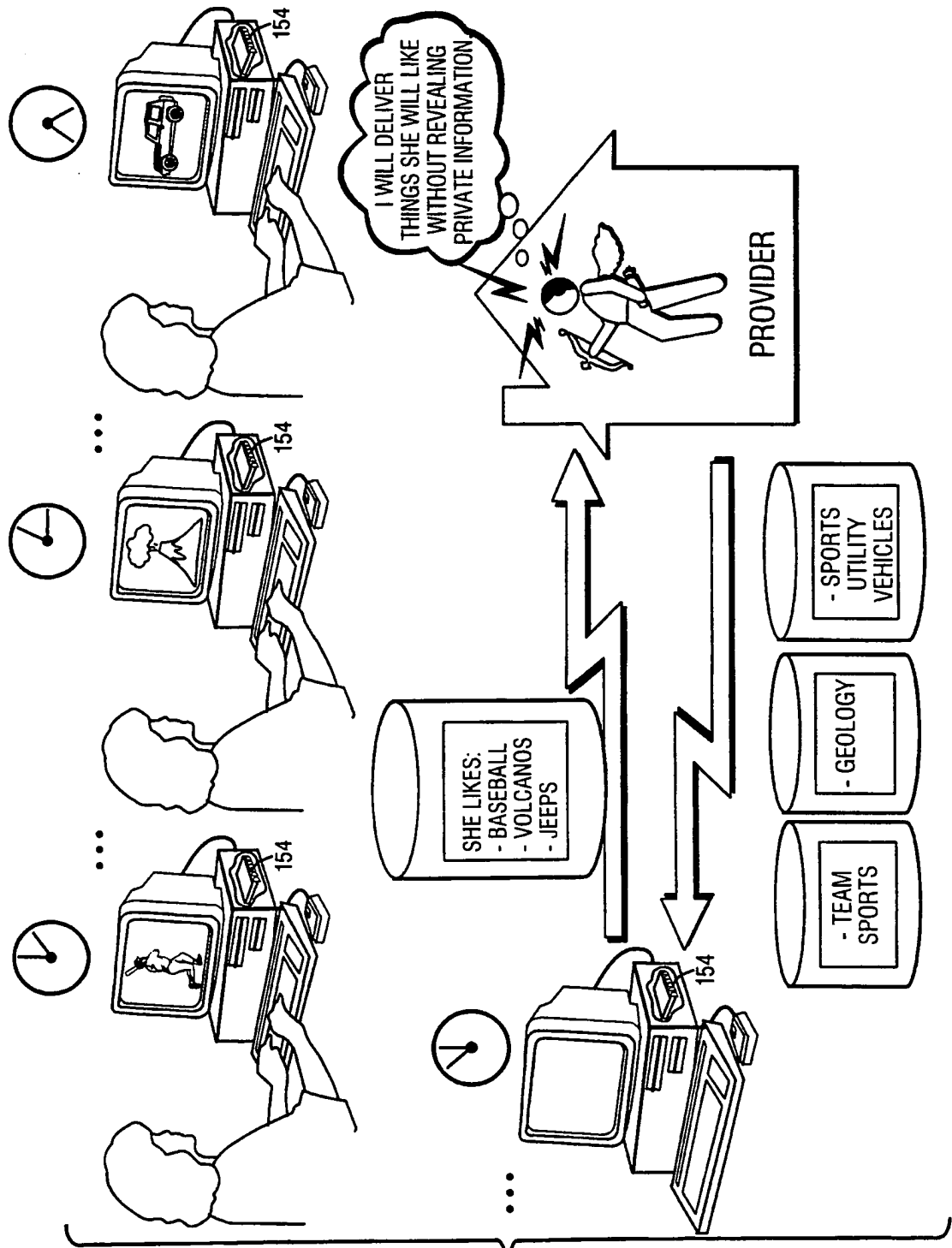
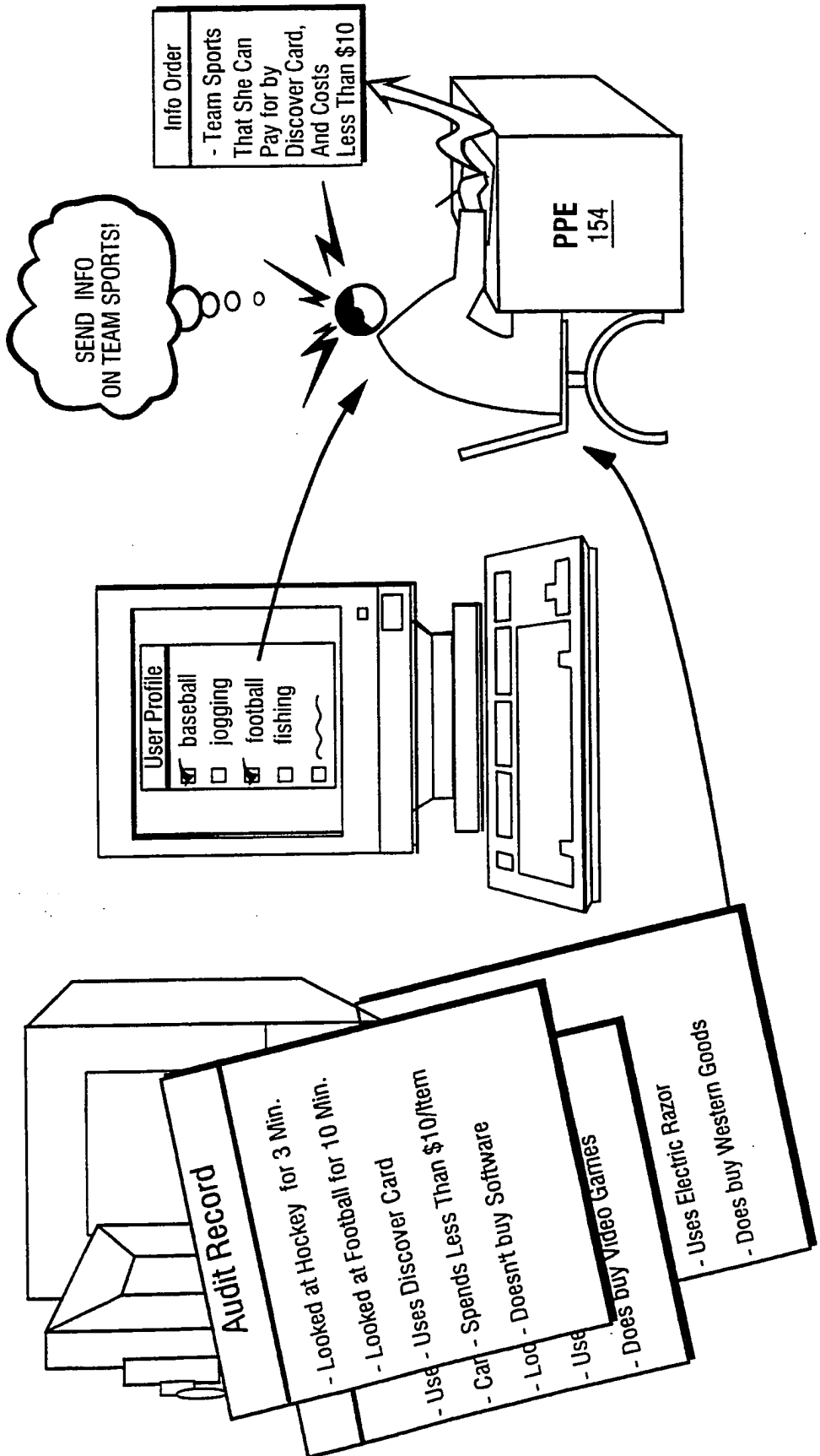


Fig. 10
Example Electronic
Matchmaking Process

Fig. 11 Example User Rights Management Information
By Electronic Matchmaker



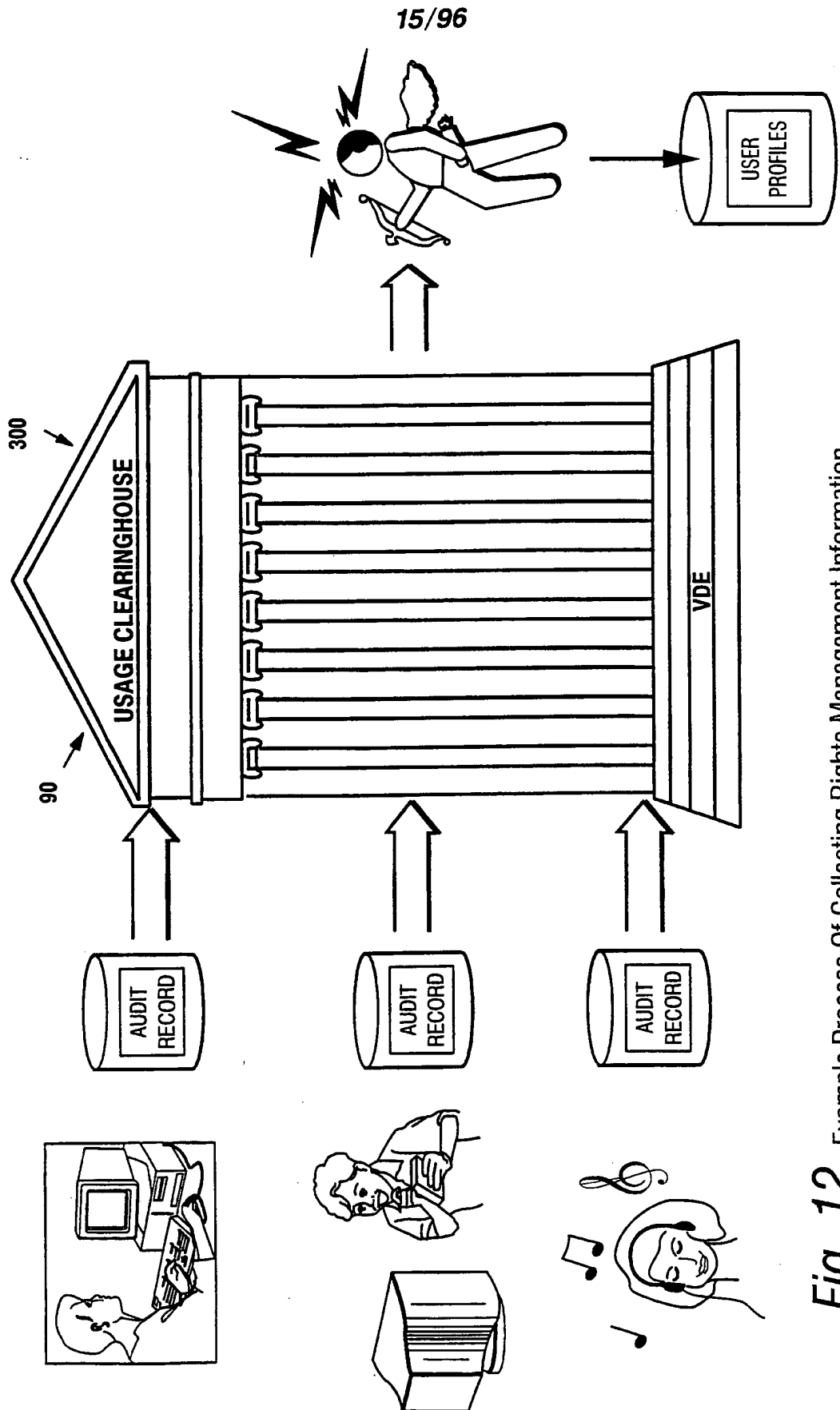


Fig. 12 Example Process Of Collecting Rights Management Information

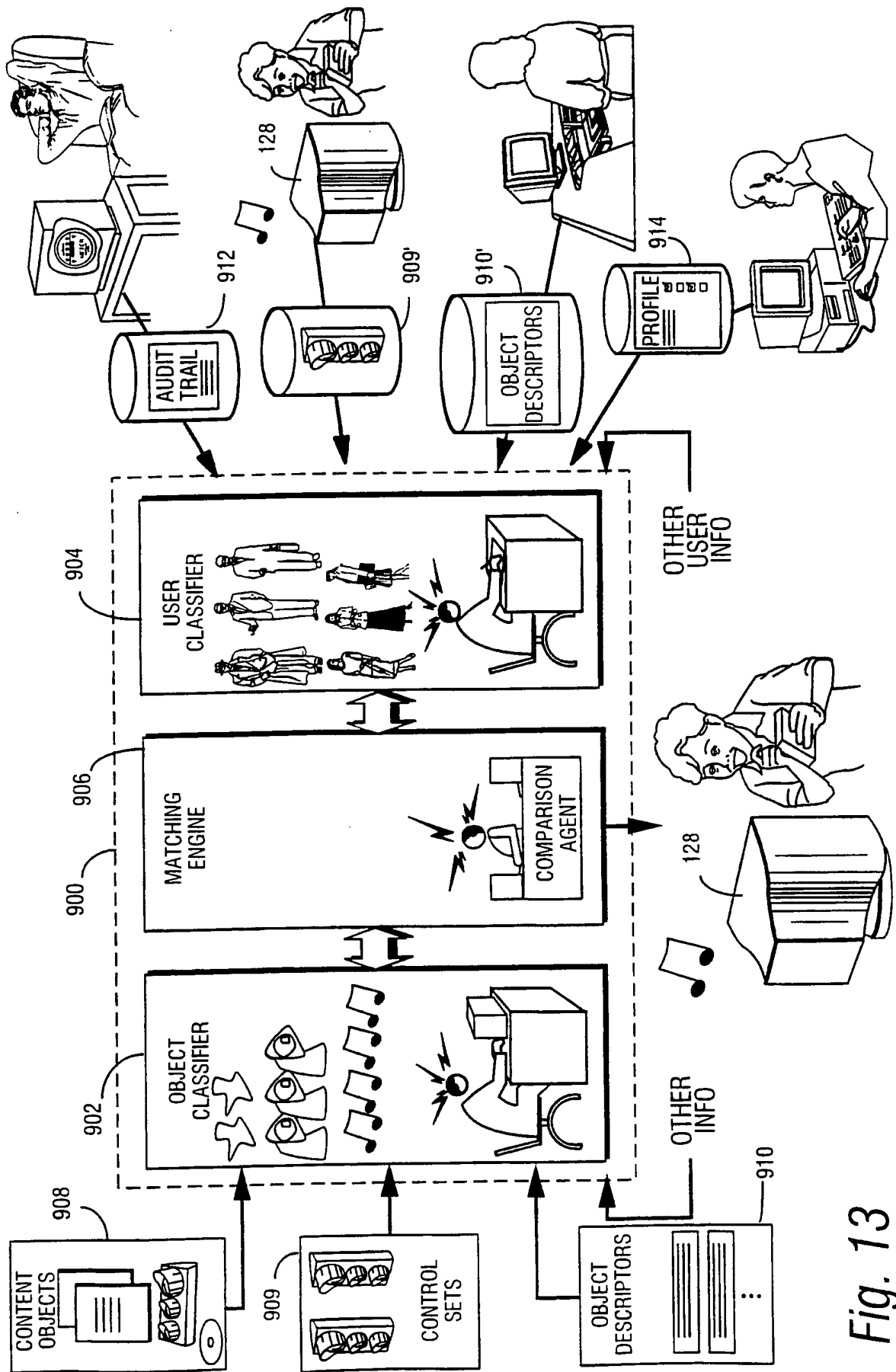


Fig. 13

SUBSTITUTE SHEET (RULE 26)

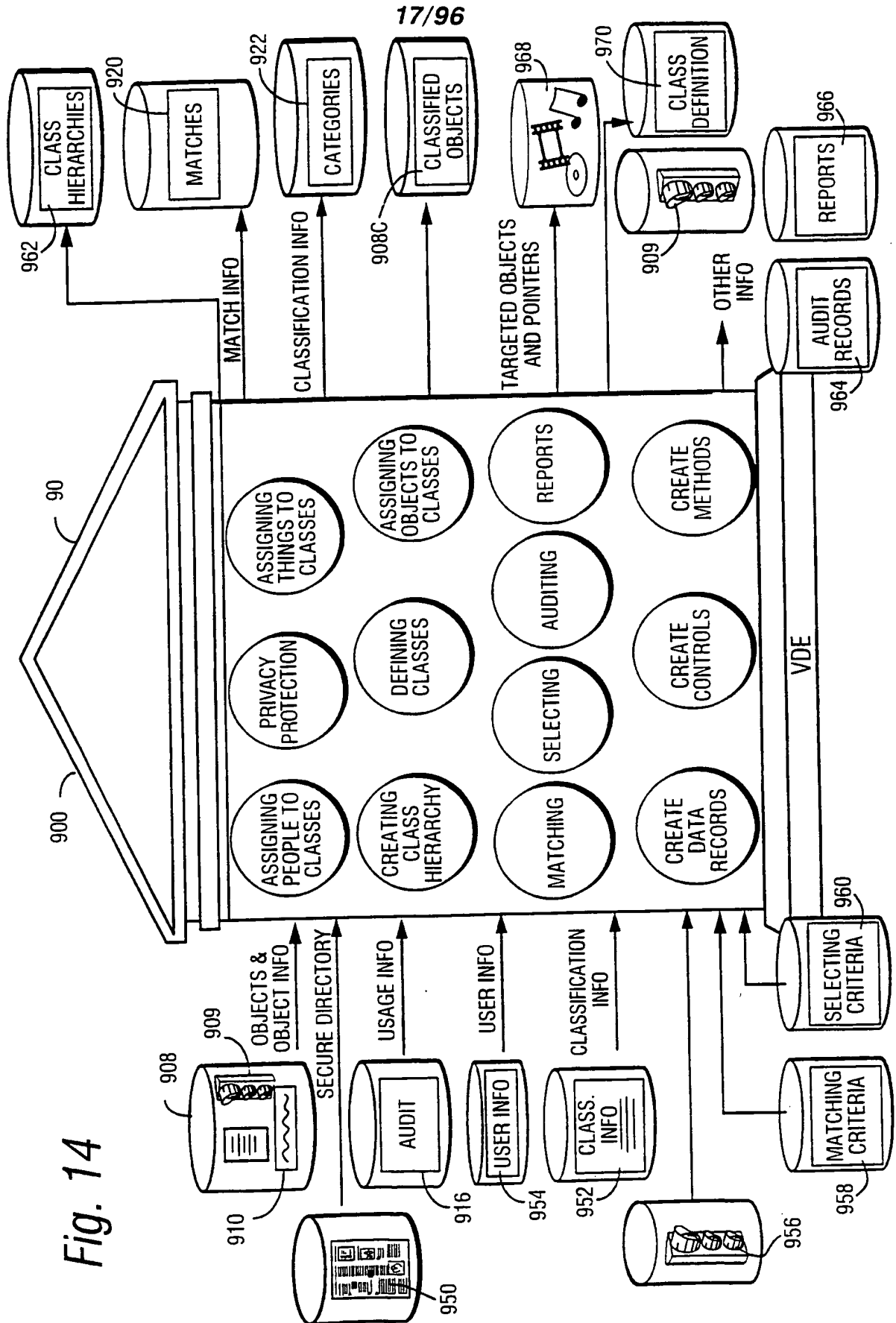
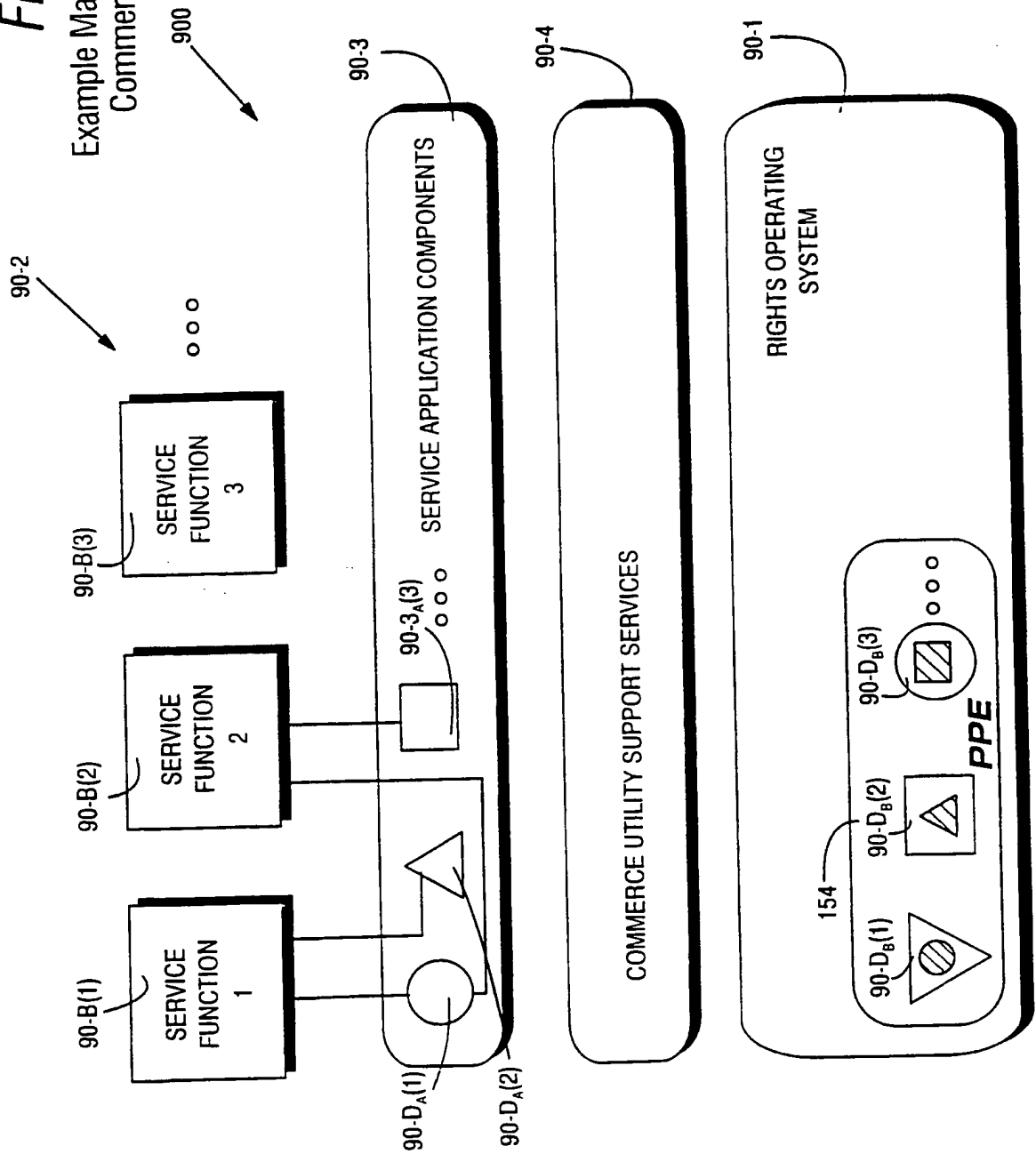


Fig. 14

Fig. 14(A)
Example Matching and Classification
Commerce Utility System 900



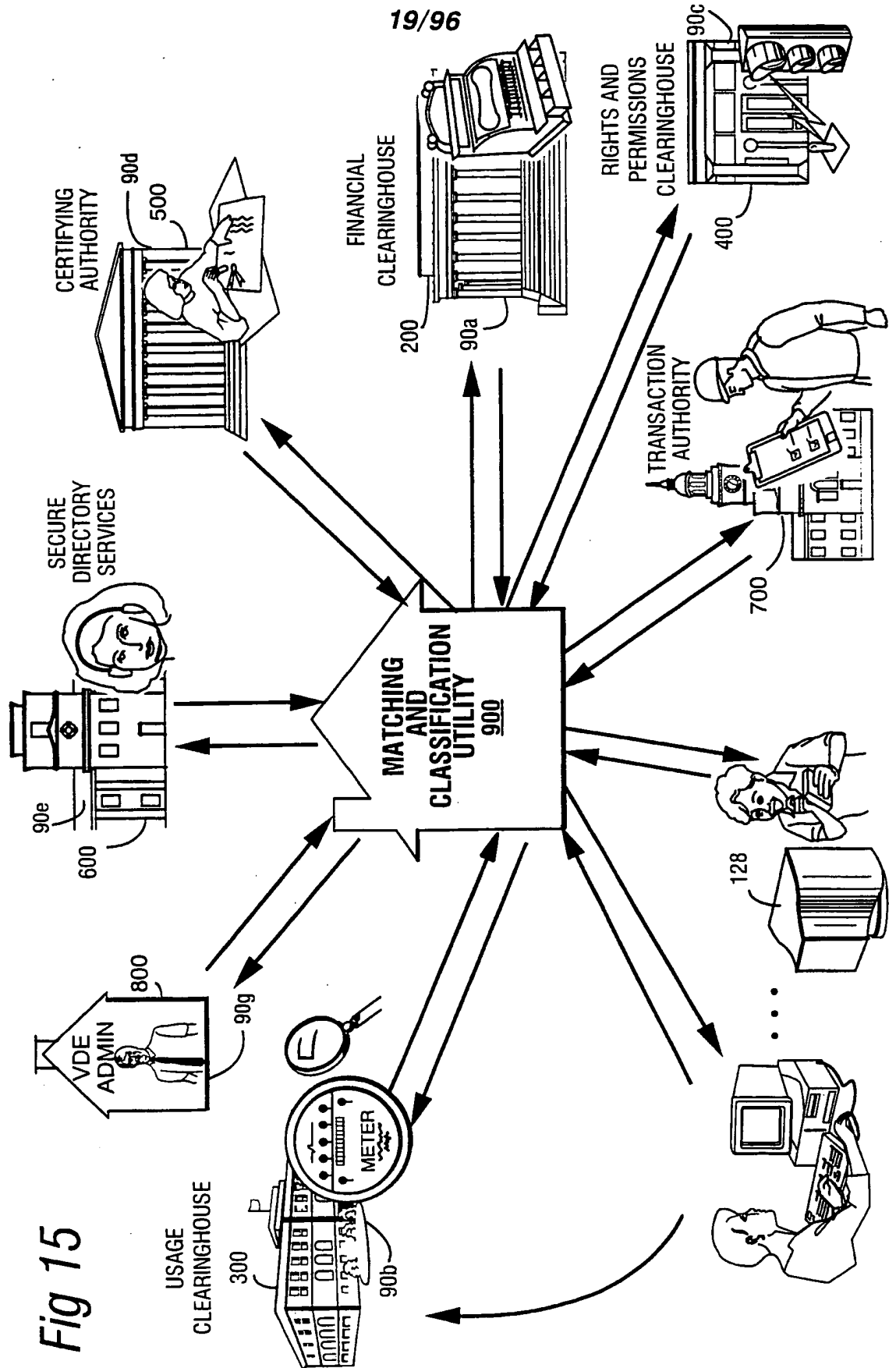


Fig 15

20/96

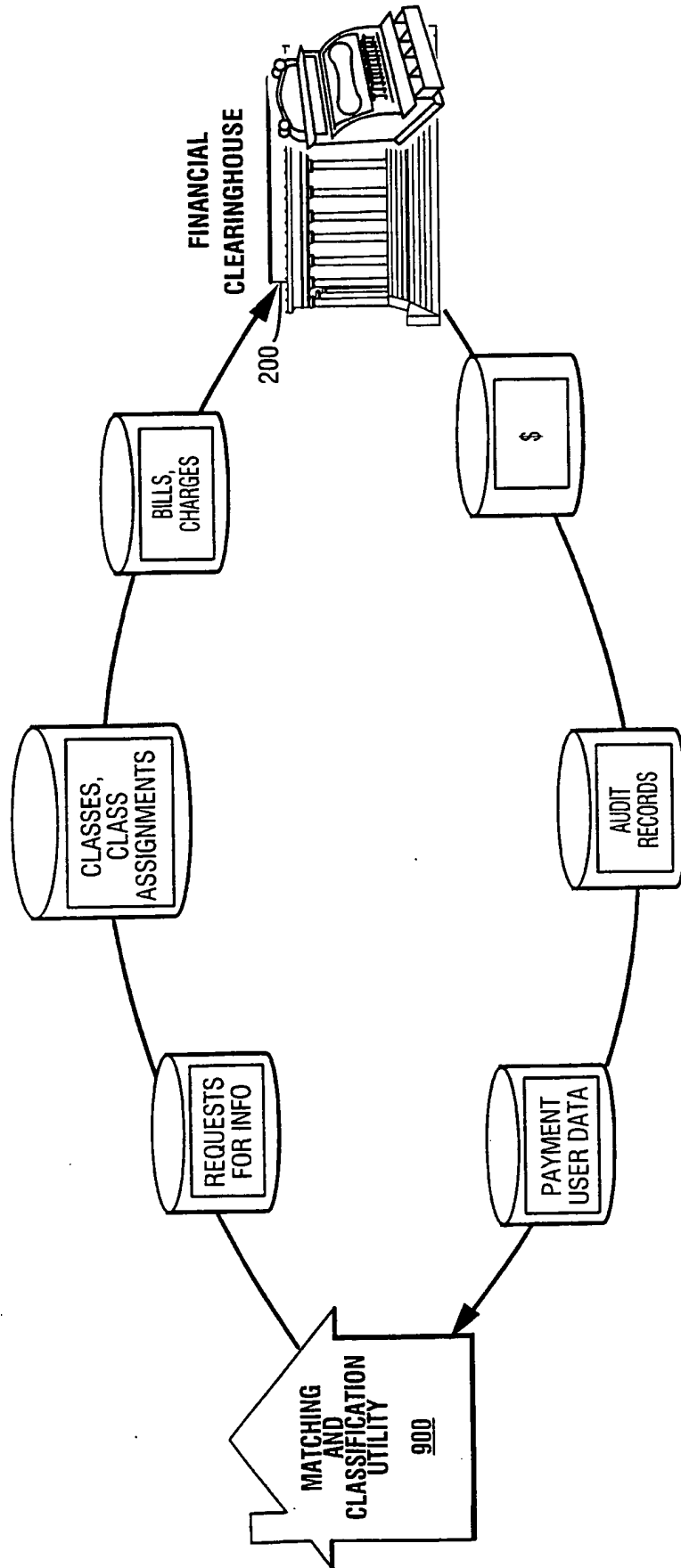


Fig. 15A

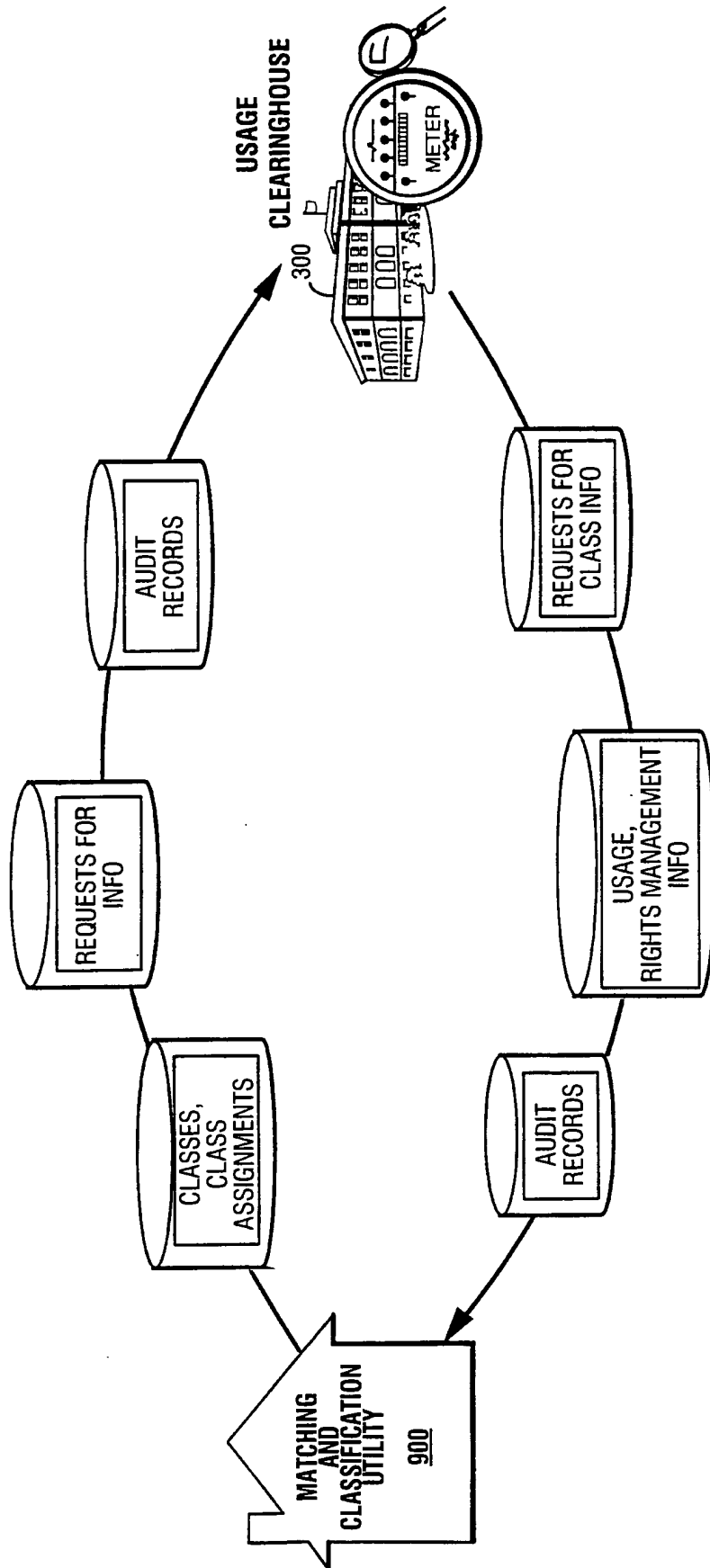


Fig. 15B

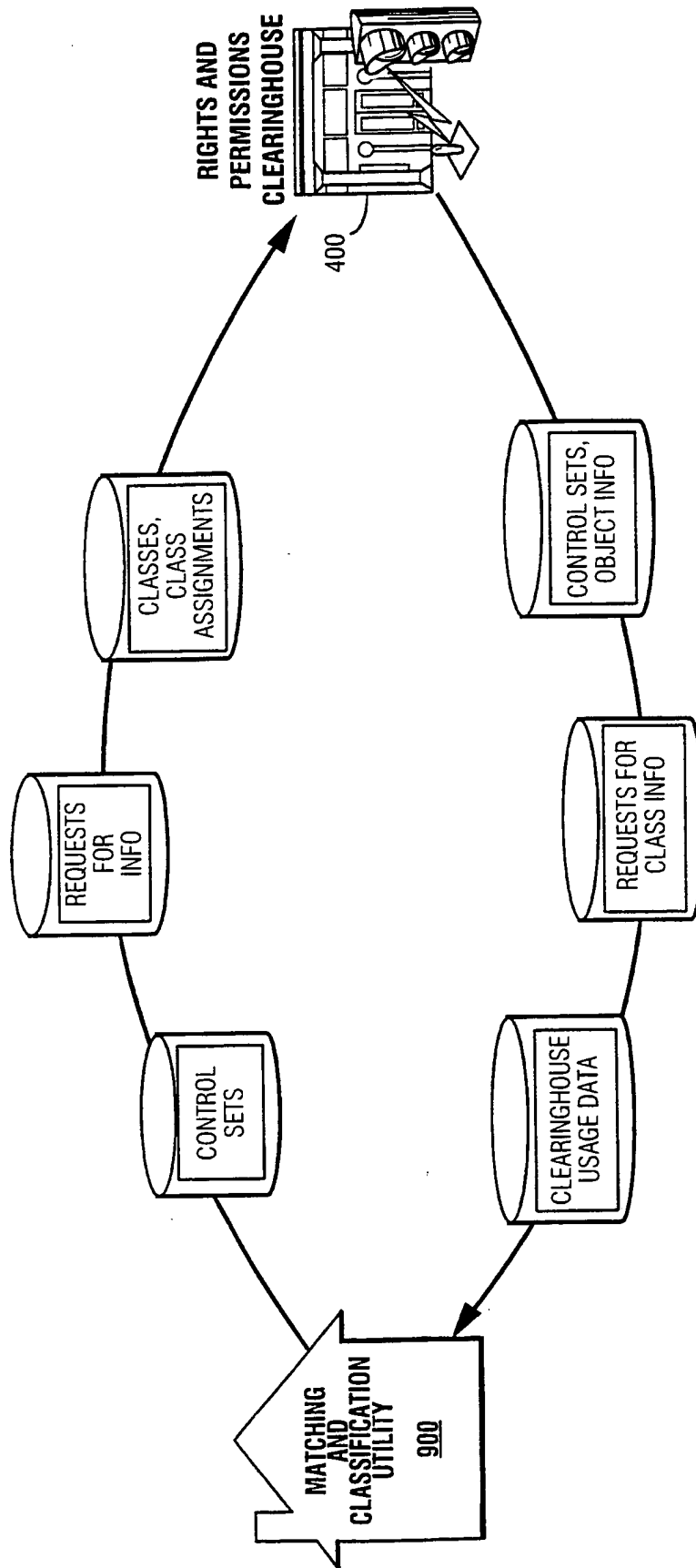
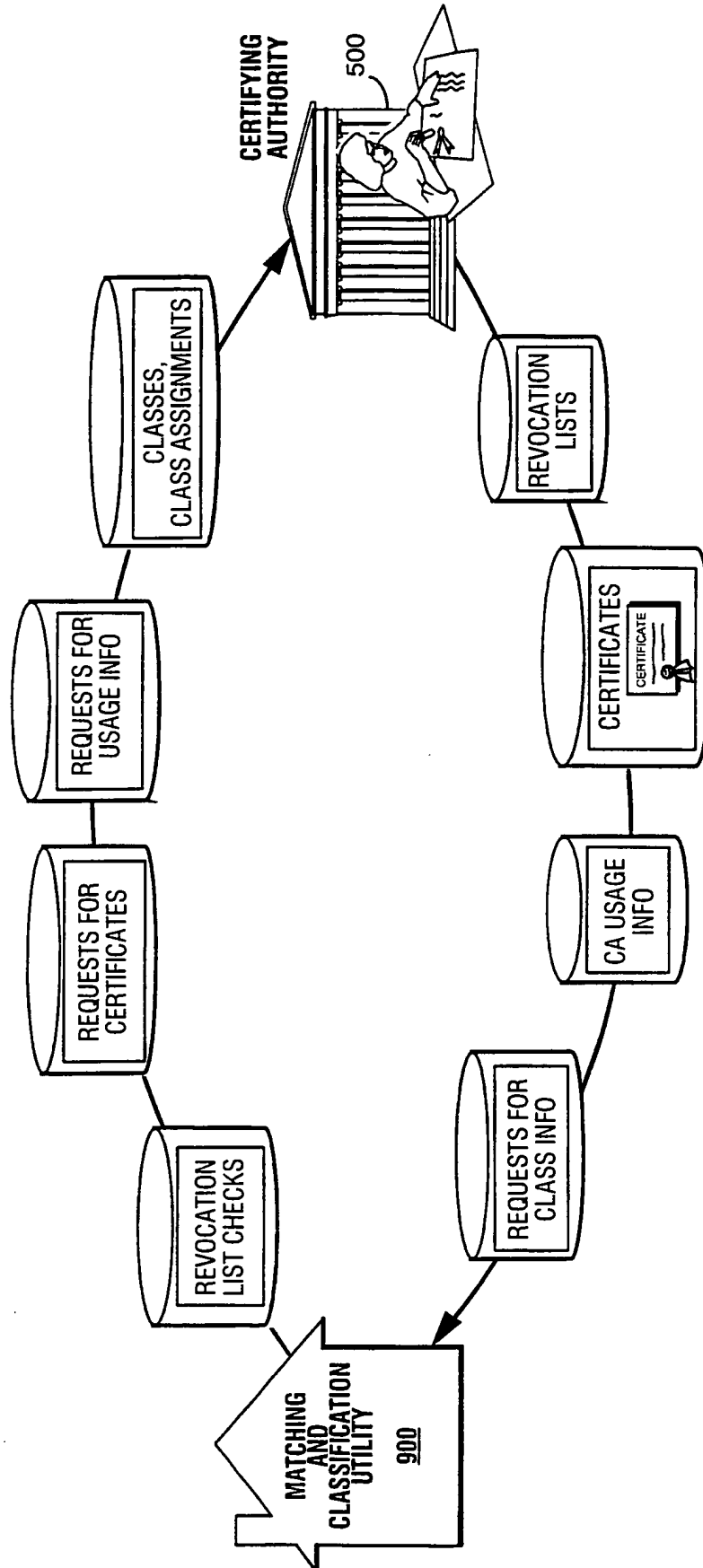


Fig. 15C

Fig. 15D



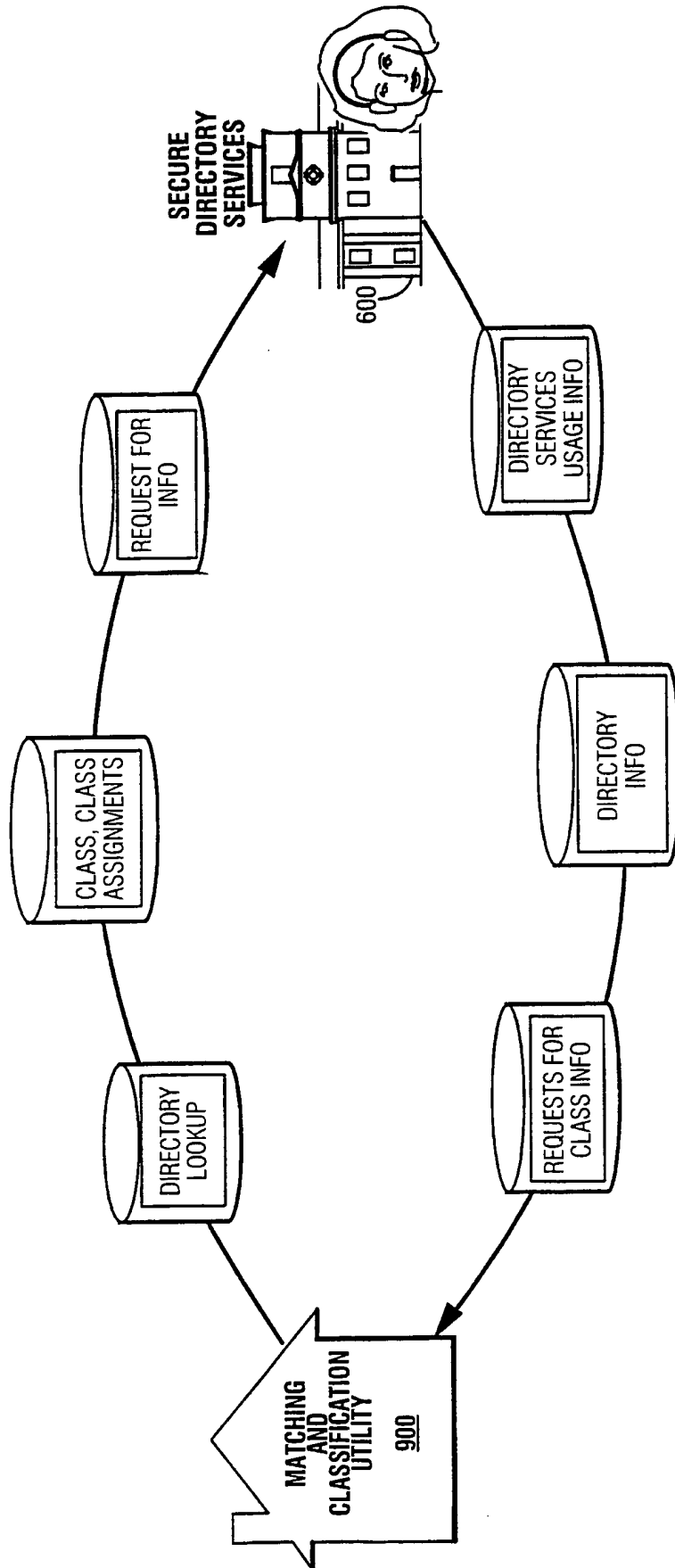


Fig. 15E

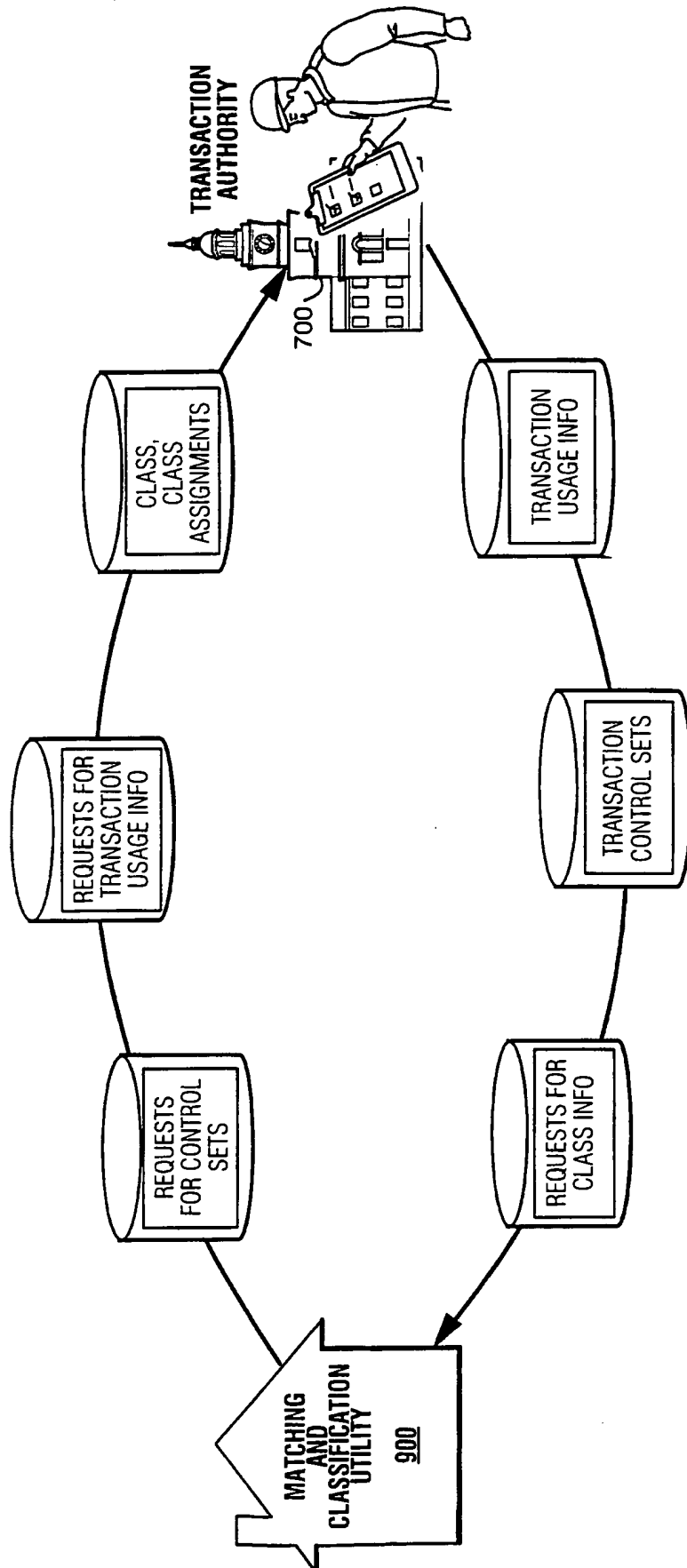


Fig. 15F

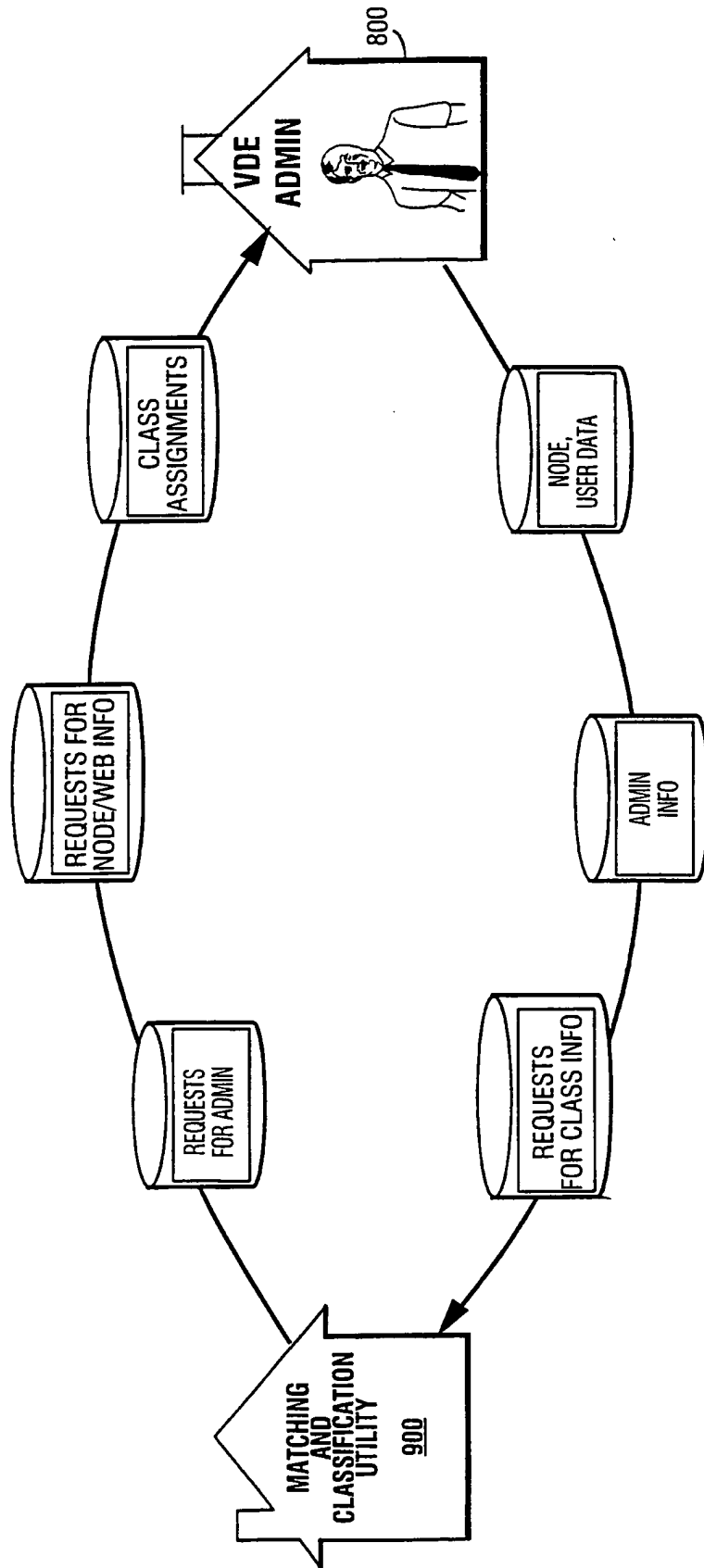
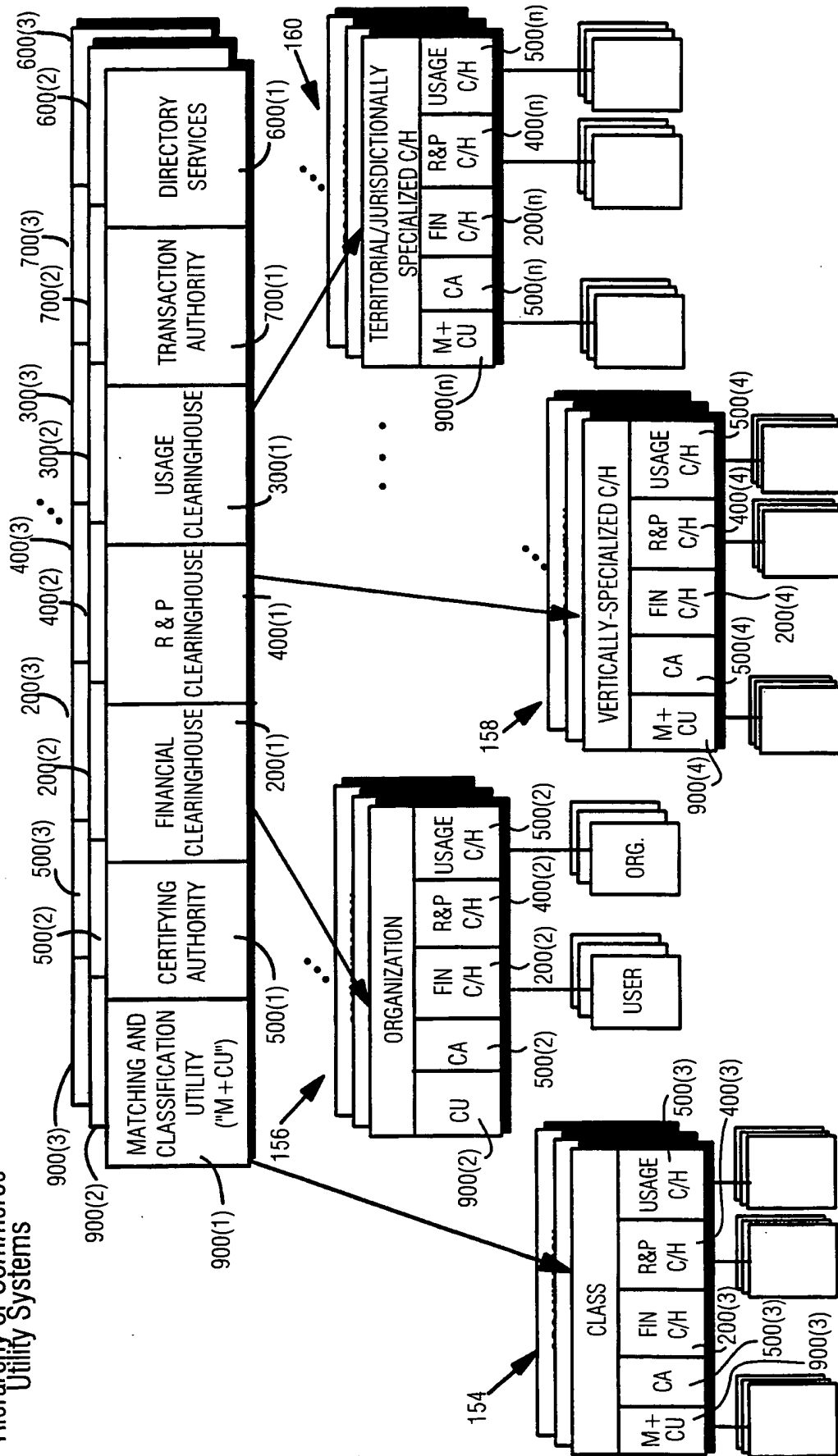


Fig. 15G

Fig. 16A
Hierarchy of Commerce
Utility Systems



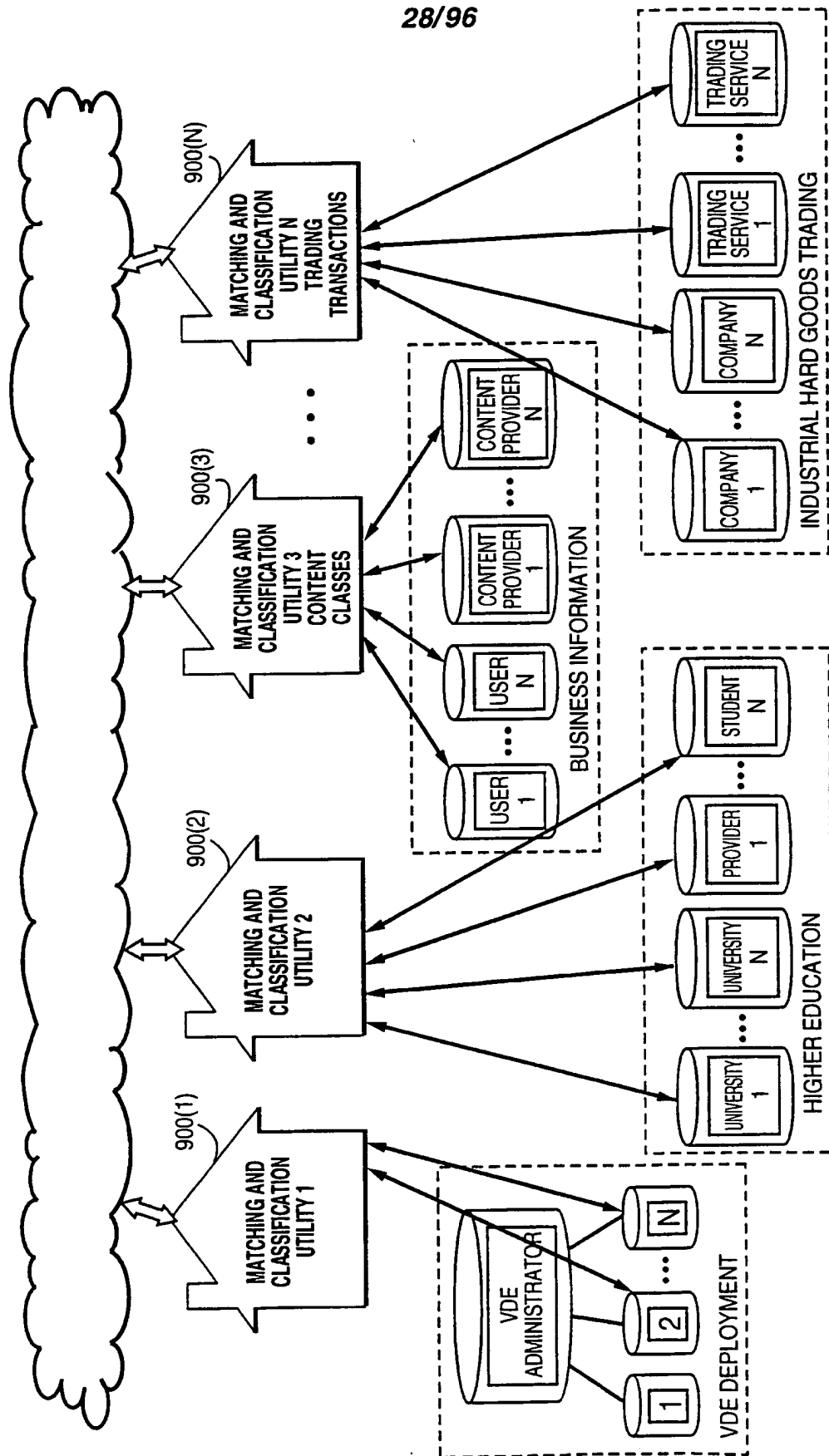
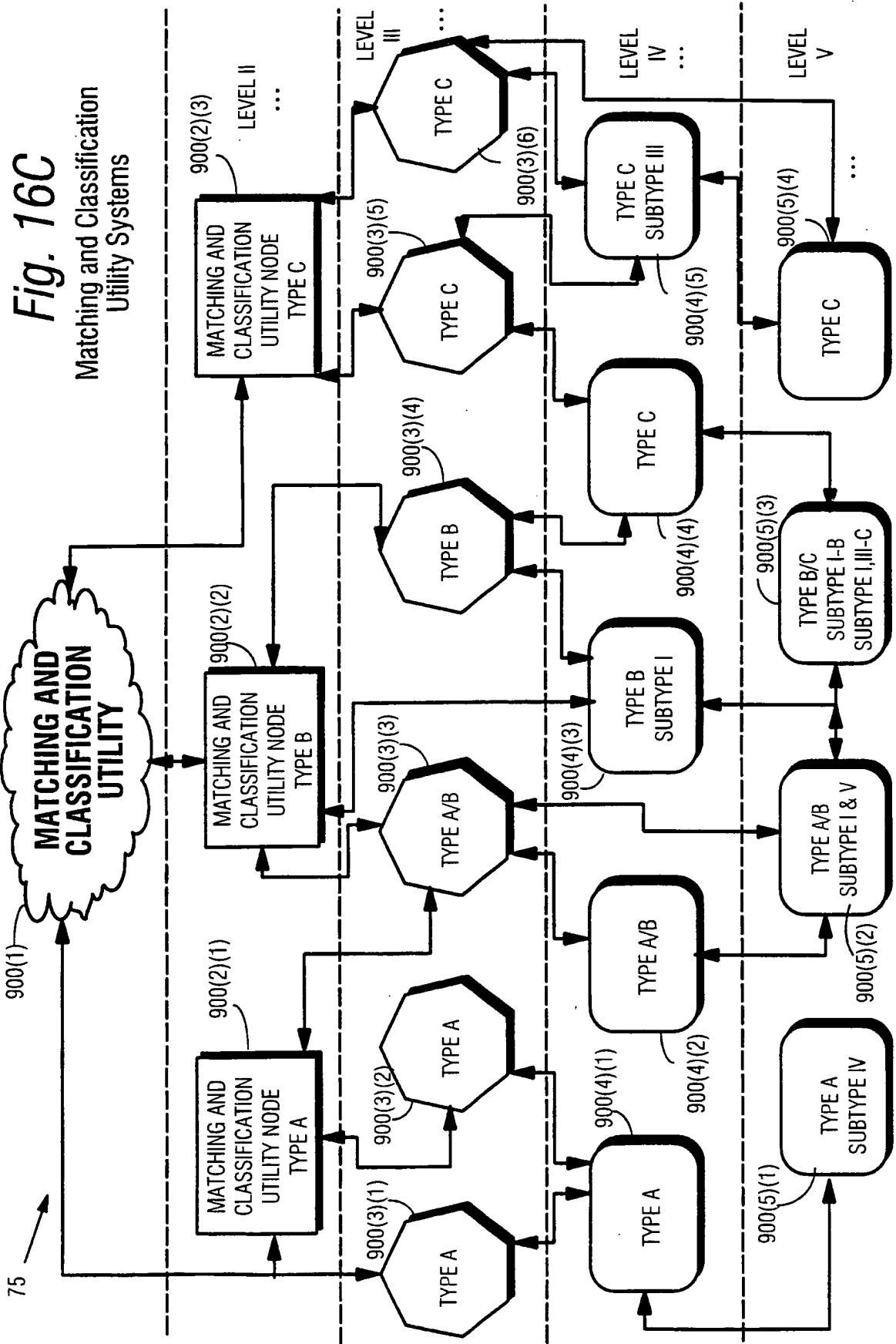


Fig. 16B Matching & Classification Utilities Provide Services To Classes Of Nodes, Users, Content Services, Transaction Services.



90B

Fig. 17

FINANCIAL CLEARINGHOUSE	USAGE CLEARINGHOUSE	RIGHTS & PERMISSIONS CLEARINGHOUSE	CERTIFICATE AUTHORITY	SECURE DIRECTORY SERVICES	TANGIBLES PURCHASE & FULFILLMENT	INTANGIBLES PURCHASE & FULFILLMENT	CONTRACT NEGOTIATIONS & EXECUTION	EDI	SECURE DOCUMENT DELIVERY	BUSINESS PROCESS INTEGRATION	ARBITRATION & MEDIATION	ELECTRONIC ORDERS	ELECTRONIC BANKING & CURRENCY MANAGEMENT	CYBERSPACE TRADING ENVIRONMENTS	CLASSIFICATION UTILITY
AUDIT BY CLASS		MAINTAINING RECORDS	STATUS NOTIFICATION	EVENT DATABASE MANAGEMENT	CONTROL SET DATABASE MGMT	NOTARY	OBJECT REGISTRY	CERTIFICATE CREATION
OVERSEEING PROCESS	CONFIRMATIONS	ROUTING DATABASE	GENERATE CONTROL SETS	SEAL GENERATOR	OBJECT IDENTIFIER ASSIGNMENT	REVOCACTION LIST MAINTENANCE
MONITORING STATUS	UNCOMPLETED EVENTS RECORD	GENERATING REQUESTS	CONTROL LOGIC	DIGITAL TIME STAMP	COPYRIGHT REGISTRATION
COMPLETE PROCESS DEFINITION	REQUIREMENTS GENERATION	REPLICATION	EVENT FLOW GENERATION	FINGERPRINT /WATERMARK	CONTROL SET REGISTRY
PROCESS CONTROL	REPORT GENERATION	PROPAGATION	ROUTING	OFFERS & COUNTER OFFERS	TEMPLATE REGISTRY	DIRECTOR DATABASE MANAGEMENT
INTERFACE(S) TO SETTLEMENT SERVICES	FUNDS TRANSFER	EVENT CONSEQUENCES	USAGE DATABASE MANAGEMENT	ARCHIVE	DATABASE QUERY & RESPONSE PROCESSING
CURRENCY CONVERSION	TAX CALCULATION & APPLICATION	ACCOUNT RECONCILIATION	BILL CREATION & PROCESSING	RIGHTS & PERMISSION DATABASE MANAGEMENT	ADVERTISING DATABASE MANAGEMENT
ACCOUNT CREATION & IDENTIFIER ASSIGNMENT	PAYMENT AGGREGATION	IDENTITY AUTHENTICATION	MARKET RESEARCH	TEMPLATE DATABASE MANAGEMENT	AUTOMATIC CLASS GENERATION	AUTOMATIC CLASS MATCHING
PAYMENT DISAGGREGATION	BUDGET PRE-AUTHORIZATION	ELECTRONIC CURRENCY CREATION	NEGOTIATION	COMMERCE MGMT LANGUAGE PROCESSING	AUTOMATIC CLASS ASSIGNMENT	CLASS BASED SEARCHING
...	RIGHTS MANAGEMENT LANGUAGE PROCESSING	CLASS BASED DIRECTORY

31/96

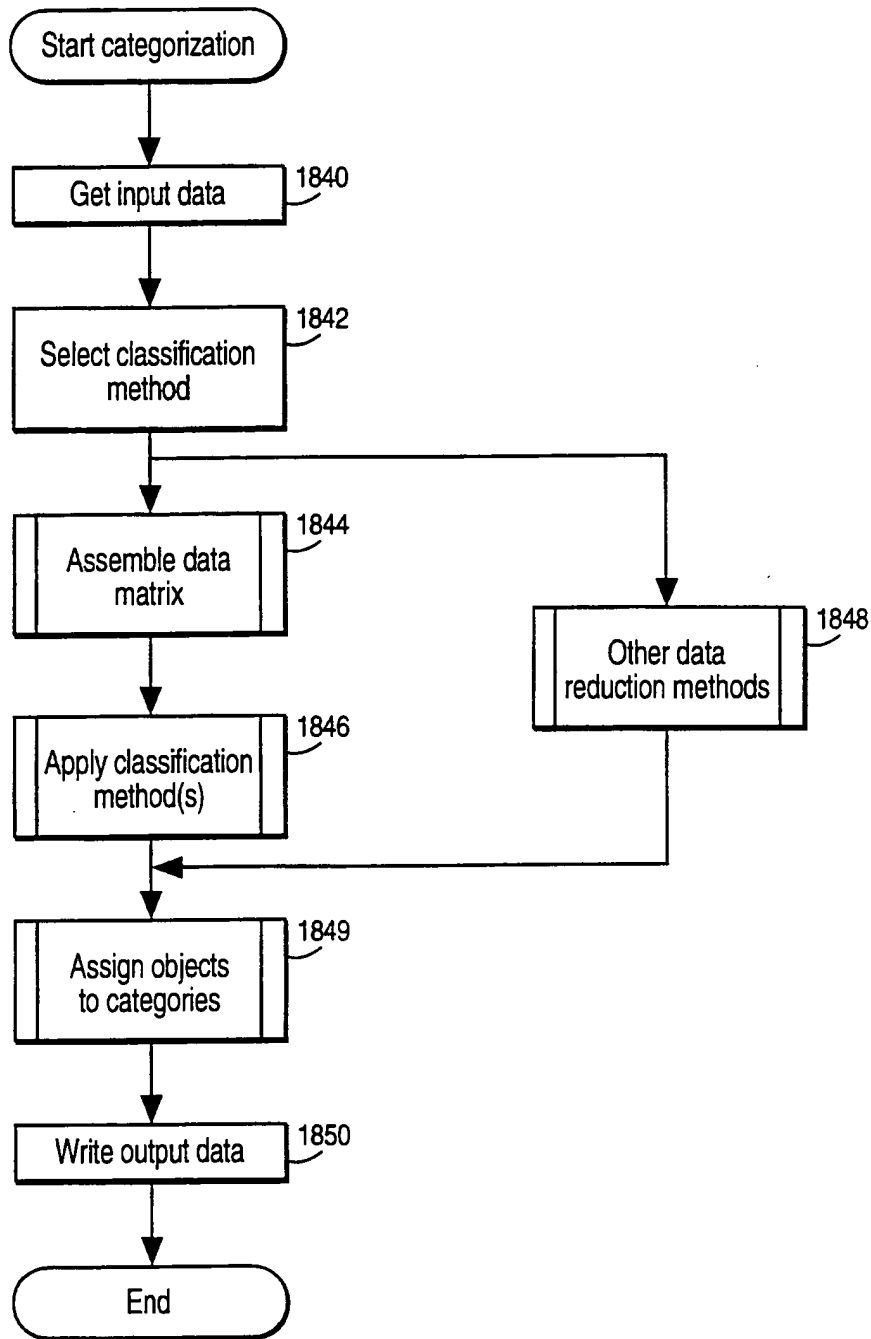


Fig. 18

Example Steps to Categorize Objects

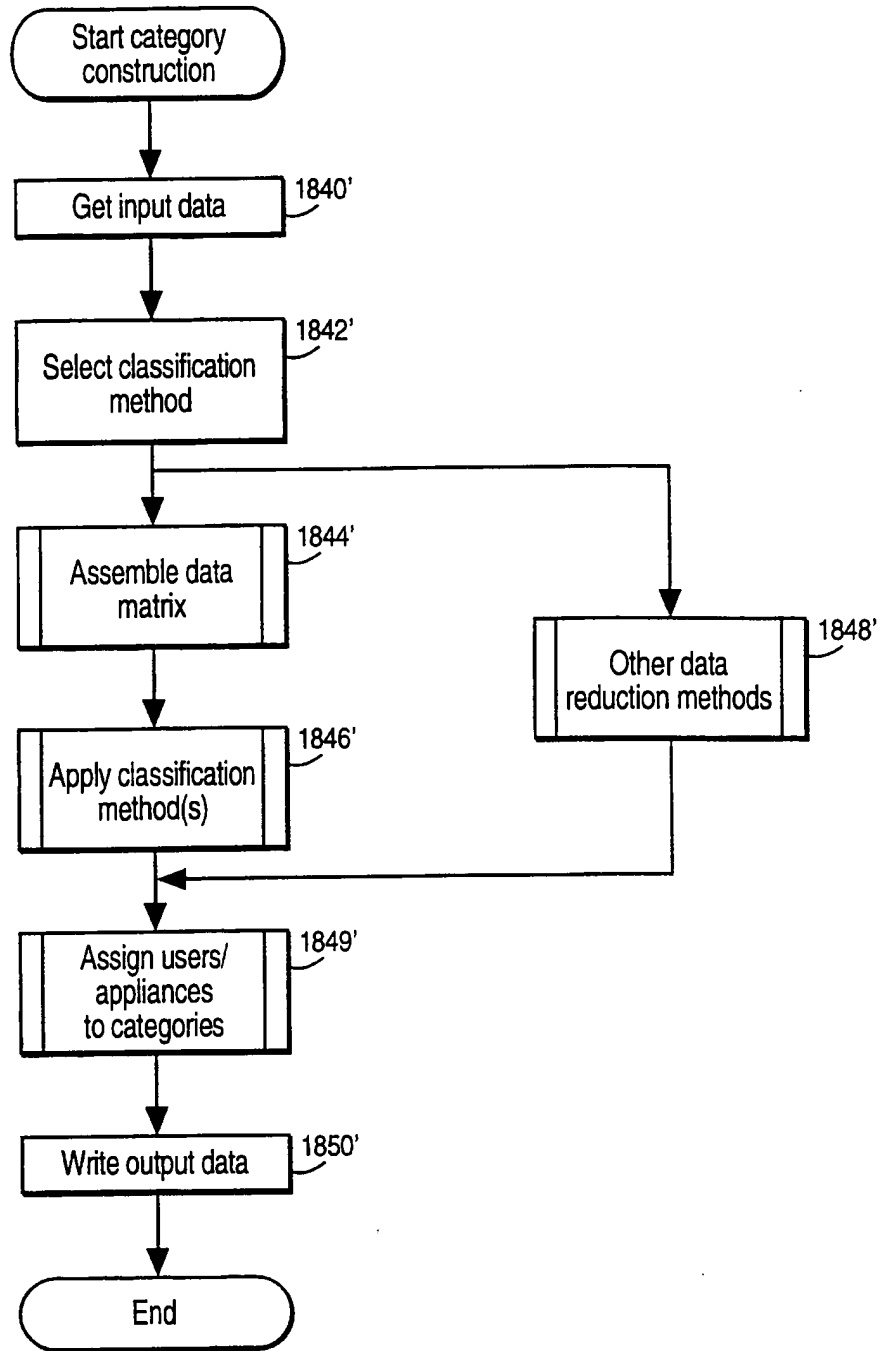


Fig. 19

Example Steps to Categorize Users/Appliances

Node ID	Operating system	Country	State	VDE Adm. Org.	VDE version	VDE maintenance level	User ID number	Gender	Age	Highest edu. level	Citizenship residence	Country of residence	City
128.1.4.132	WIN95	USA	CA	VDEADM	1.5	02	FF98C48A	Female	32	14	UK	UK	London

1852

User ID	Myers-Briggs Categories				SRI internet iVALS category
	Extroversion or introversion	Sensing or intuition	Thinking or feeling	Judging or perceiving	
FF98C48A	I	N	T	J	Worker

Fig. 20
 Example Composite Record-Input To Classification Process

User ID number	Object ID	Right ID	Method	Right ID	Method	Right ID	Method
CF129CD5	1227-33-1298-2	Use	Open	Meter	Each time	Budget	Simple purchase
						Bill	\$1.00
							VISA

34/96

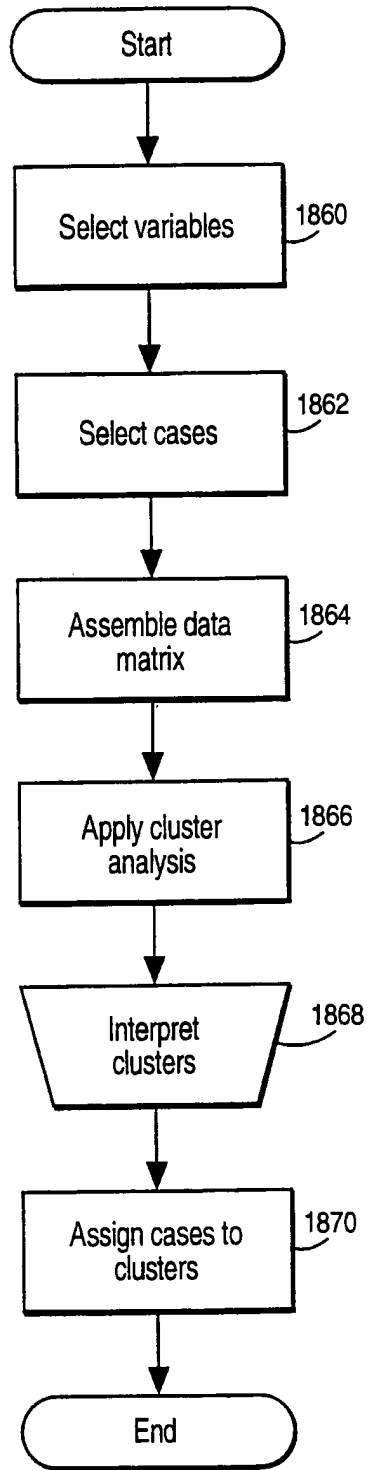


Fig. 21

Example Cluster Analysis Process

35/96

Variables	Typical Class 1-Profile	Typical Class 2-Profile
City	Washington, DC	Knoxville, TN
Av. price of content purchased last 30 days	\$8.79	\$1.95
Number of trips abroad in last 2 years	3	0
Type of content most frequently purchased	National and international news	Sports
2nd most frequently purchased	Business information	Religious
Third most frequently purchased	Travel information	Movies
Pay per view	No	Yes
Add new controls to content	Yes	No
Stated religious affiliation	None	Methodist
SRI internet lifestyle category	Surfer	Worker
Modification rights purchased	20% of text items	5% of text items

Fig. 22 Example Classification Output Illustrating Different Classes Based Upon Differing Profiles

36/96

Variables	Factor 1 Loadings	Factor 2 Loadings
Region of US	.82	.11
Family income	.90	-.09
Av. price of content purchased last 30 days	.72	.15
Number of trips abroad in last 2 years	.91	.09
Percent news, business	.79	-.12
Percent entertainment	-.69	.21
Add new controls to content	.88	.19
Religiosity	-.60	-.22
Participates in sports	-.21	.87
Watches team/individual sports on TV	-.11	.62
Owns a sports utility vehicle	.12	.72
Consumes beer/wine	-.18	.83
Male/female	.21	.92
Education beyond college	.45	-.45
Buys pay per view sports events	-.25	.77
Number of TVs in house	-.11	.66

Fig. 23 Example Classification Output Illustrating Principal Components Analysis On Parameter Data And Categories Data

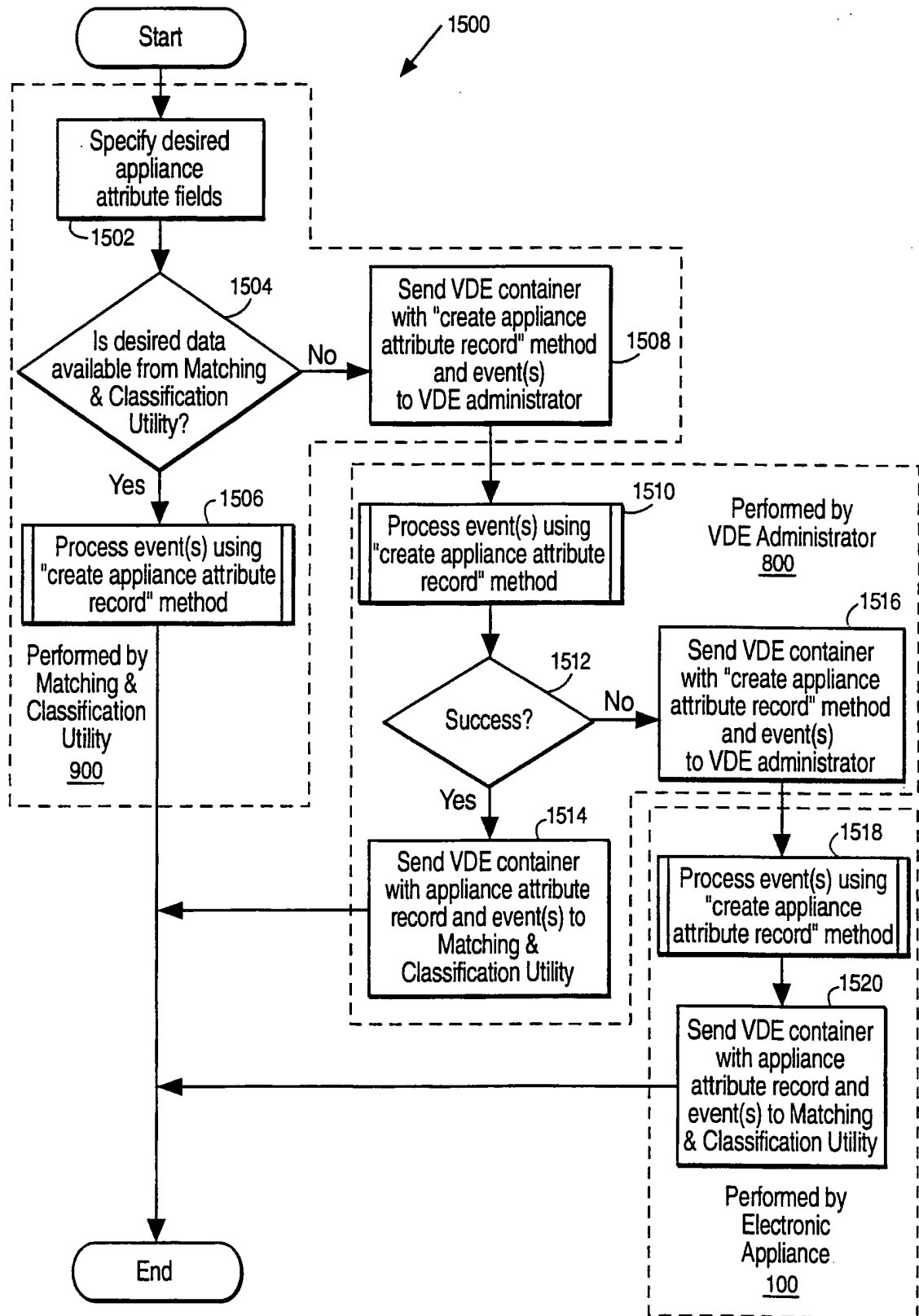


Fig. 24

Example Steps for Collecting Appliance Attribute Data

SUBSTITUTE SHEET (RULE 26)

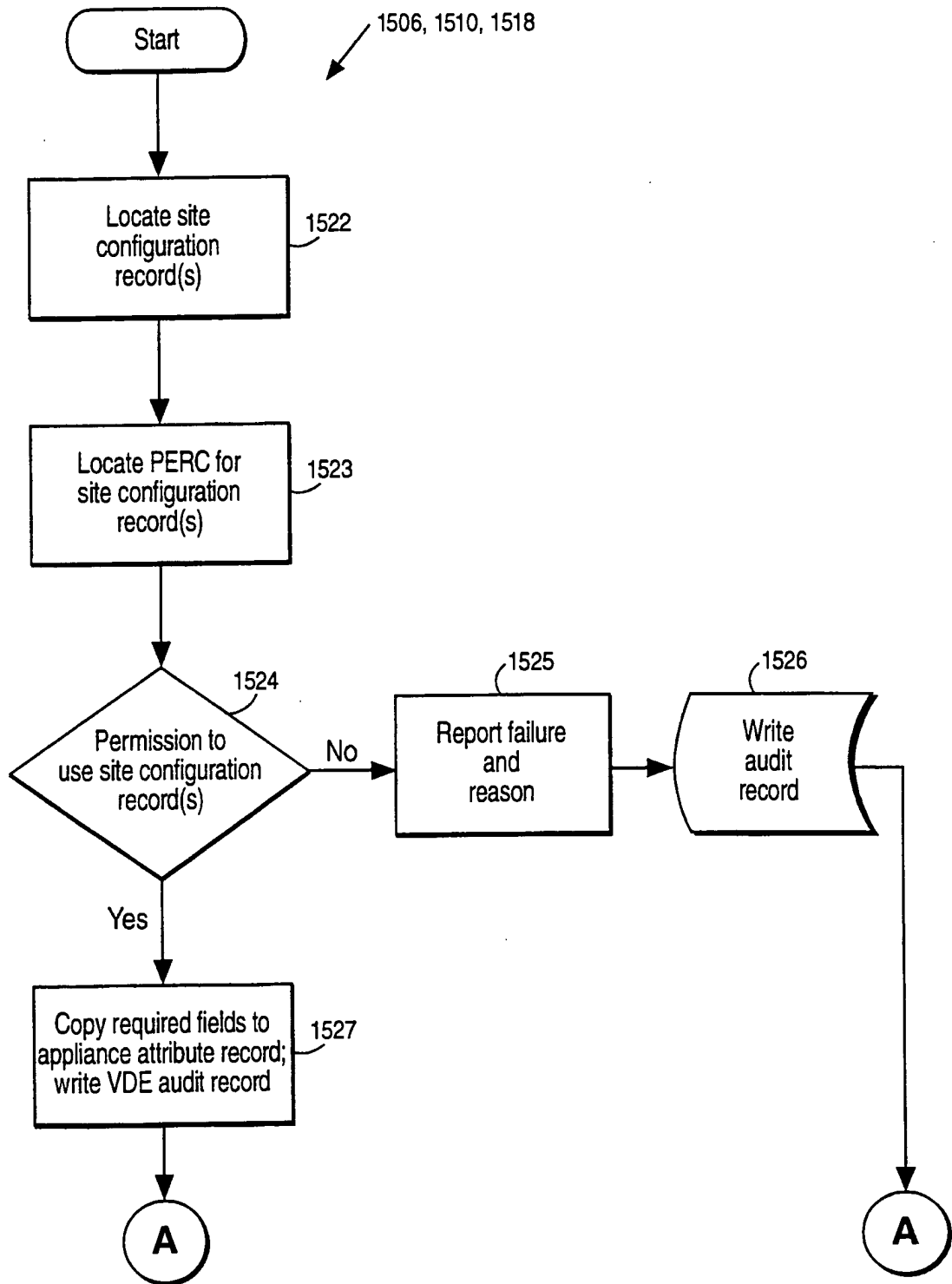


Fig. 25(A)

Example Create Appliance Attribute Data Method steps

Fig. 25(B)

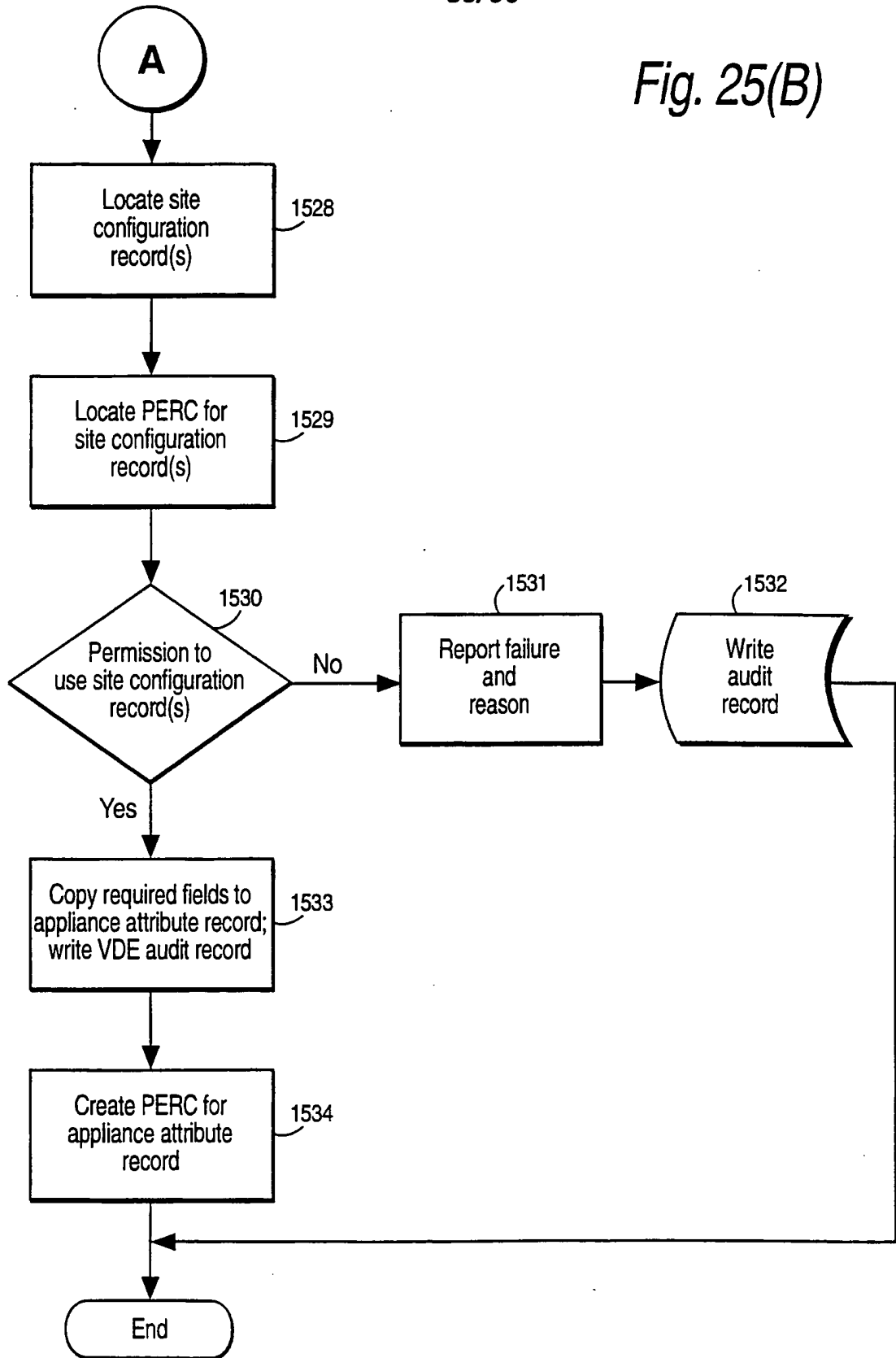


Fig. 26(A) Example Appliance Attribute Record

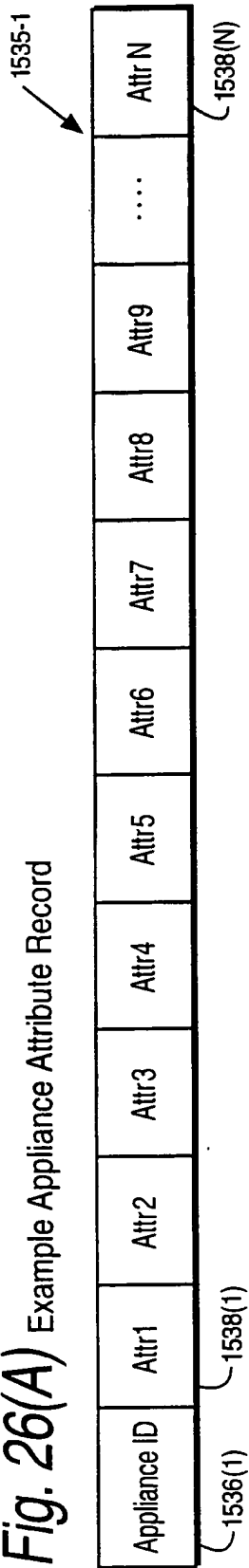


Fig. 26(B)

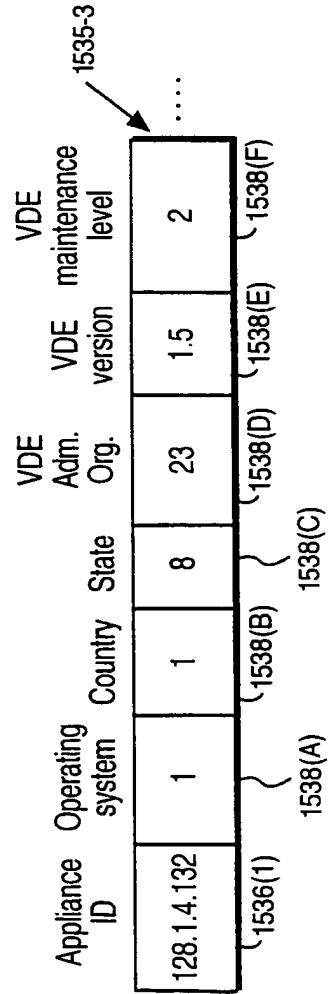
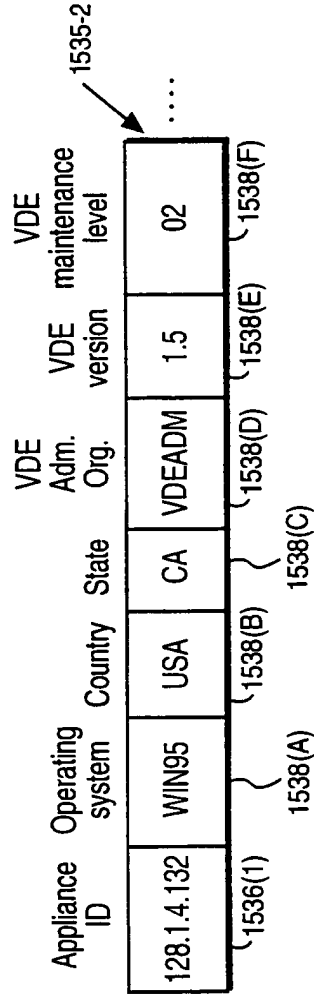


Fig. 26(C)
Example Appliance Attribute Record

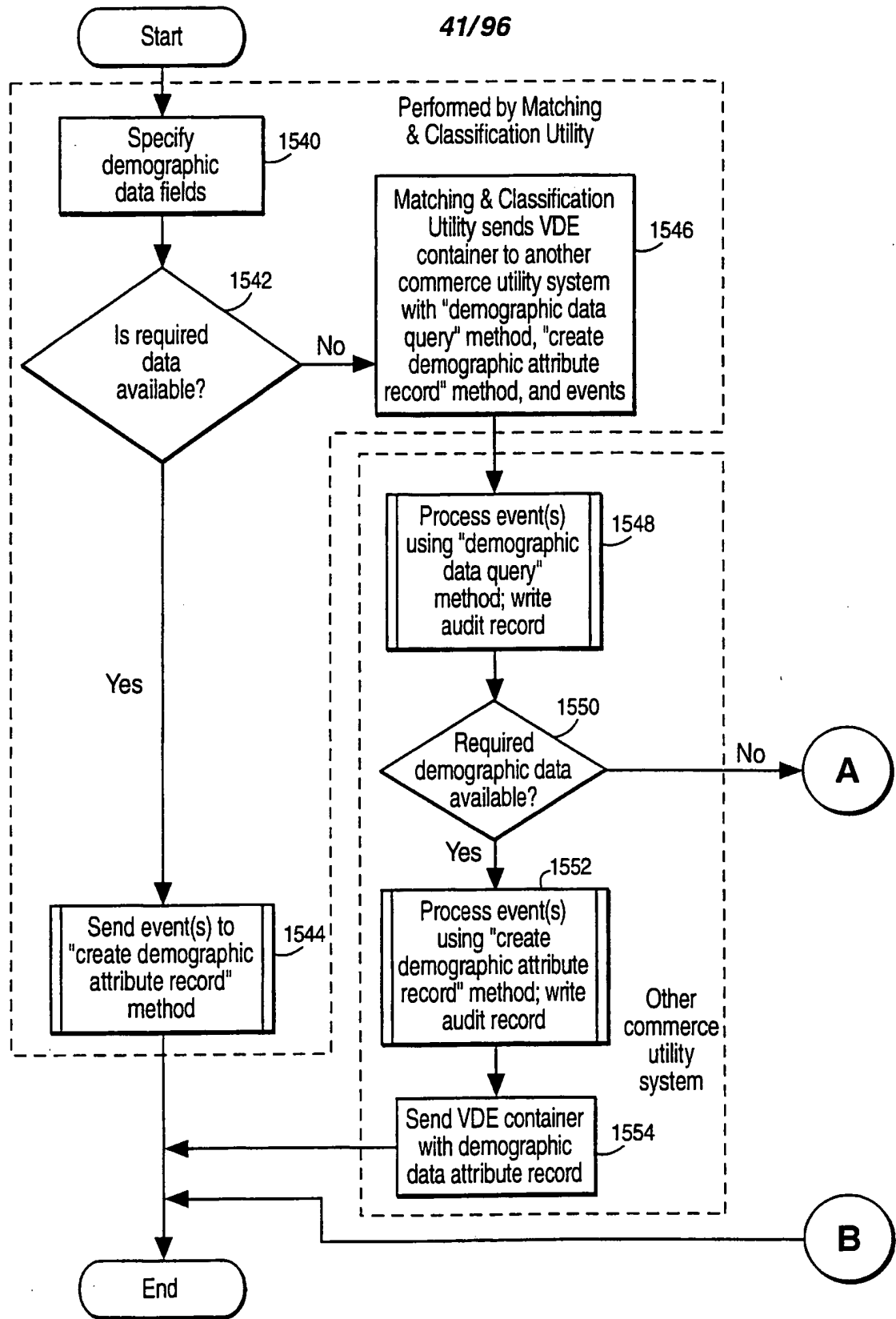


Fig. 27(A) Example Steps for Collecting Demographic Data

42/96

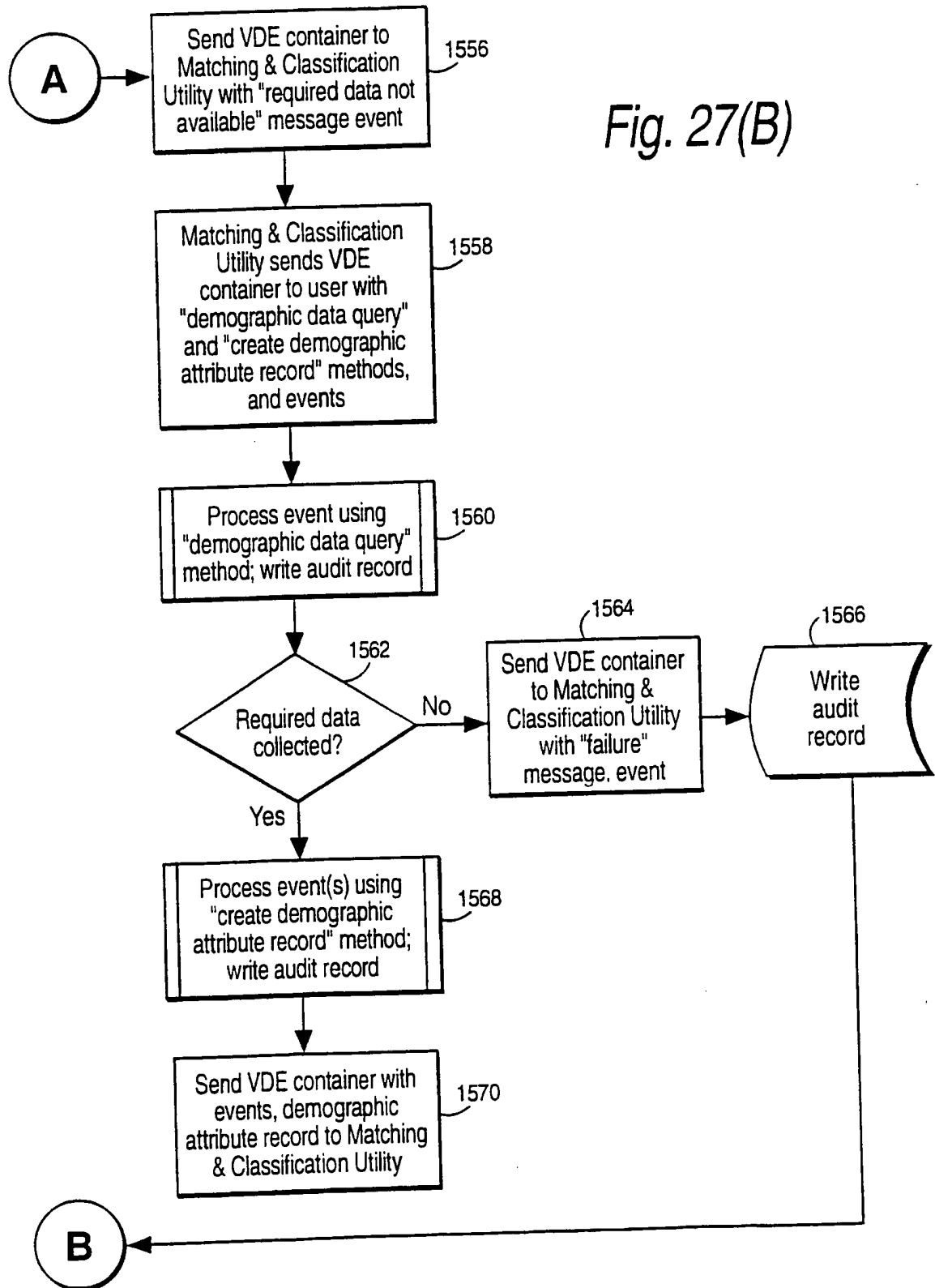


Fig. 27(B)

Demographic Information Questionnaire

Name: _____

Address: _____

Address: _____

City: _____ State: _____ Zip: _____ - _____

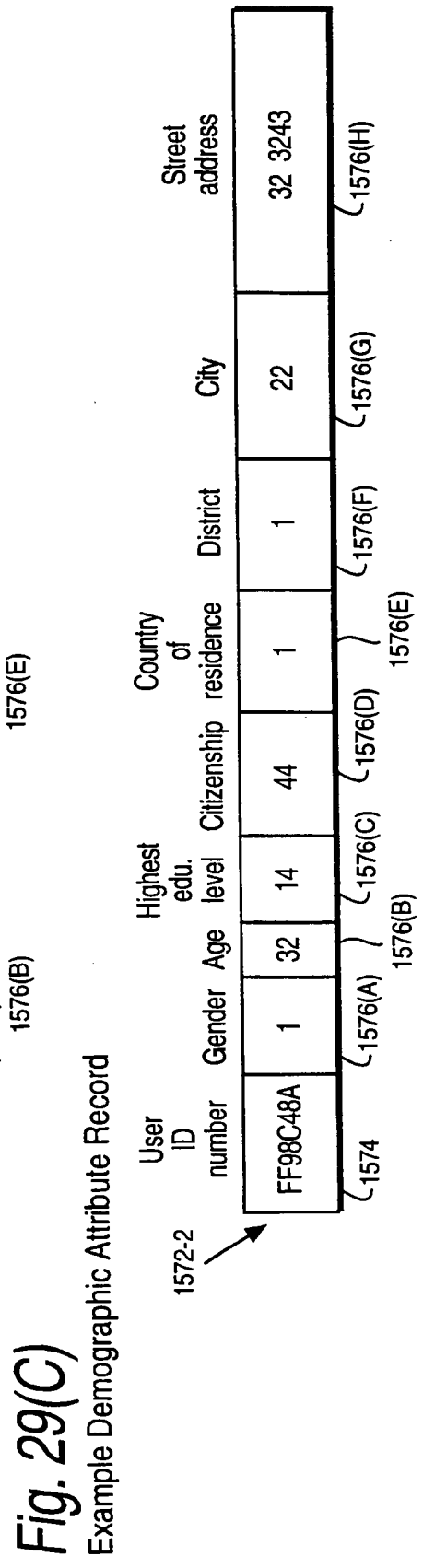
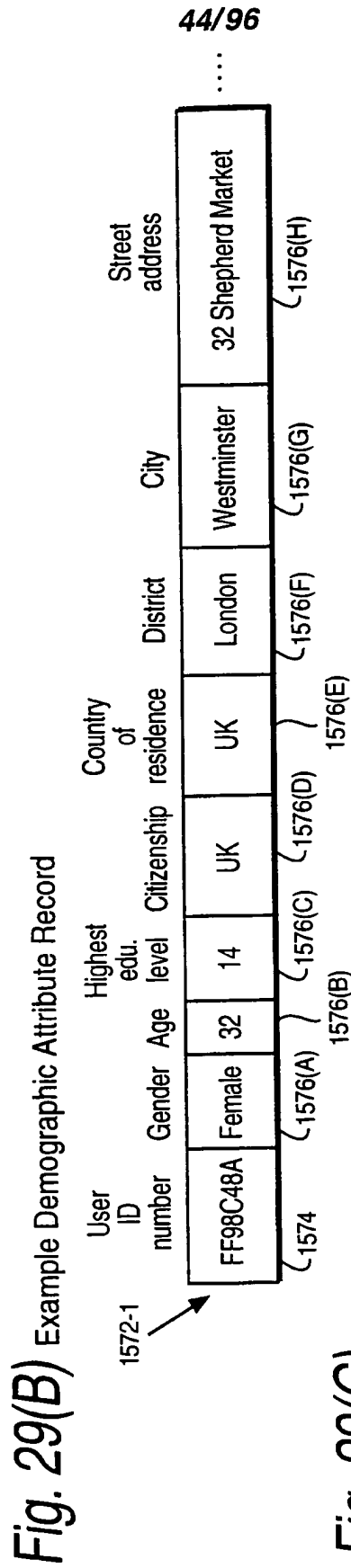
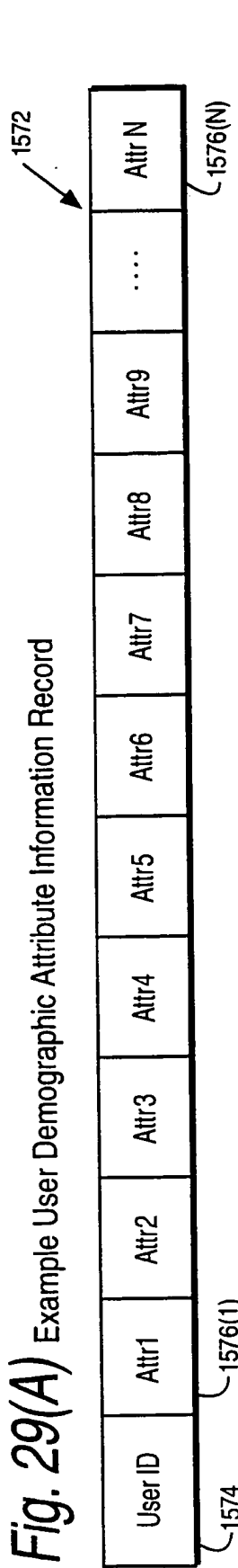
Gender (M/F) _____ Date of birth: _____ / _____ / _____

Education:

- Have not graduated high school
- High school graduate
- Some college
- College degree
- Some graduate school
- Advanced degree

All Information Will Be Treated As Confidential

Fig. 28 Example Demographic Questionnaire "Pop-Up" Screen



45/96

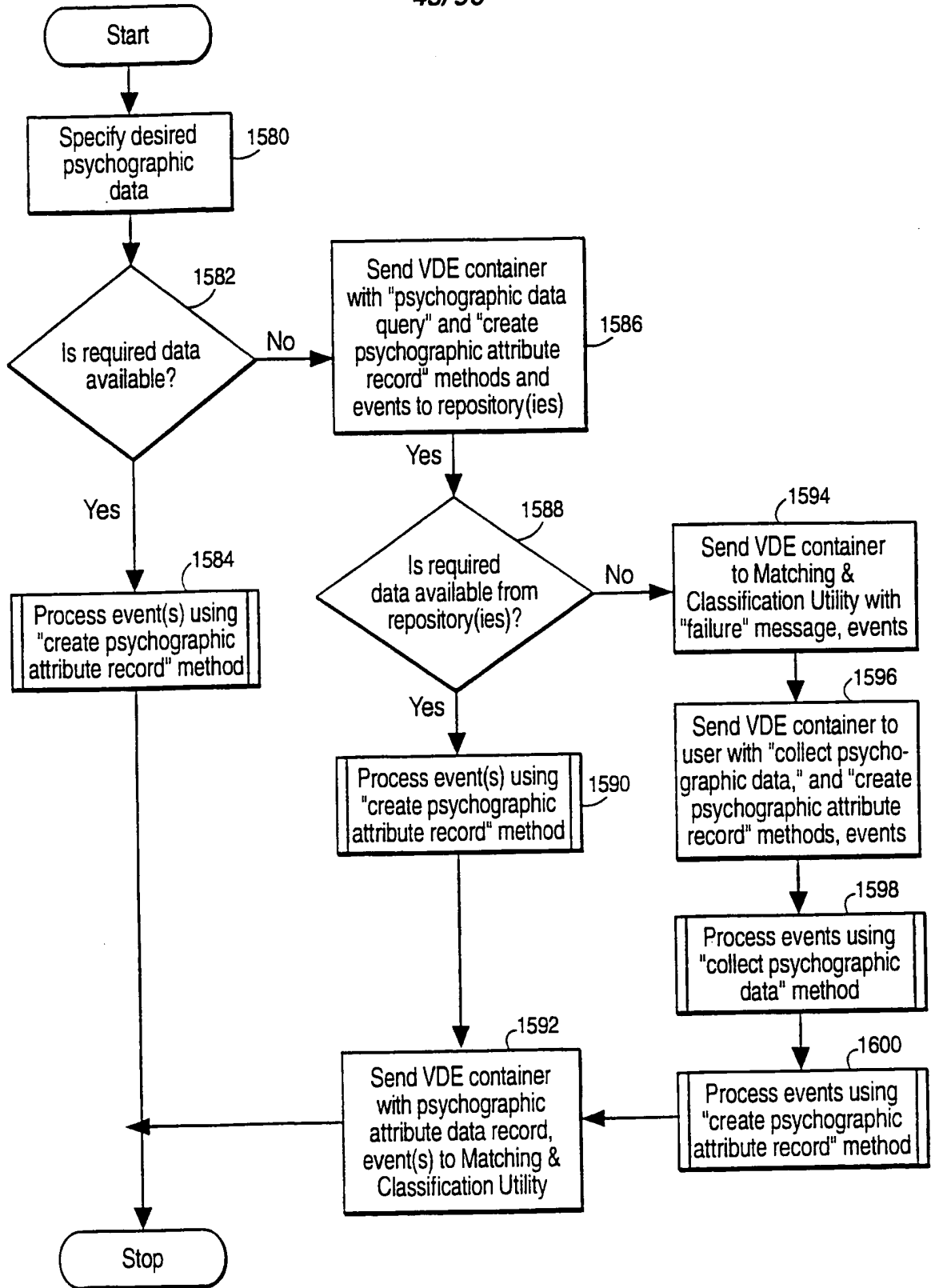


Fig. 30 Example Steps for Collecting Psychographic Data

46/96

Today's Anonymous Questionnaire
Thanks for taking the time to answer these questions
We'll put \$2.00 in your VDE budget

1. Do you feel sad, blue, unhappy or "down in the dumps"?

A. Never
 B. Rarely
 C. Sometimes
 D. Very Often
 E. Most of the time

2. Do you feel tired, having little energy, unable to concentrate?

A. Never
 B. Rarely
 C. Sometimes
 D. Very Often
 E. Most of the time

3. Do you feel uneasy, restless or irritable?

A. Never
 B. Rarely
 C. Sometimes
 D. Very Often
 E. Most of the time

4. Do you have trouble sleeping or eating (too little or too much)?

A. Never
 B. Rarely
 C. Sometimes
 D. Very Often
 E. Most of the time

[Click here for more questions](#)

All Information Will Be Treated As Confidential

Fig. 31 Example Psychographic Questionnaire "Pop-Up" Screen

47/96

Fig. 32(A) Example User Psychographic Attribute Information Record

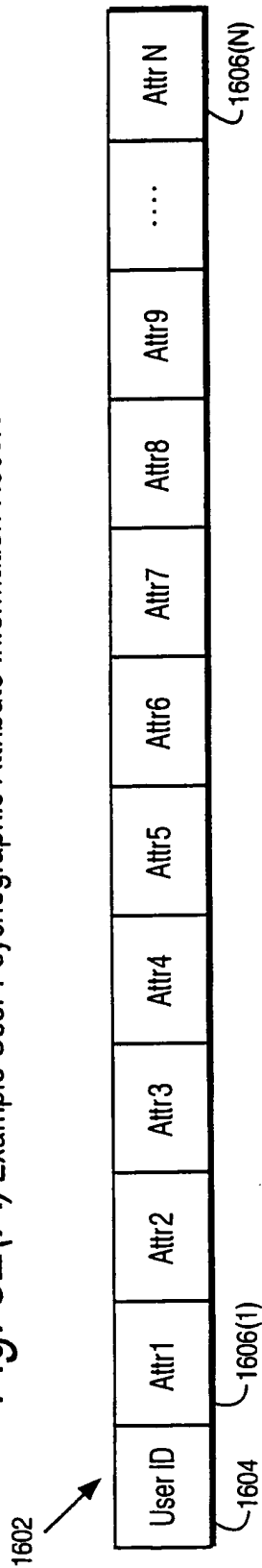
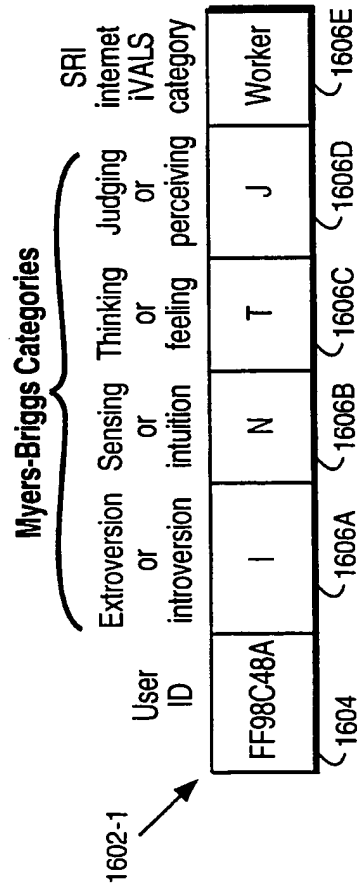


Fig. 32(B)
Example User Psychographic Attribute Record



48/96

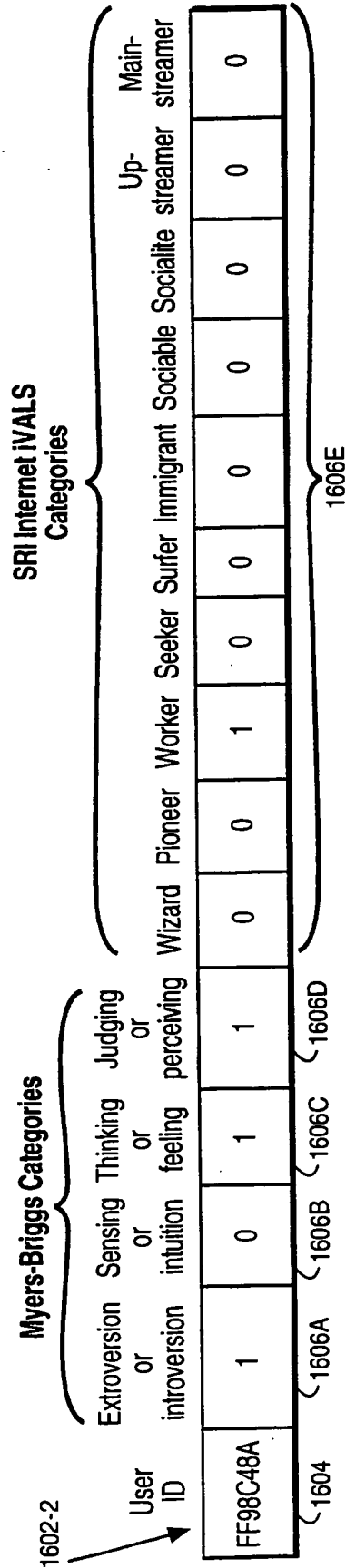
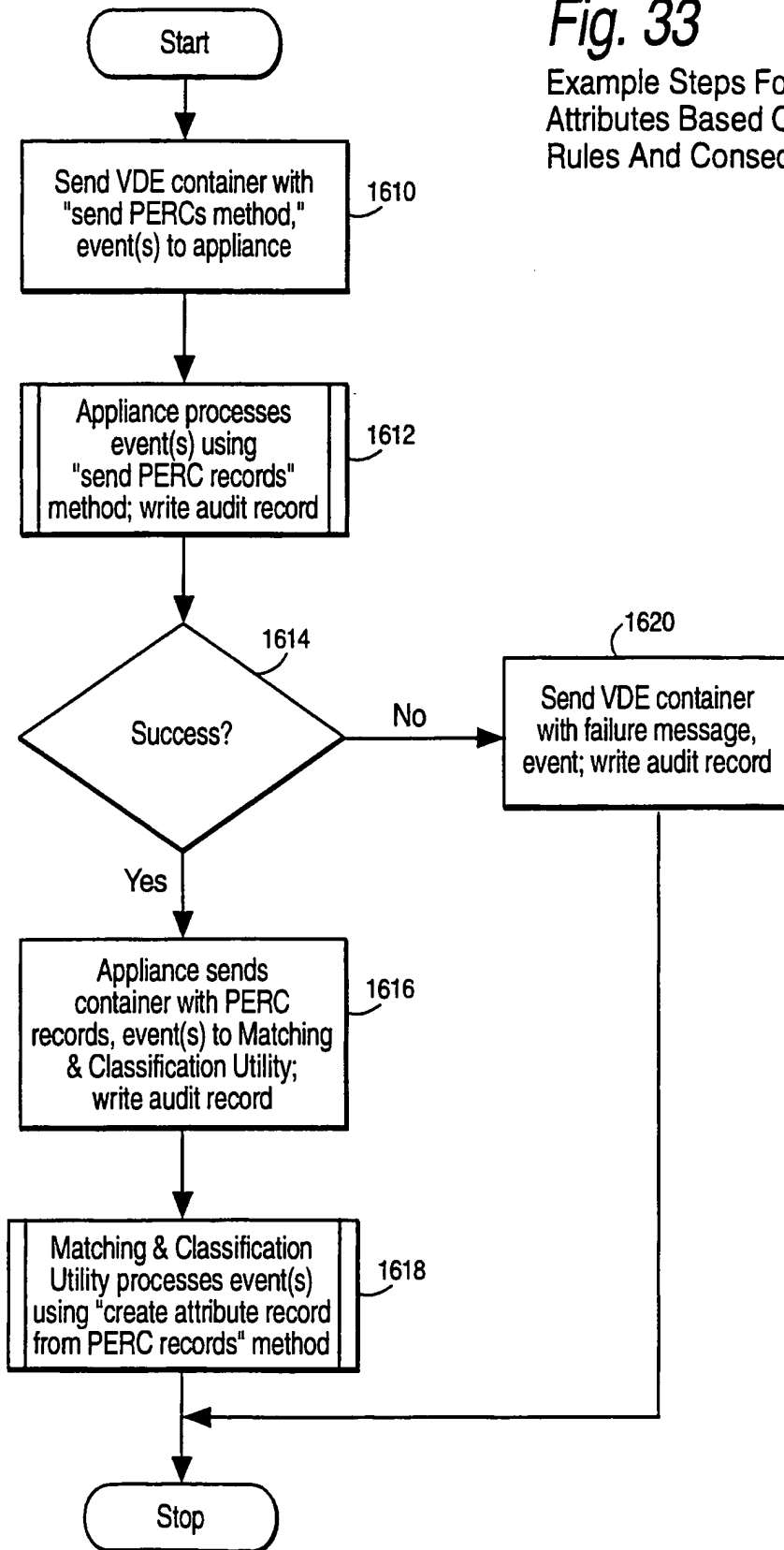


Fig. 32(C) Example Psychographic Attribute Record

49/96

Fig. 33

Example Steps For Determining Attributes Based On Available Rules And Consequences



50/96

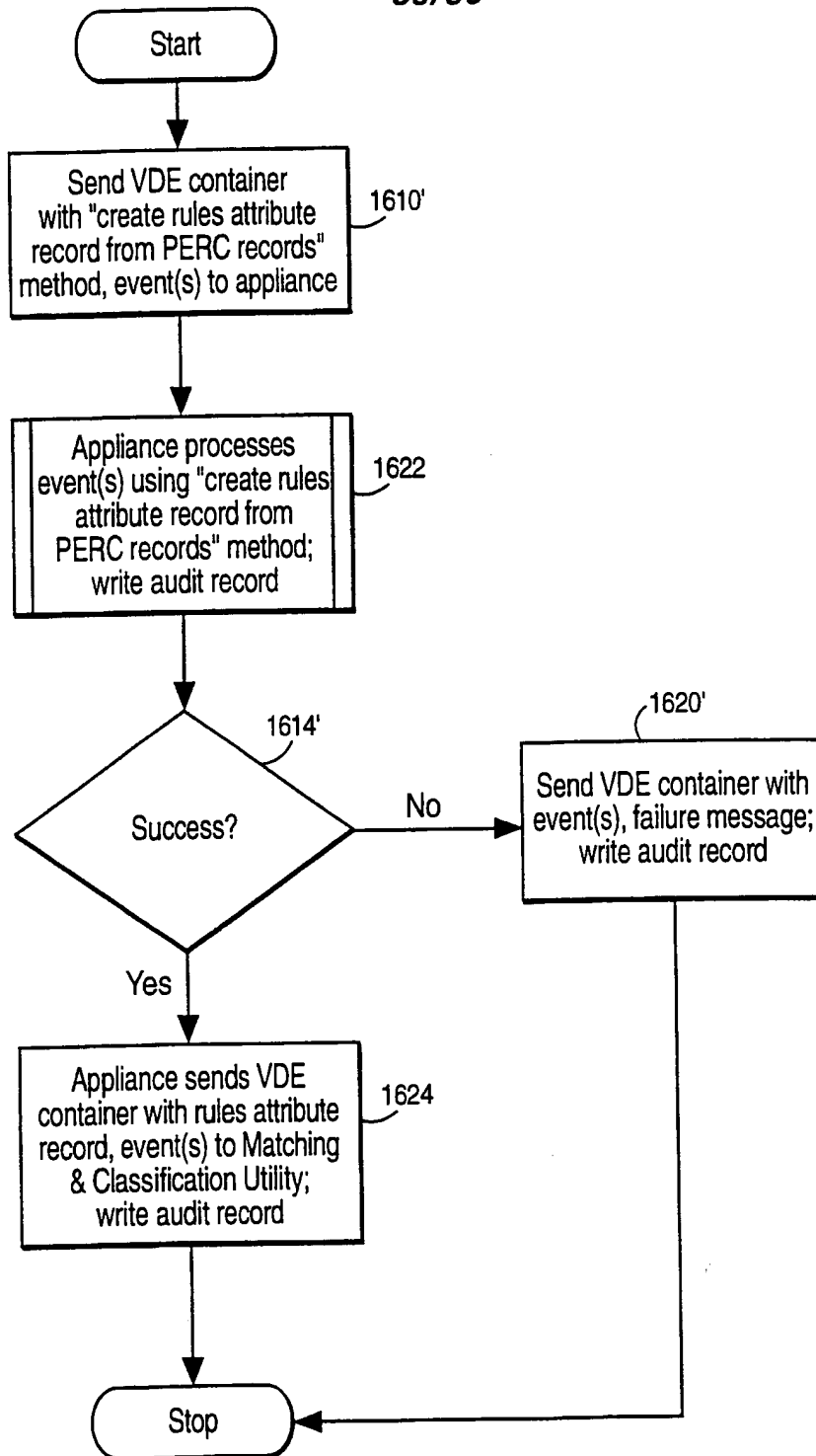


Fig. 34

Example Steps For Determining Attributes Based On Available Rules And Consequences

51/96

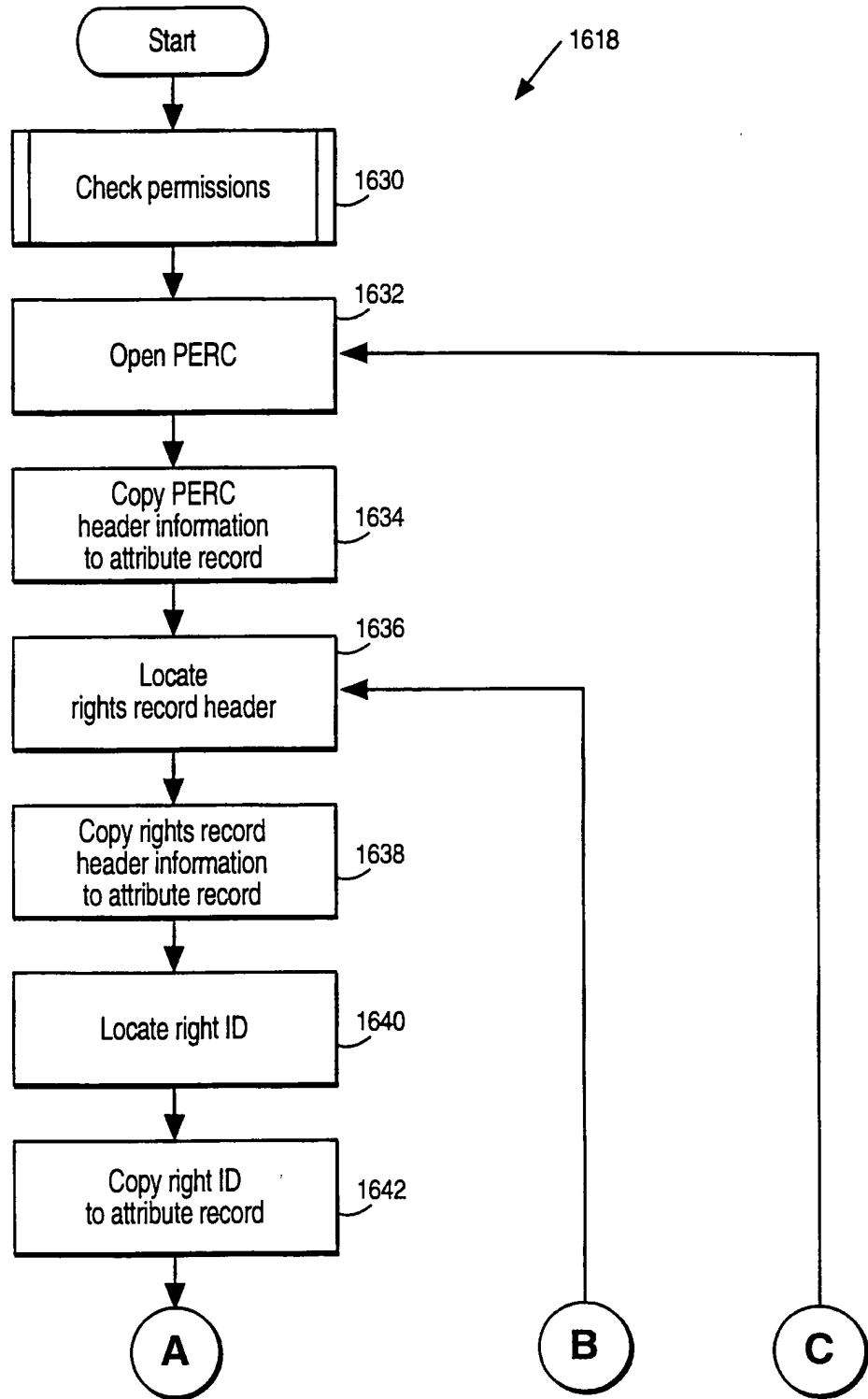


Fig. 35(A)

Construct Attribute Records From PERC Records Example Method

52/96

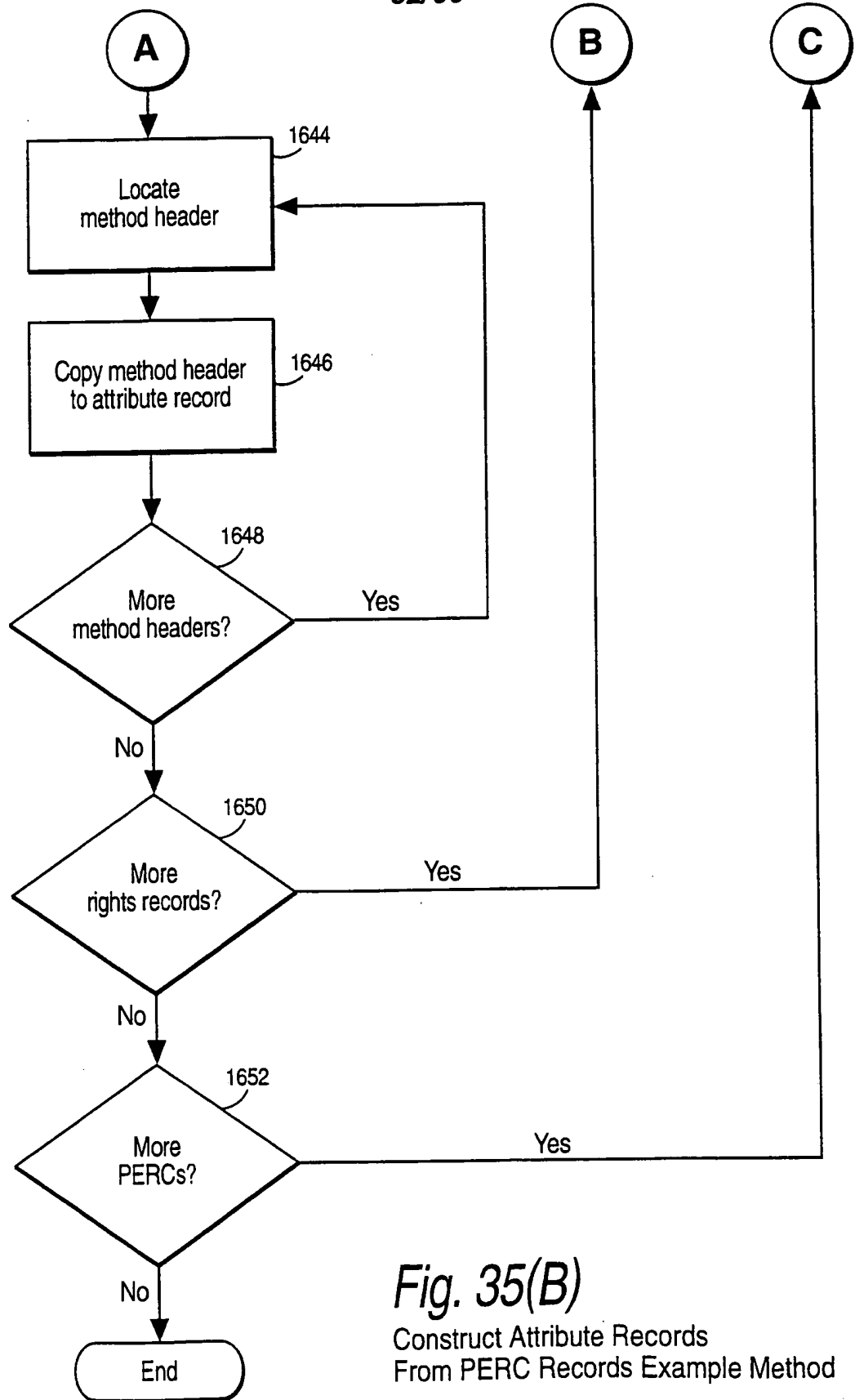


Fig. 35(B)

Construct Attribute Records
From PERC Records Example Method

53/96

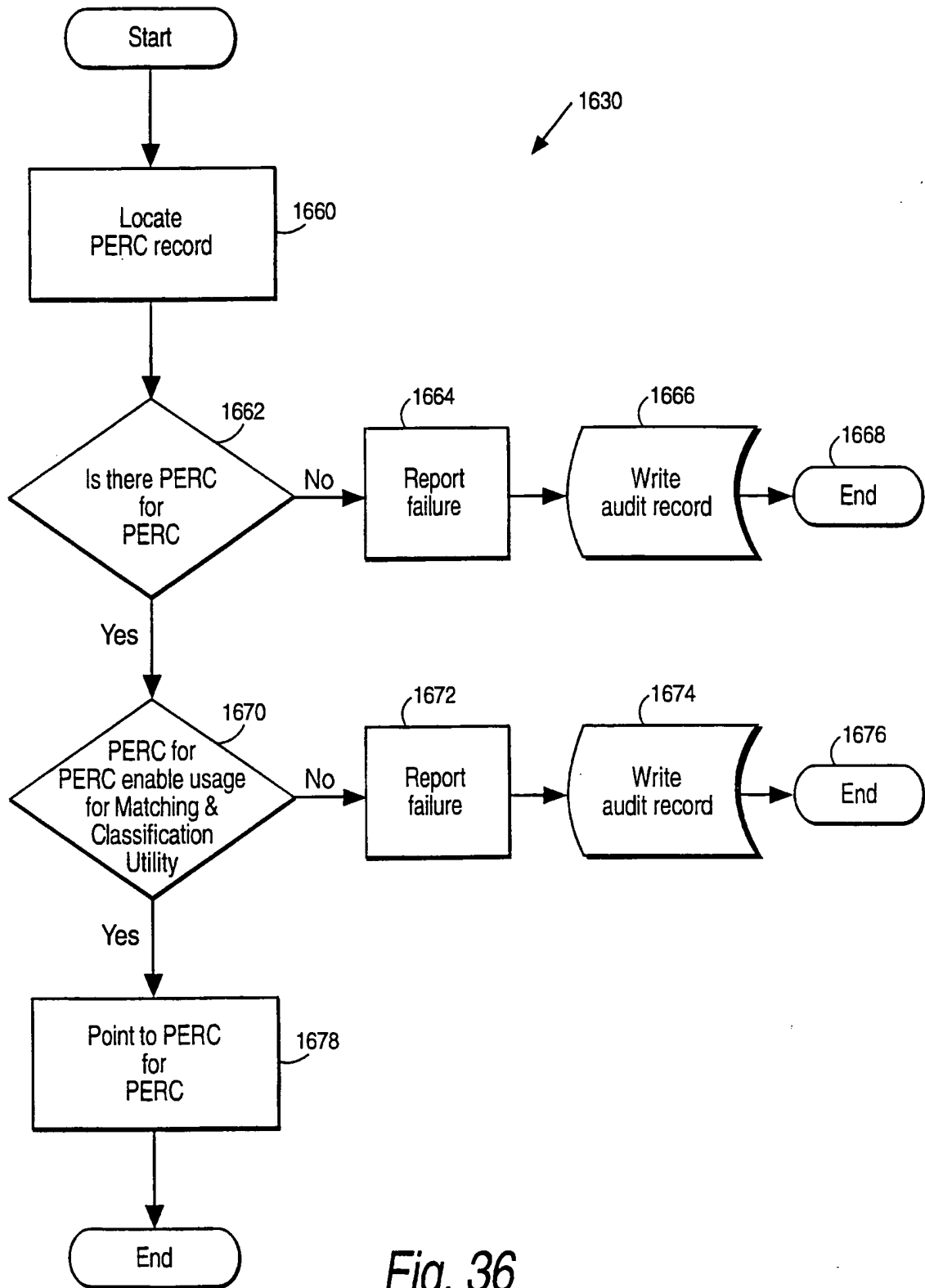


Fig. 36

Check Permissions Record Example Steps

Fig. 37(A) Example Rights Attribute Record From PERCs

1680-1	1682	1684	1686(1)	Attr1	Attr2	Attr3	Attr4	Attr5	Attr6	Attr7	Attr8	Attr9	Attr N	1686(N)
	User ID	Object ID		Attr1	Attr2	Attr3	Attr4	Attr5	Attr6	Attr7	Attr8	Attr9	Attr N	

Fig. 37(B) Example Attribute Record

1680-2	1682	1684	1686A	1686B	1686C	1686D	1686E	1686F	1686G	1686H	1686I	1686J	54/96
	User ID number	Object ID	Right ID	Method	Right ID	Method	Right ID	Method	Right ID	Method	Right ID	Method	Method	
	CF129CD5	1227-33-1298-2	Use	Open	Meter	Each time	Budget	One time purchase	\$1.00	Bill	VISA	AMEX		

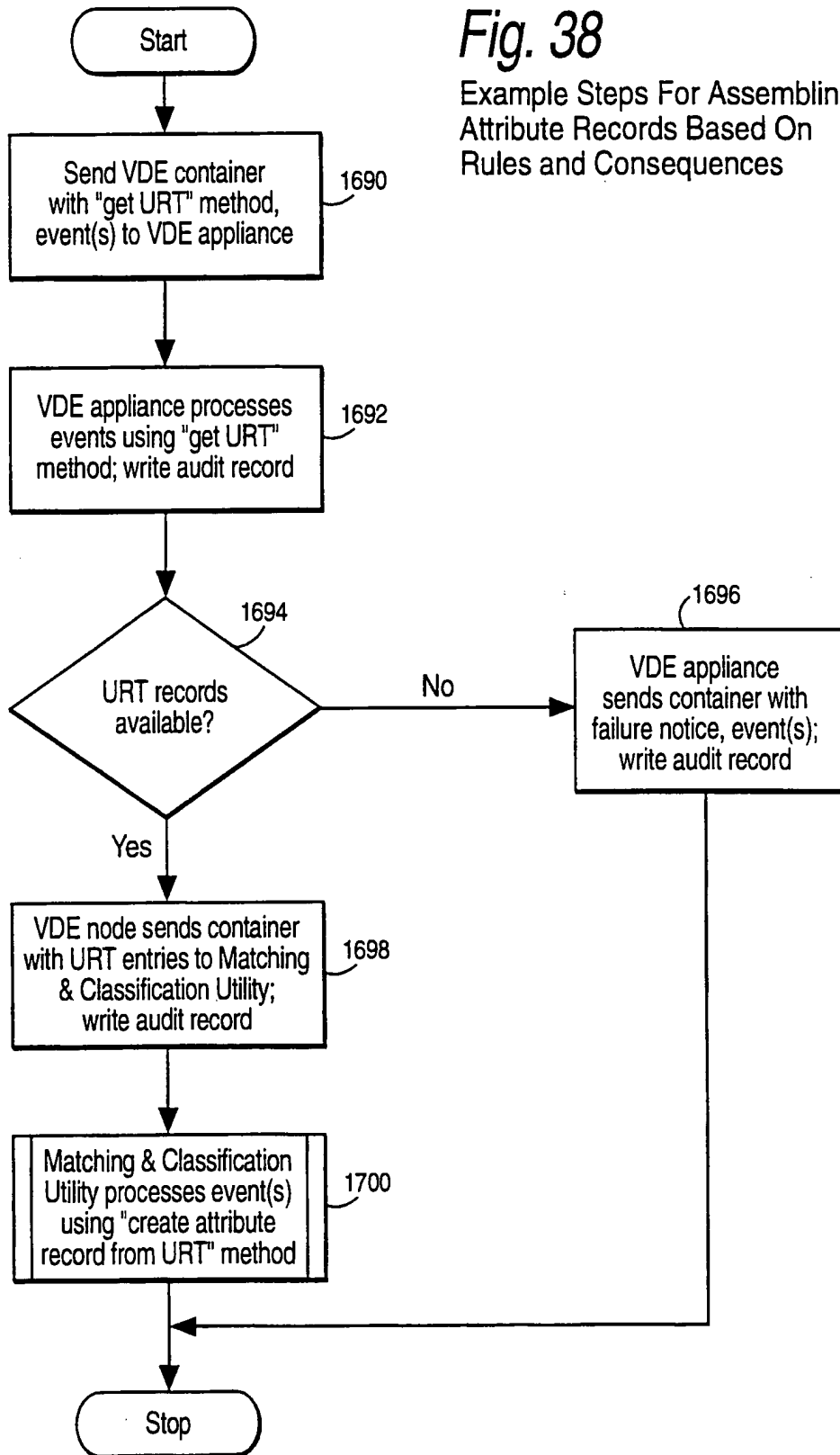
Fig. 37(C) Example Attribute Records From PERC Record

1680-3	1682	1684	1686A	1686B	1686C	1686D	1686E	1686F	1686G	1686H	1686I	1686J	
	User ID number	Object ID	Right ID	Method	Right ID	Method	Right ID	Method	Right ID	Method	Right ID	Method	Method	
	CF129CD5	1227-33-1298-2	27	239	15	546	81	423	1.00	02	666	601		

55/96

Fig. 38

Example Steps For Assembling Attribute Records Based On Rules and Consequences



56/96

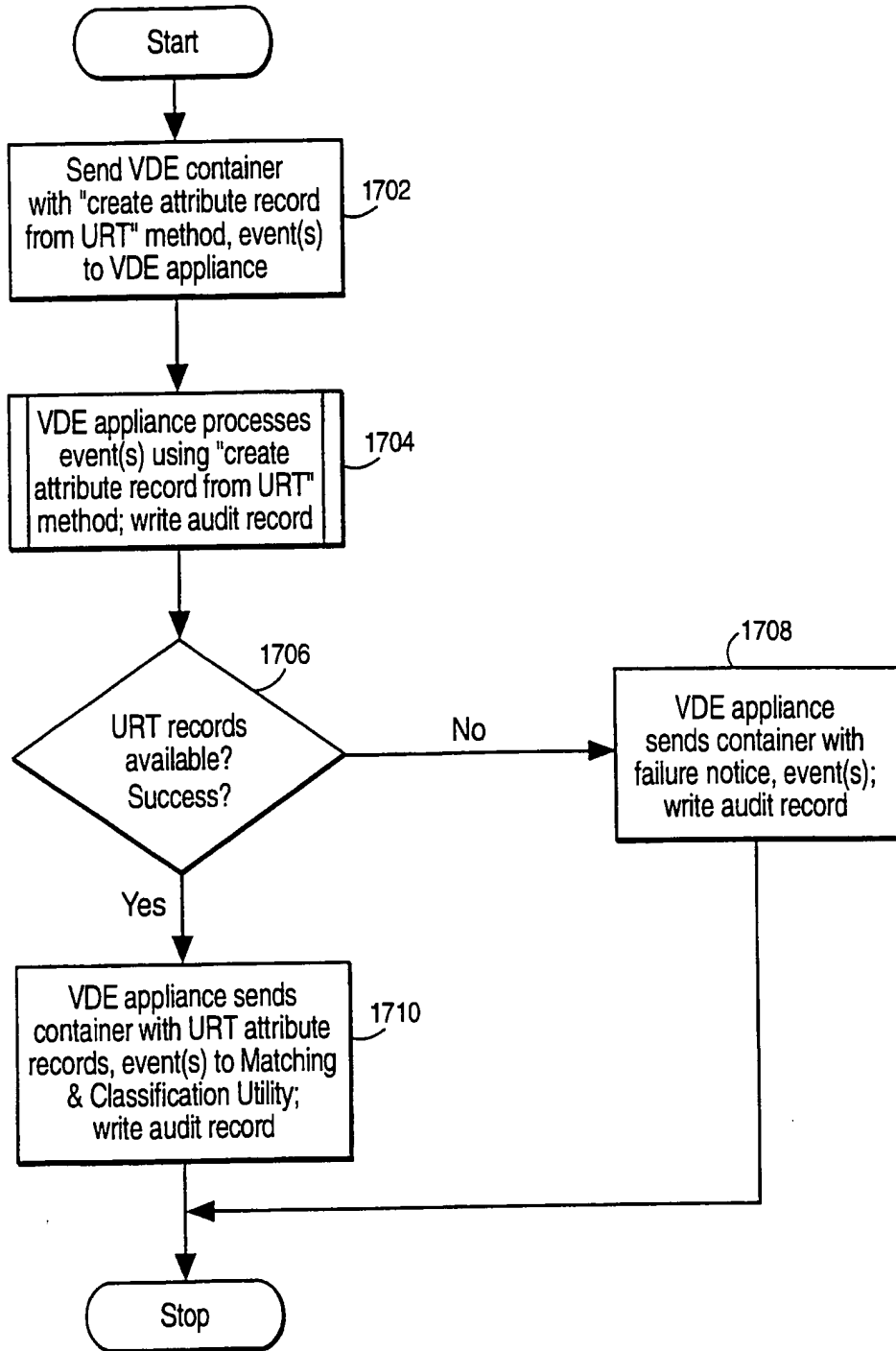


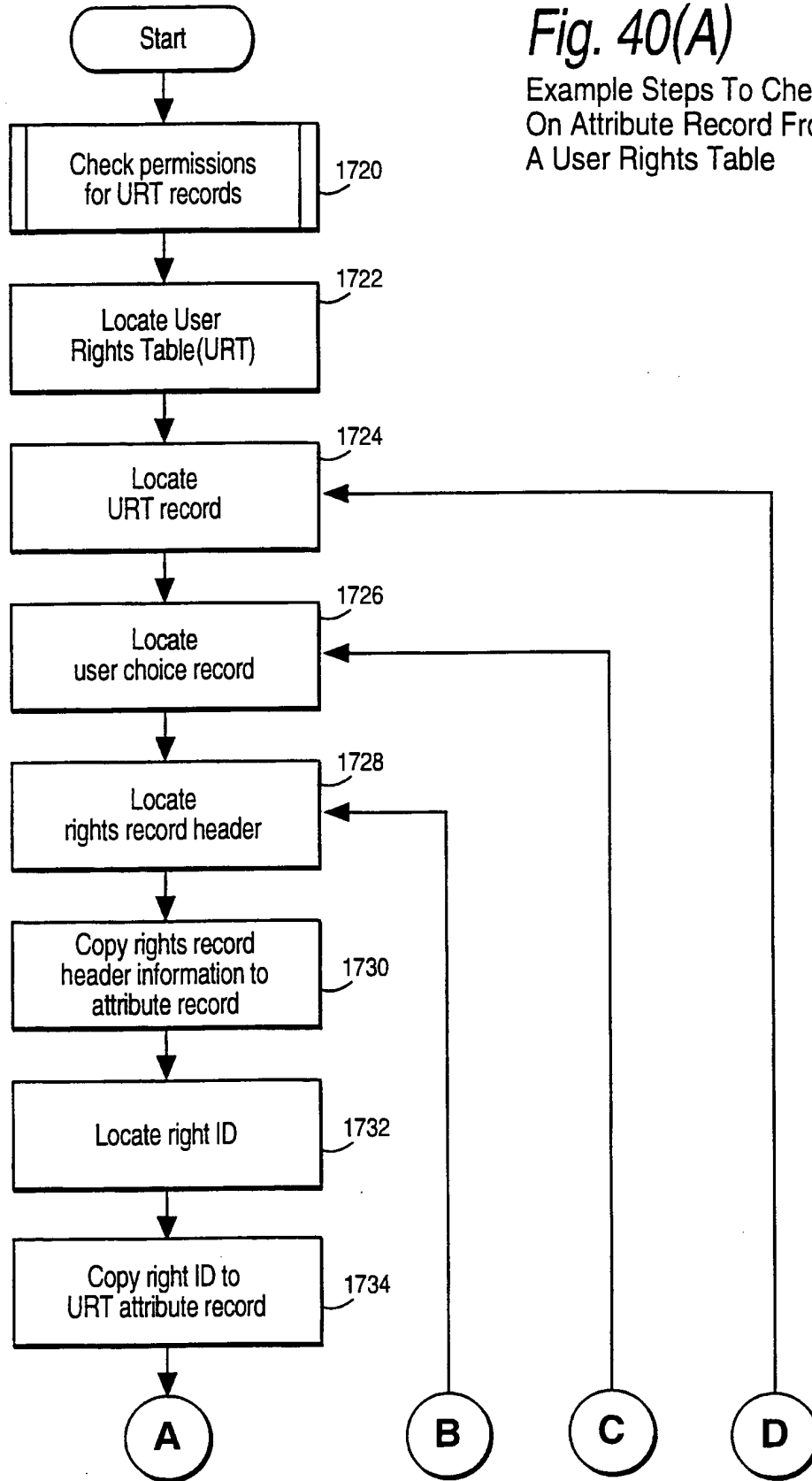
Fig. 39

Example Steps For Assembling Attribute Records Based On Rules and Consequences

57/96

Fig. 40(A)

Example Steps To Check
On Attribute Record From
A User Rights Table



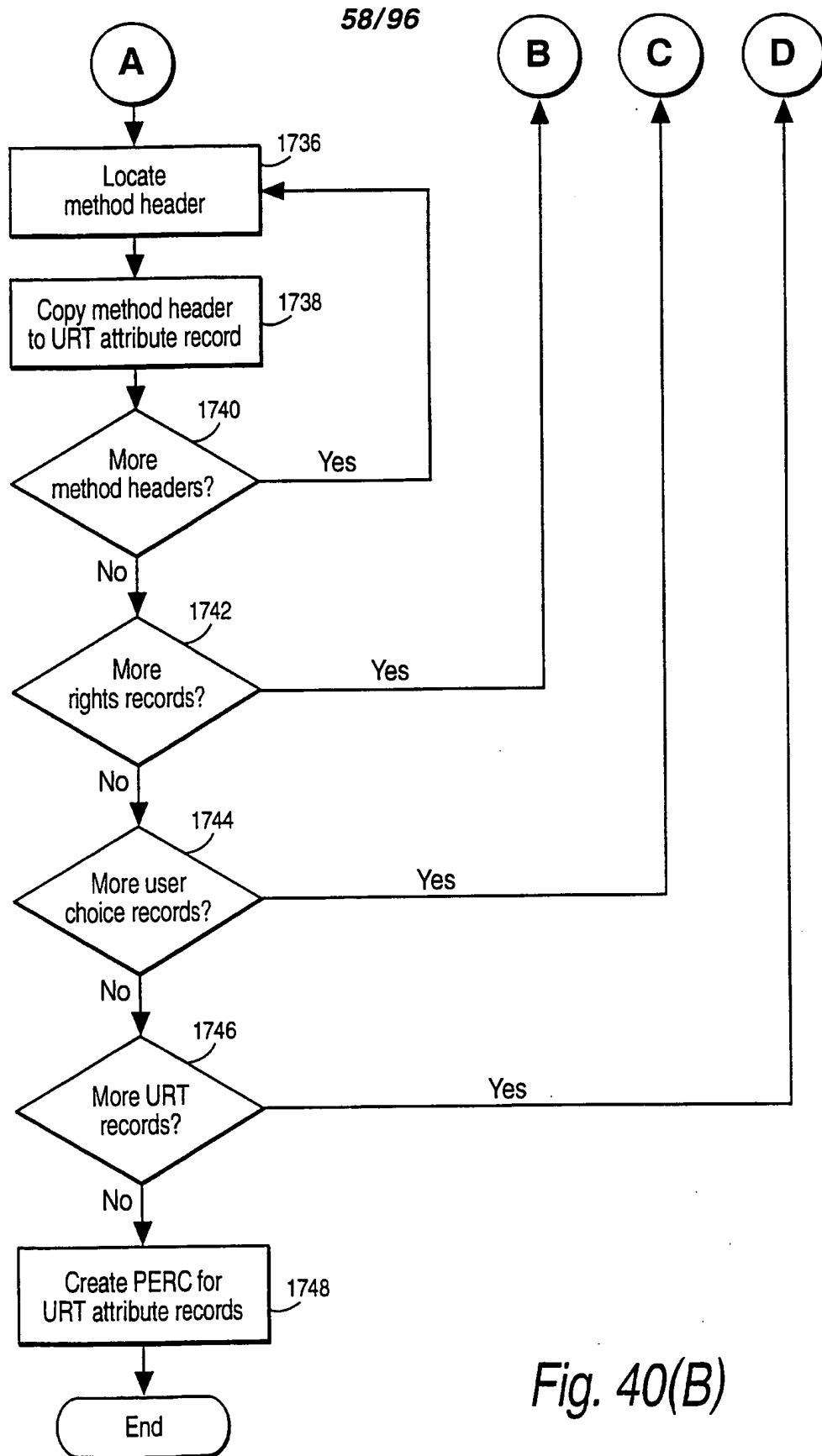


Fig. 40(B)

59/96

Fig. 41

Contract attribute records from PERC records method example

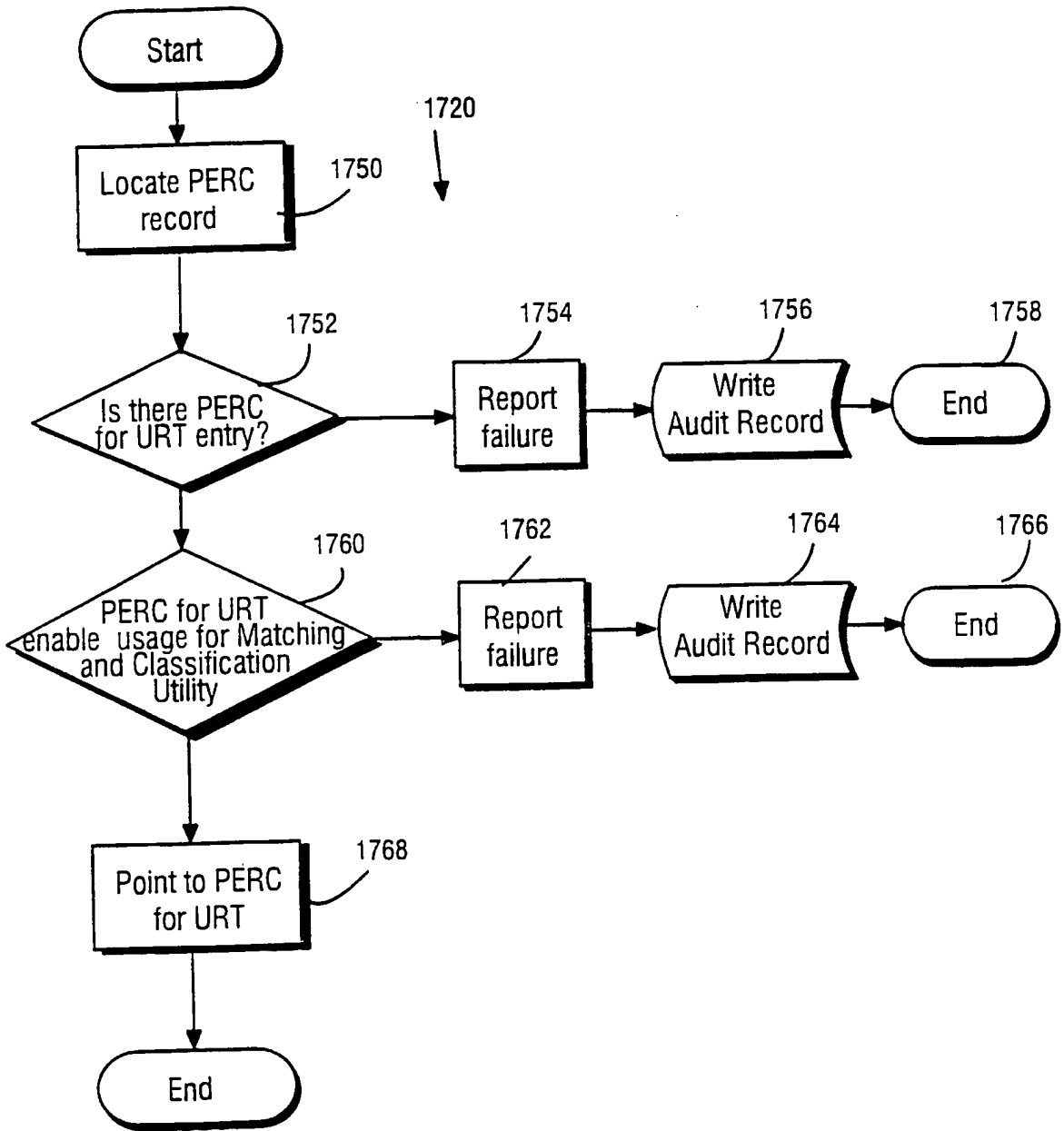


Fig. 42(A) Example Rights Attribute Record From URT

1770-1	User ID	Object ID	Attr1	Attr2	Attr3	Attr4	Attr5	Attr6	Attr7	Attr8	Attr9	Attr N
	1772	1774	1776(1)										1776(N)

60/96

Fig. 42(B) Example Rights Attribute Record From URT

User ID number	Object ID	Right ID	Method	Right ID	Method	Right ID	Method	Right ID	Method
1770-2	1774	1776A	Open	1776B	Meter	1776C	Each time	1776D	Budget
		1776E	One time purchase	1776F	\$1.00	1776G	Bill	1776H	VISA
		1776I						

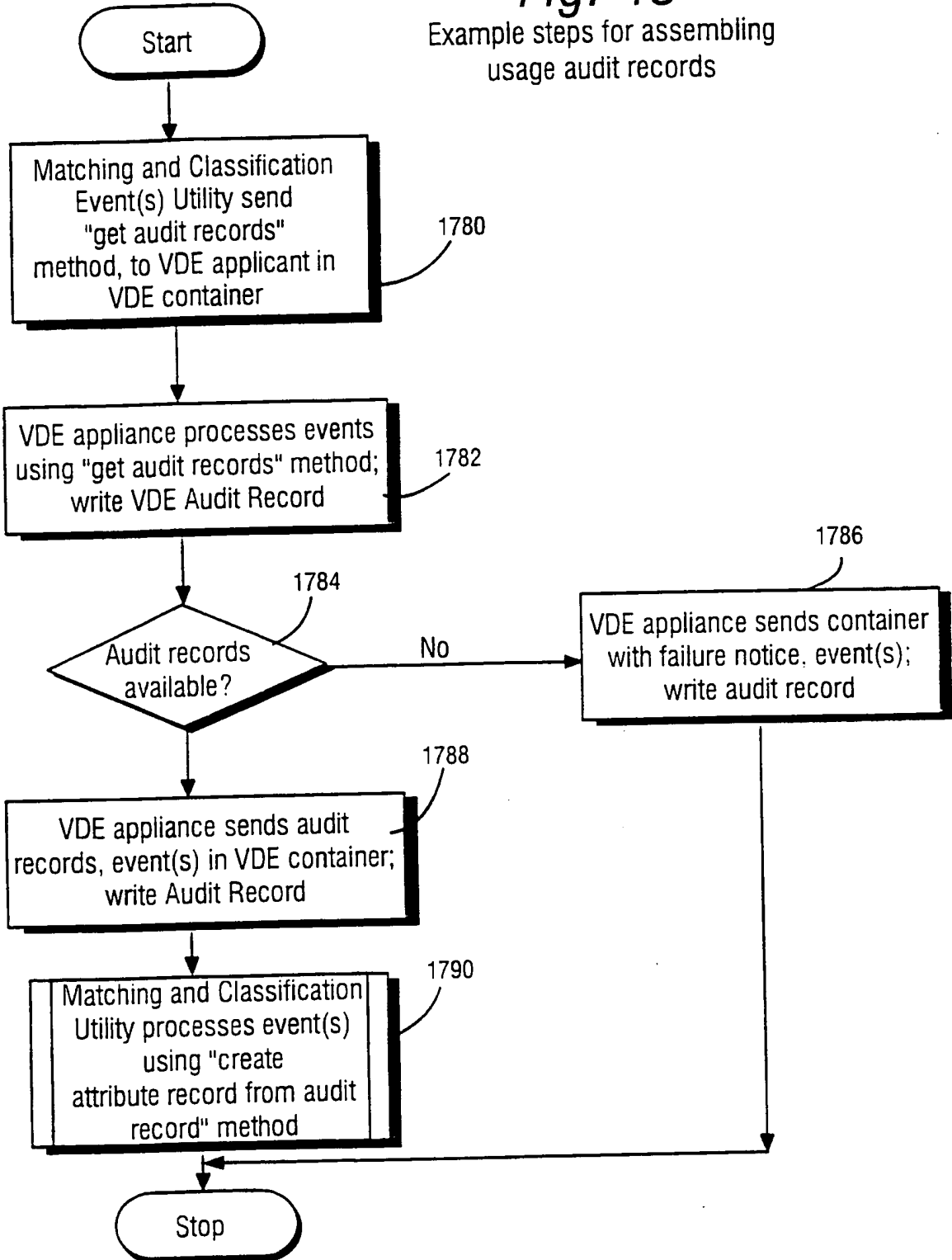
Fig. 42(C) Example User Choice Attribute Records (From URT)

User ID number	Object ID	Right ID	Method	Right ID	Method	Right ID	Method
1770-3	1774	1776A	27	1776B	239	1776C	15
		1776D	546	1776E	81	1776F	423
		1776G	1.00	1776H	02	1776I	666

61/96

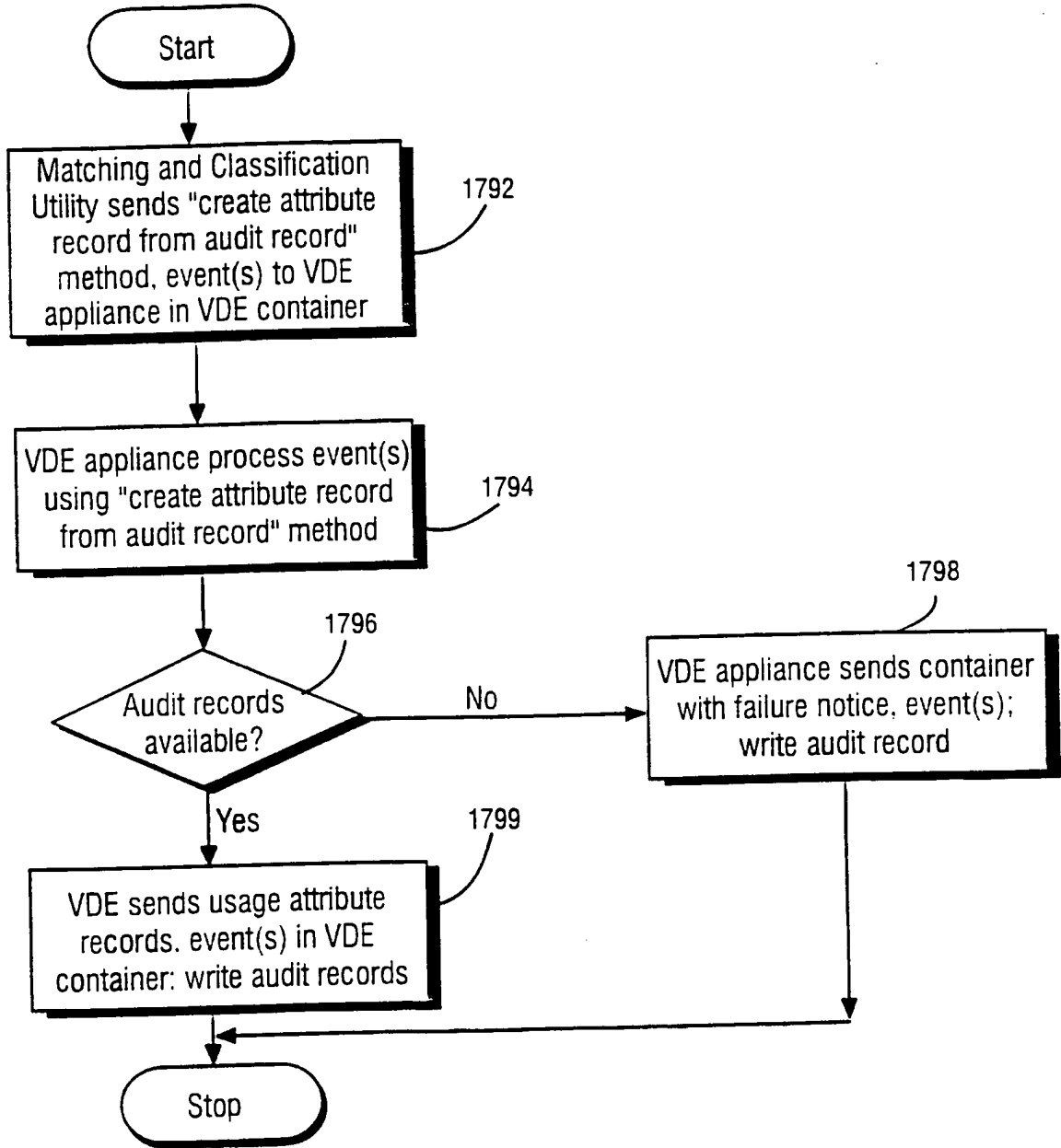
Fig. 43

Example steps for assembling usage audit records



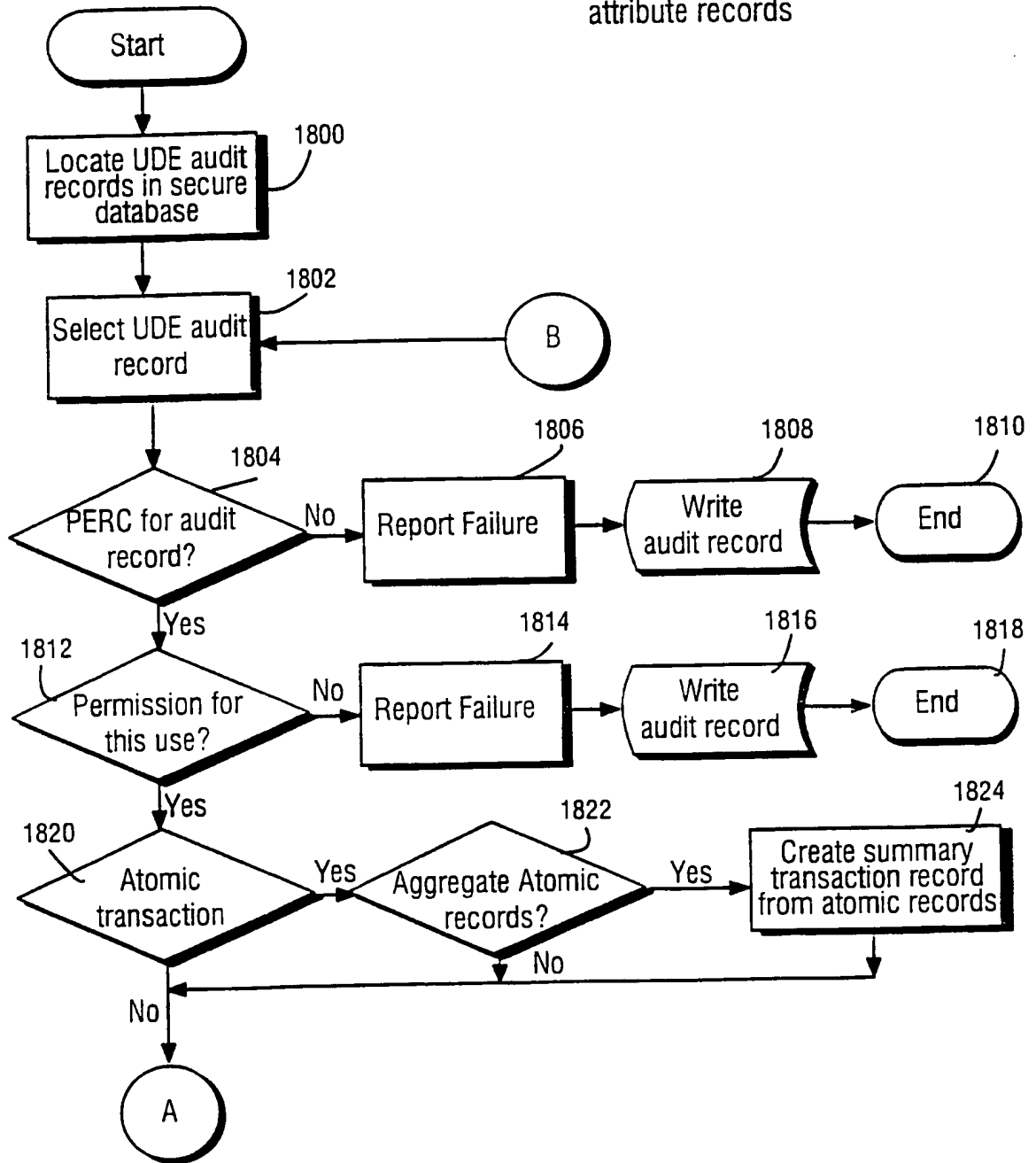
62/96

Fig. 44
Example steps for assembling usage
audit records



63/96

Fig. 45(A)
Example steps to create audit attribute records



64/96

Fig. 45(B)
Example steps to create audit
attribute records

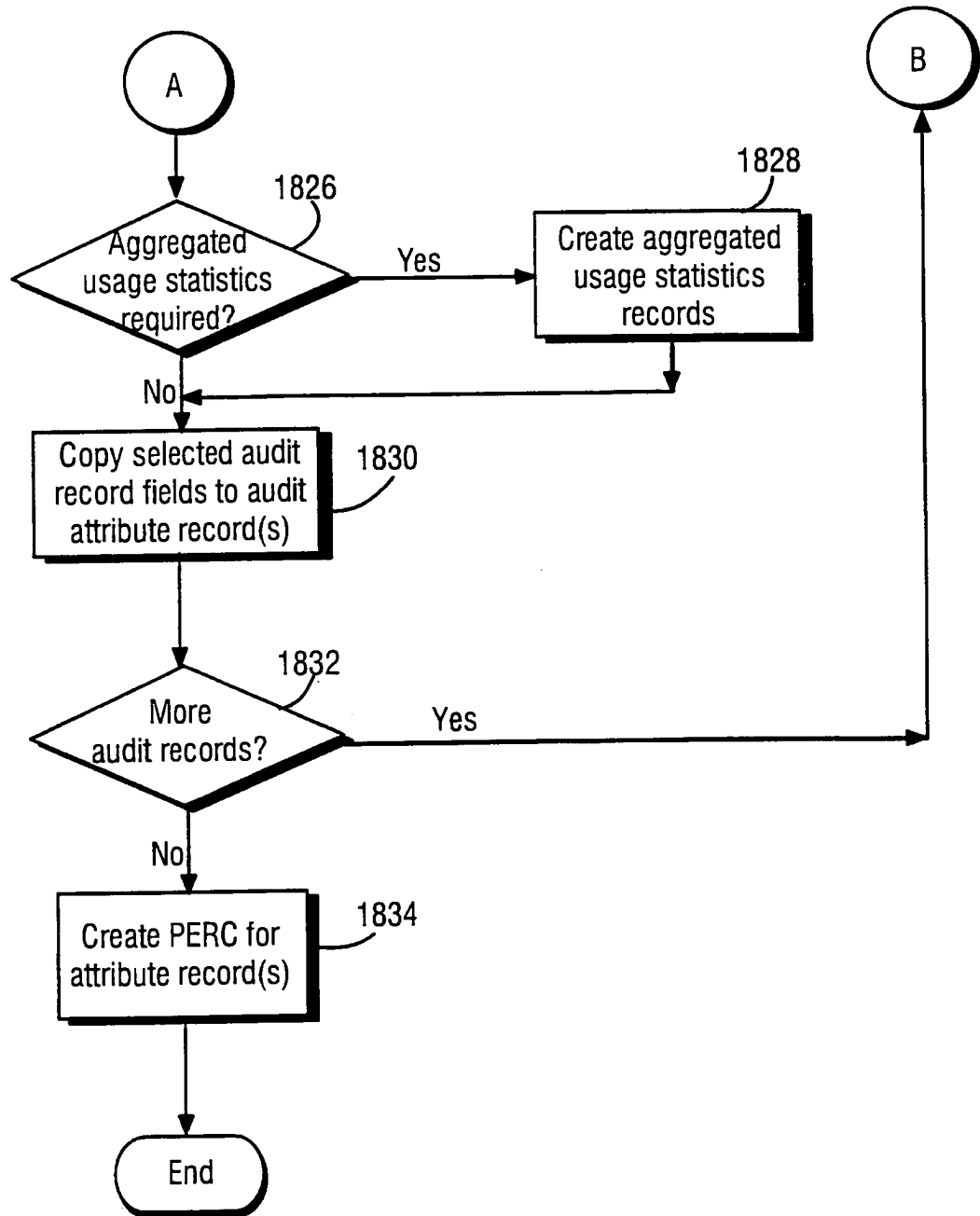


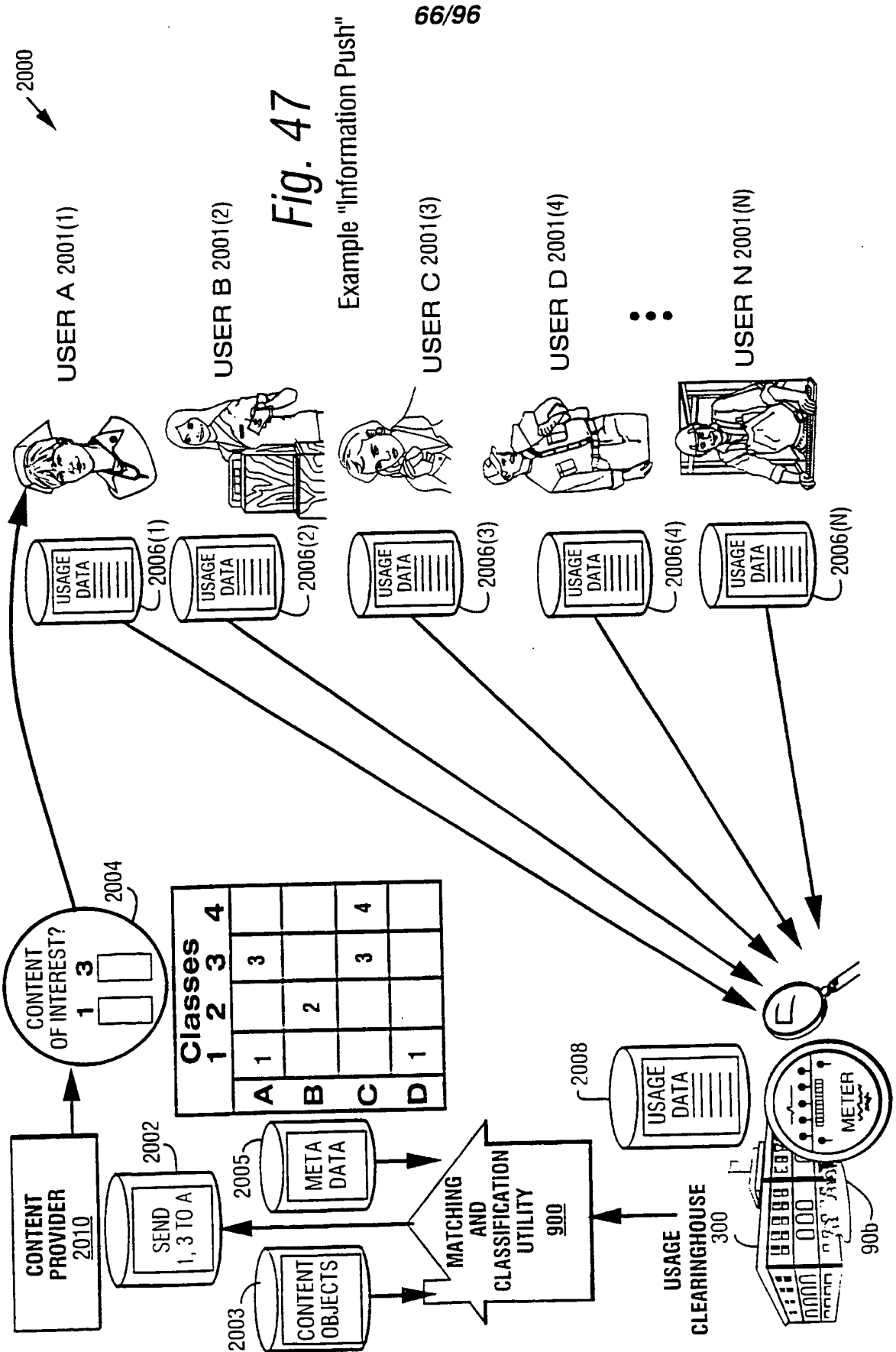
Fig. 46(A) Example Usage Attribute Record From UDE Audit Record

1830-1	1832	1834	1836(1)	1836(N)
User ID	Object ID	Attr1	Attr2	Attr3
		Attr4	Attr5	Attr6
		Attr7	Attr8	Attr9
			Attr N

Fig. 46(B) Example Usage Attribute/Statistics Records From UDE Audit Records

65/96	1830-2	1832	1834	1836A	1836B	1836C	1836D	1836E	1836F	1836G	1836H	1836I	1836J
	User ID number	Object ID	Right ID	Method	Right ID	Method	Right ID	Number opens	Right ID	Method	Right ID	Method	Method
	CF129CD5	1227-33-1298-2	Use	Open	Meter	Each time	4	Budget	One time purchase	Bill	\$1.00	VISA

1836K	1836L	1836M	1836N	1836O
Account number	First use date	First use time	Most recent use date	Most recent use time
xxx-YYYYYY-zzzzzzzzzzzzz	12/30/95	18:22:30 EST	01/04/96	20:14:01



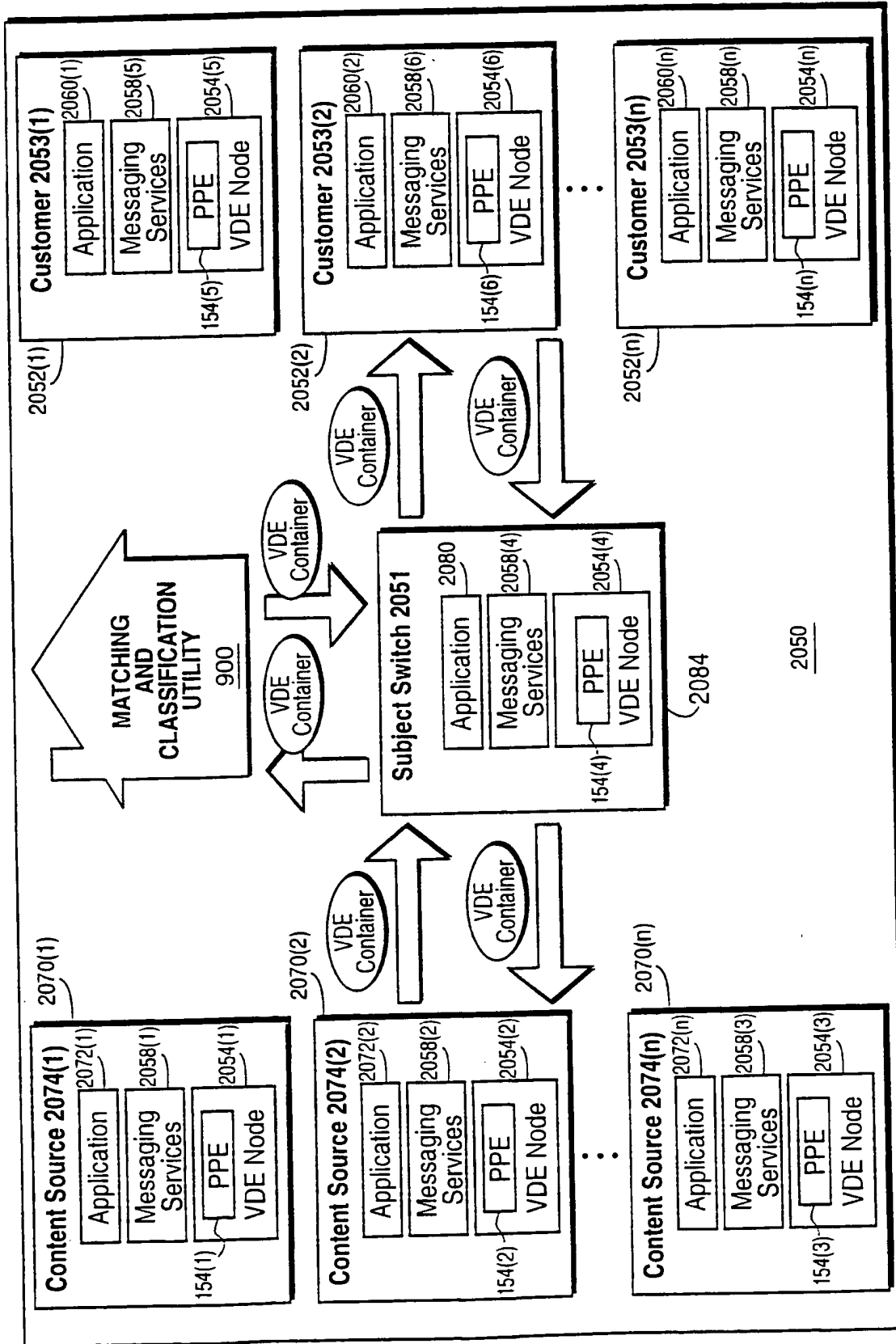


Fig. 47(A) Matching and Classification Utility 900 Supports "Push" models using Subject Switching and Messaging Services

68/96

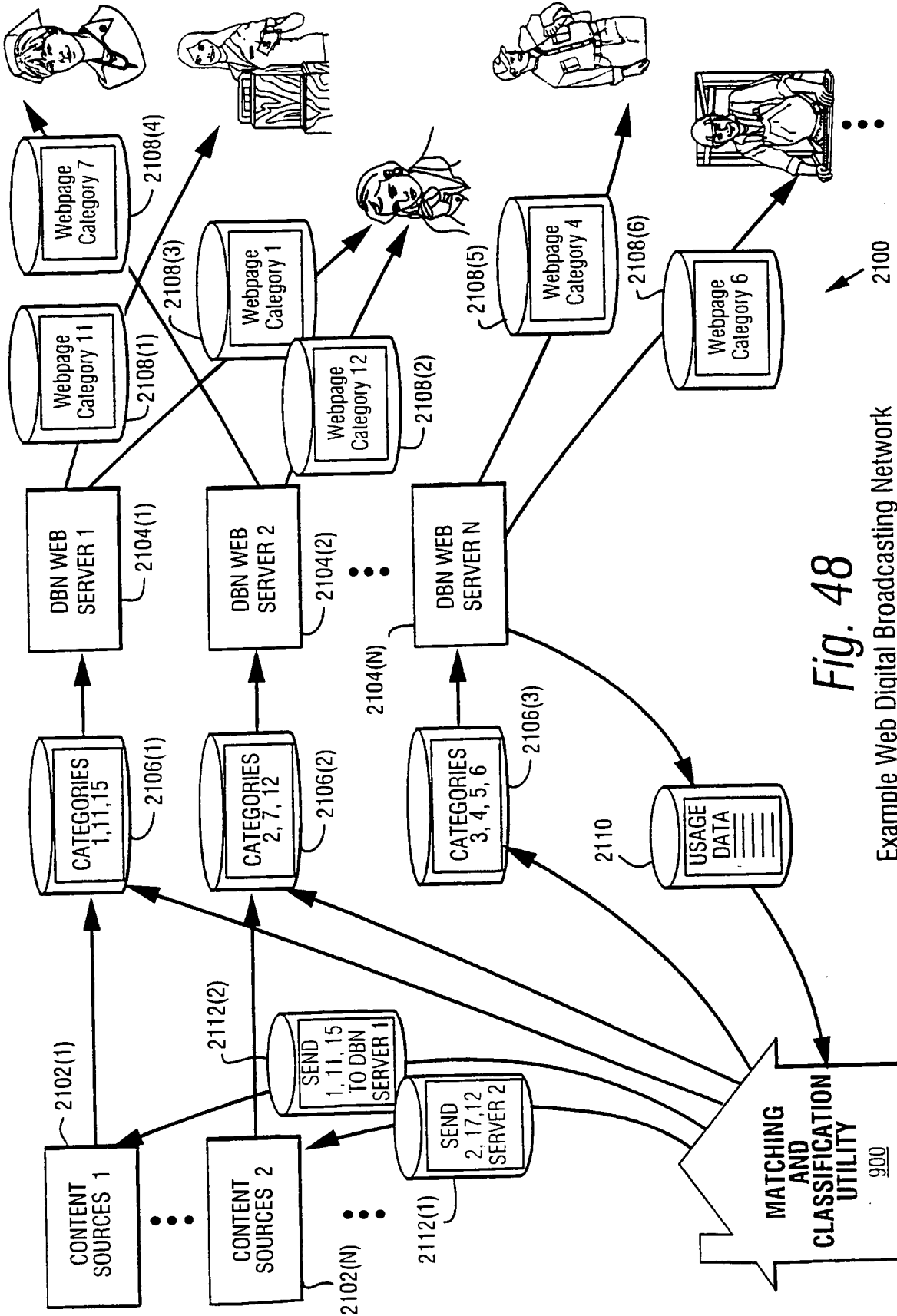


Fig. 48

Example Web Digital Broadcasting Network

69/96

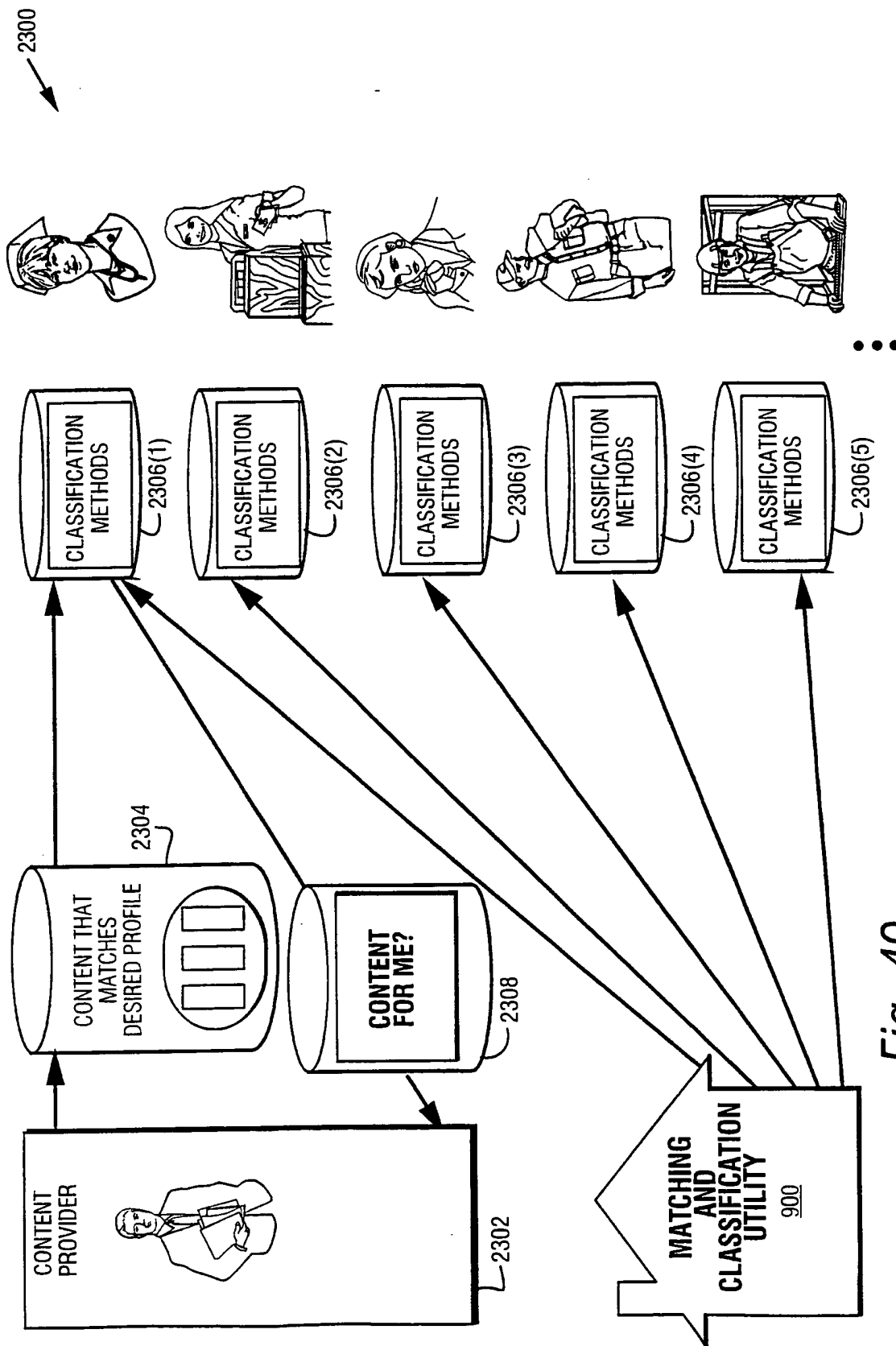


Fig. 49 Example "Consumer Pull"

70/96

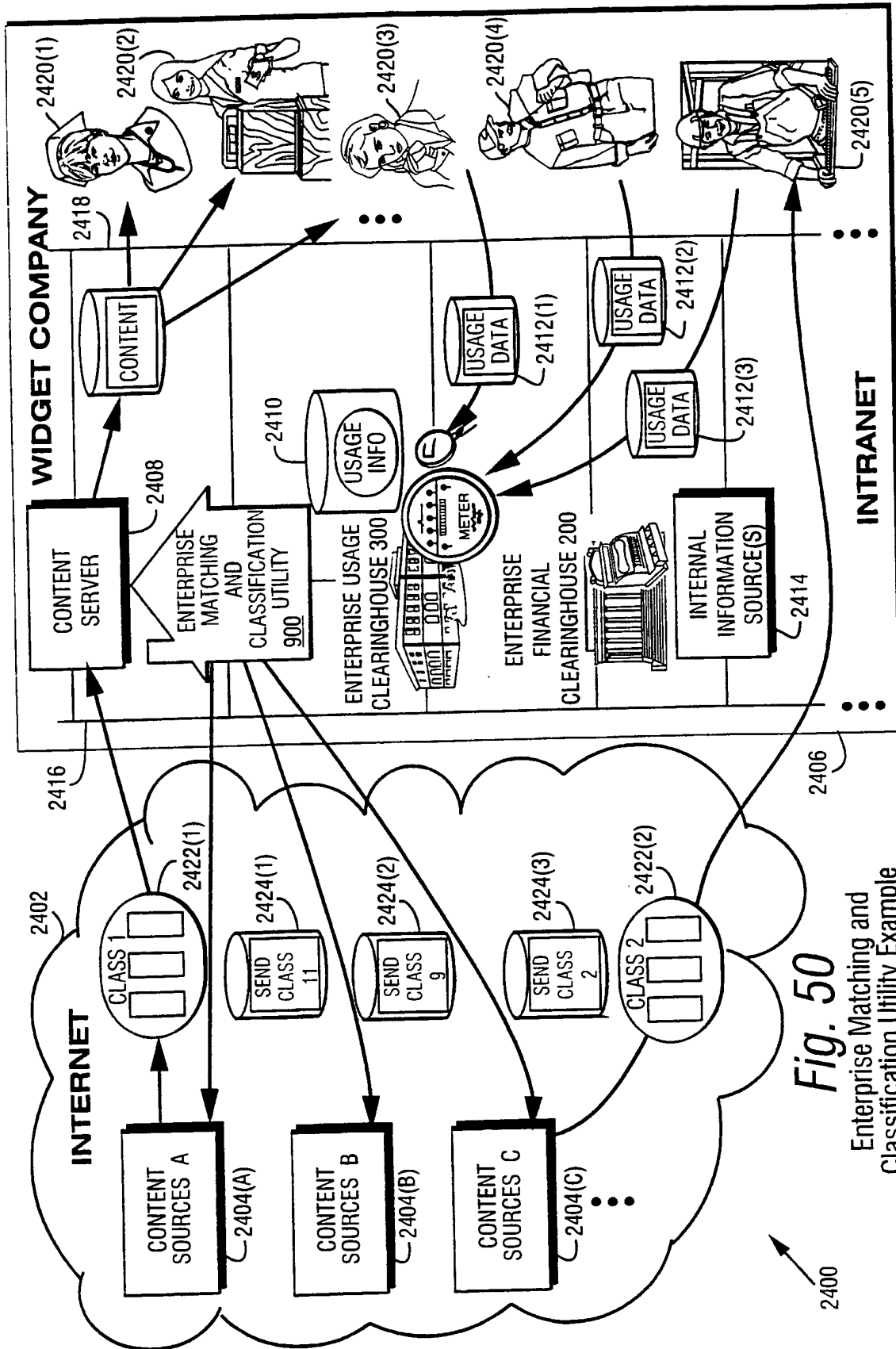
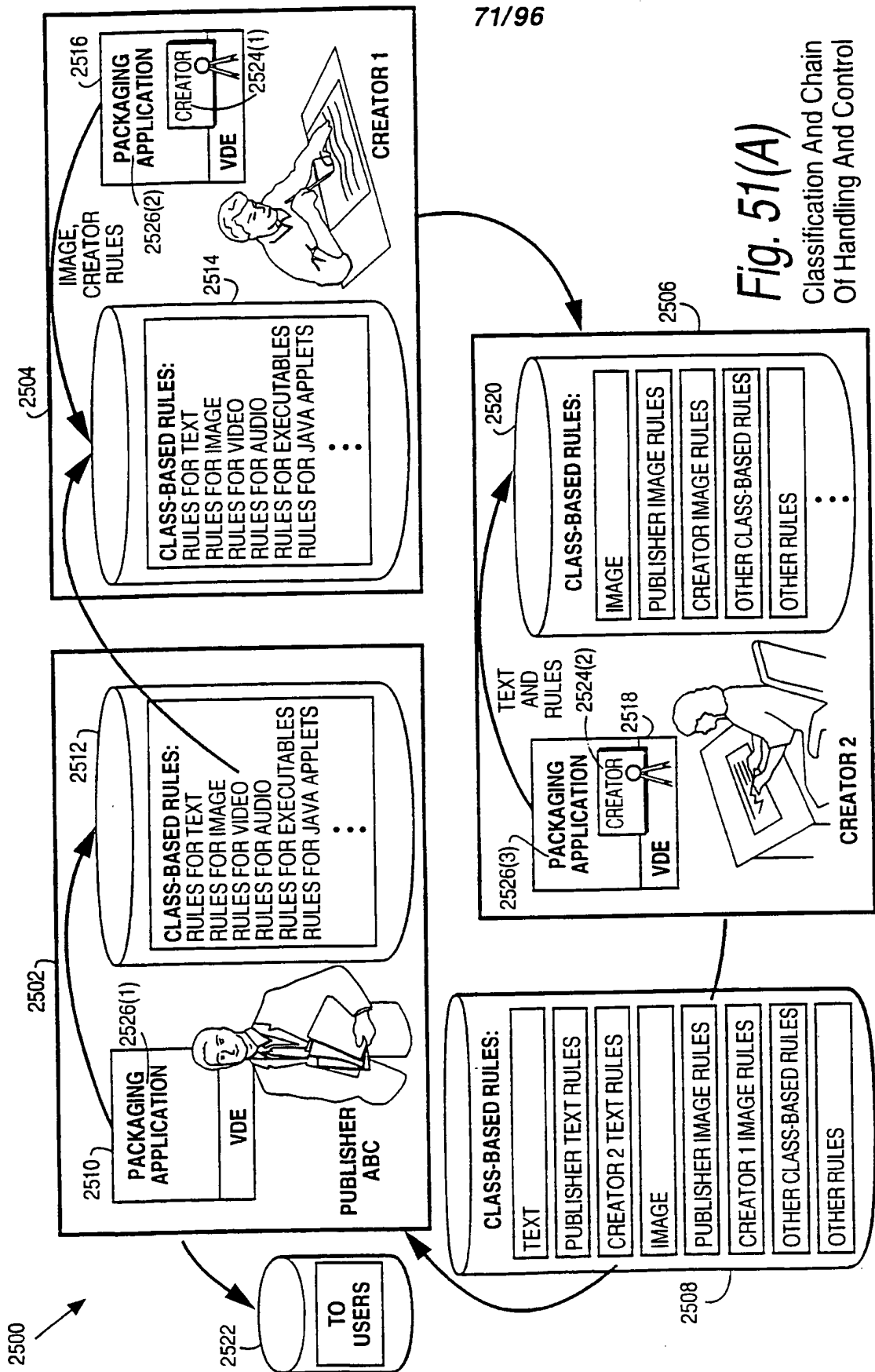


Fig. 50
Enterprise Matching and
Classification Utility Example



71/96

Fig. 51(A)
Classification And Chain
Of Handling And Control

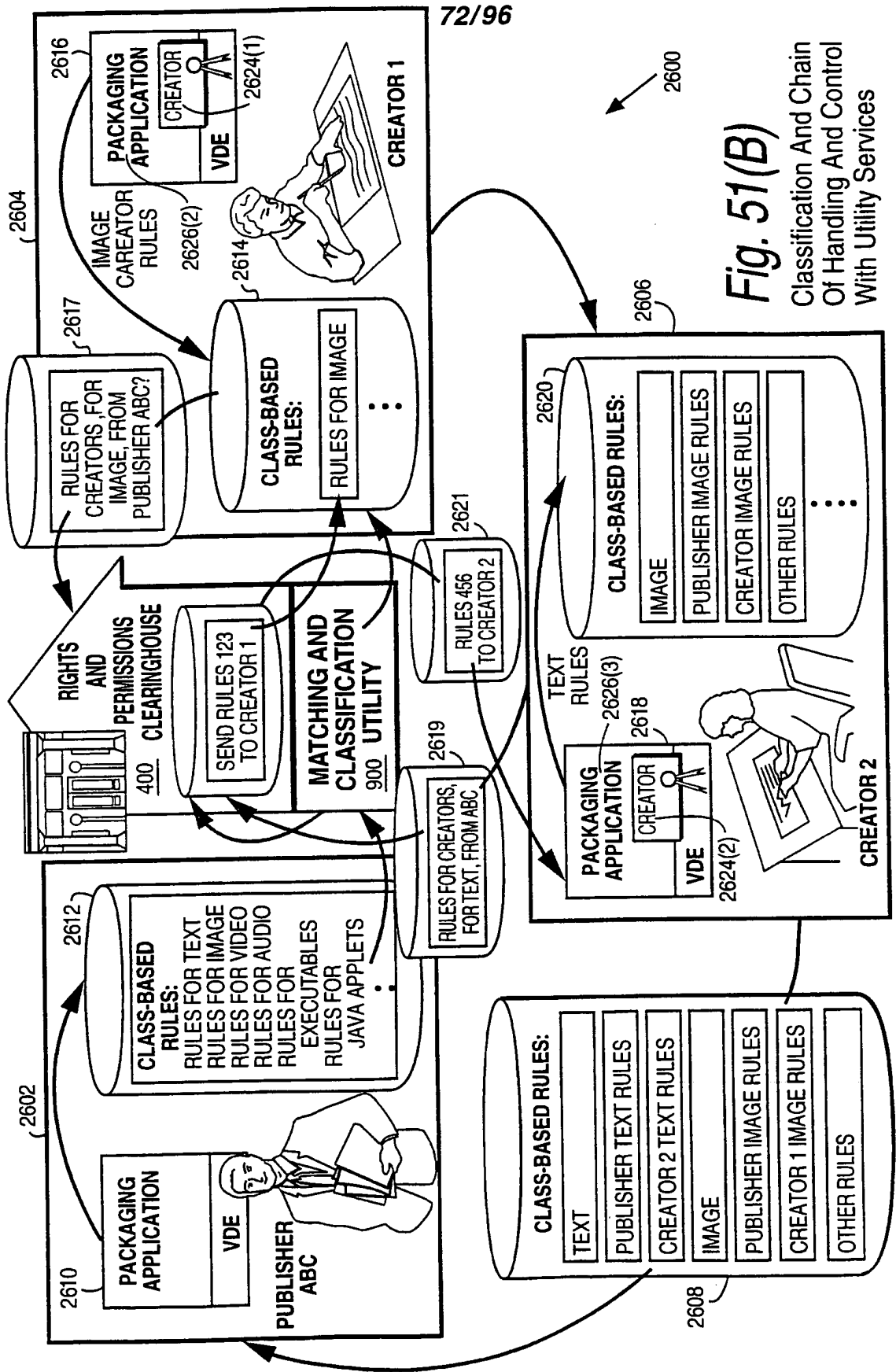


Fig. 51(B)
 Classification And Chain
 Of Handling And Control
 With Utility Services

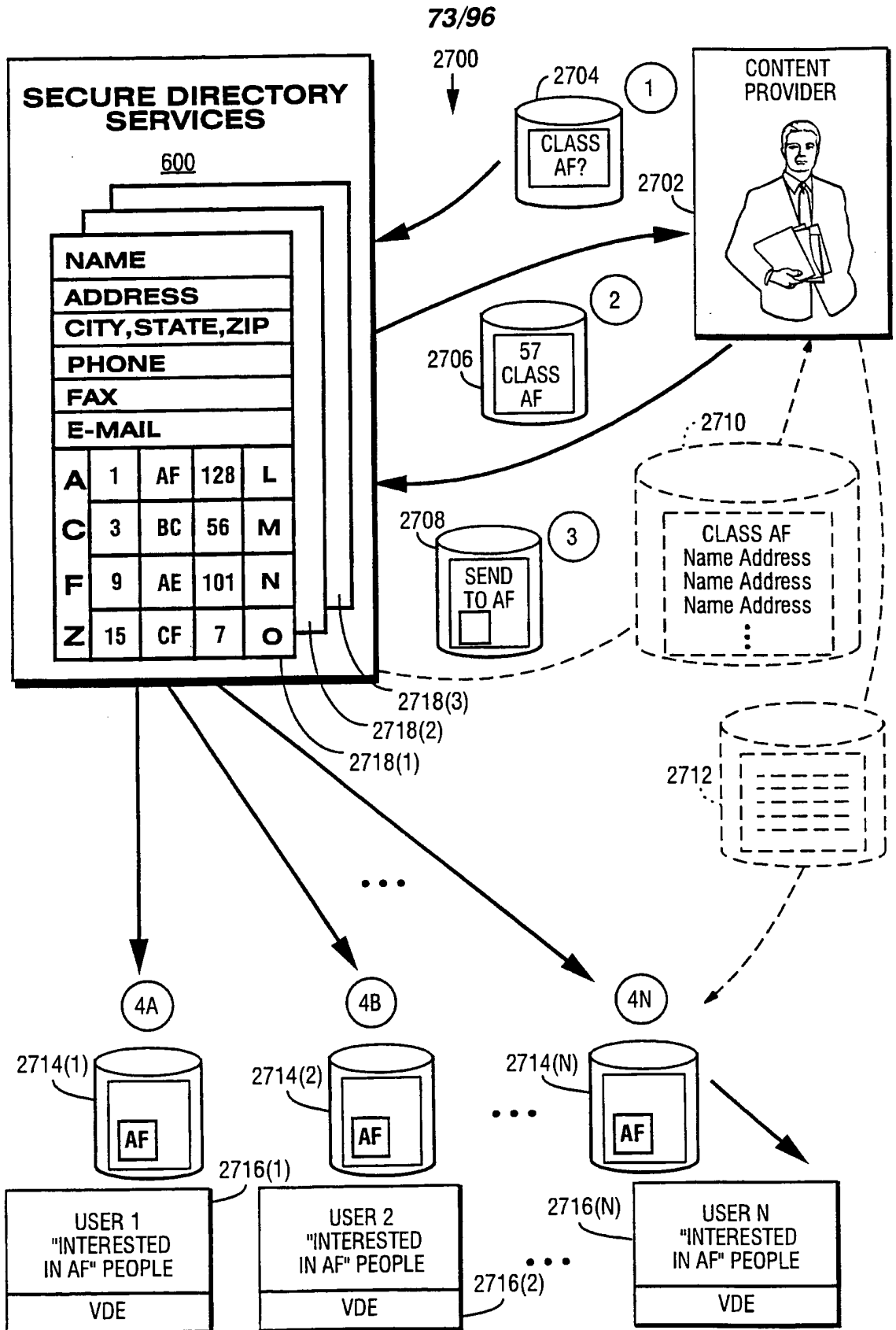


Fig. 52 Secure Directory Services

74/96

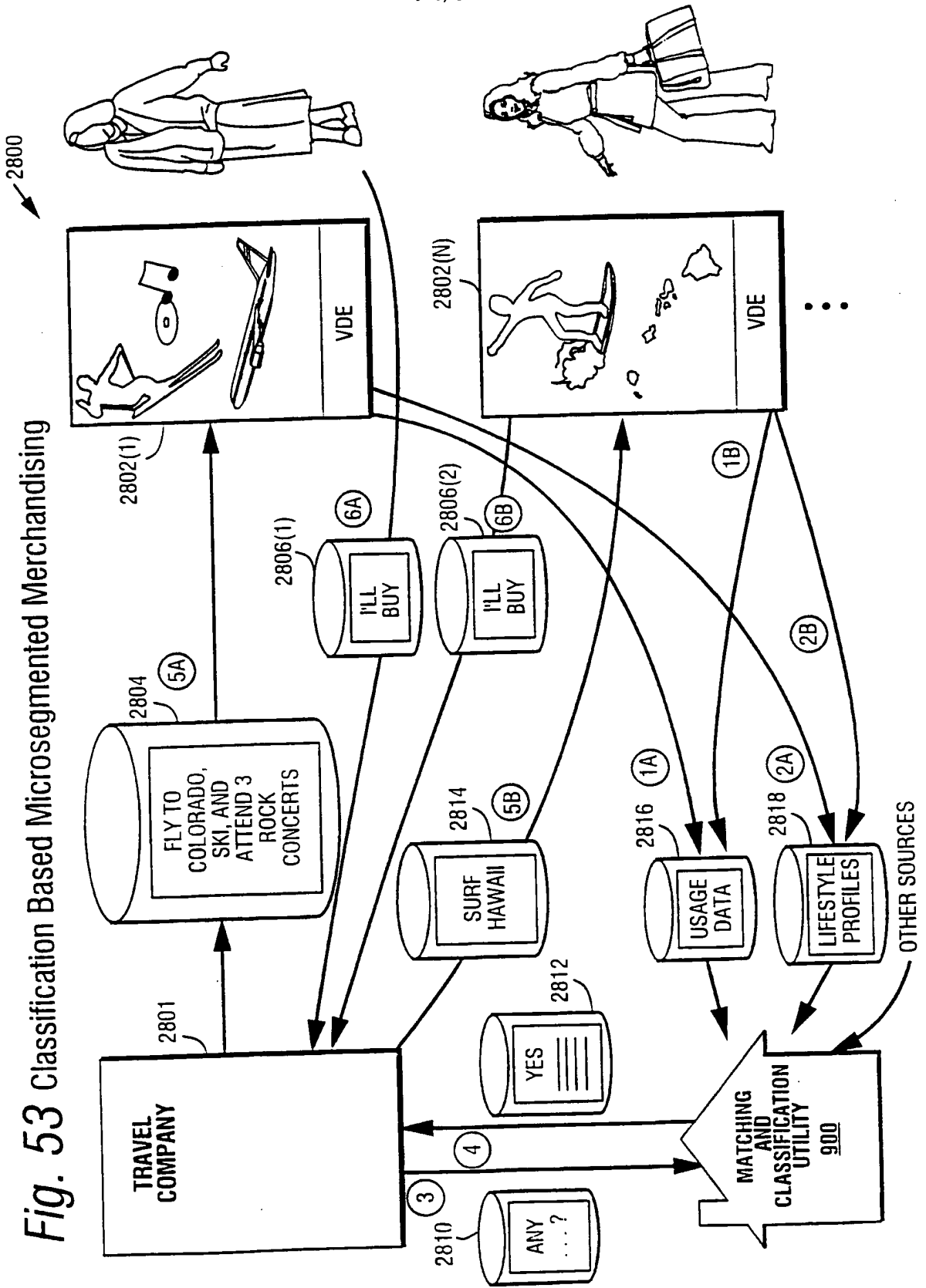


Fig. 53 Classification Based Microsegmented Merchandising

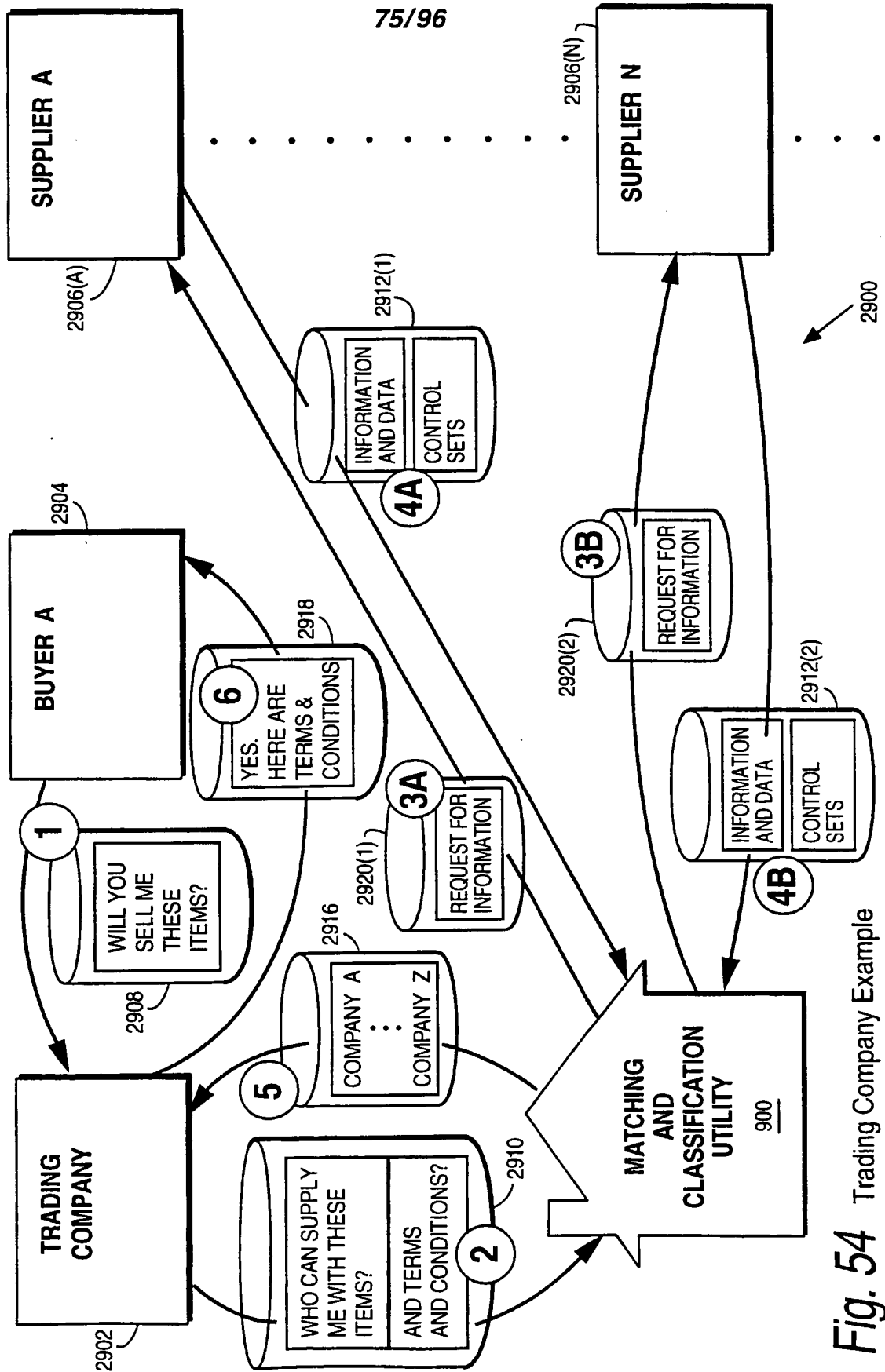


Fig. 54 Trading Company Example

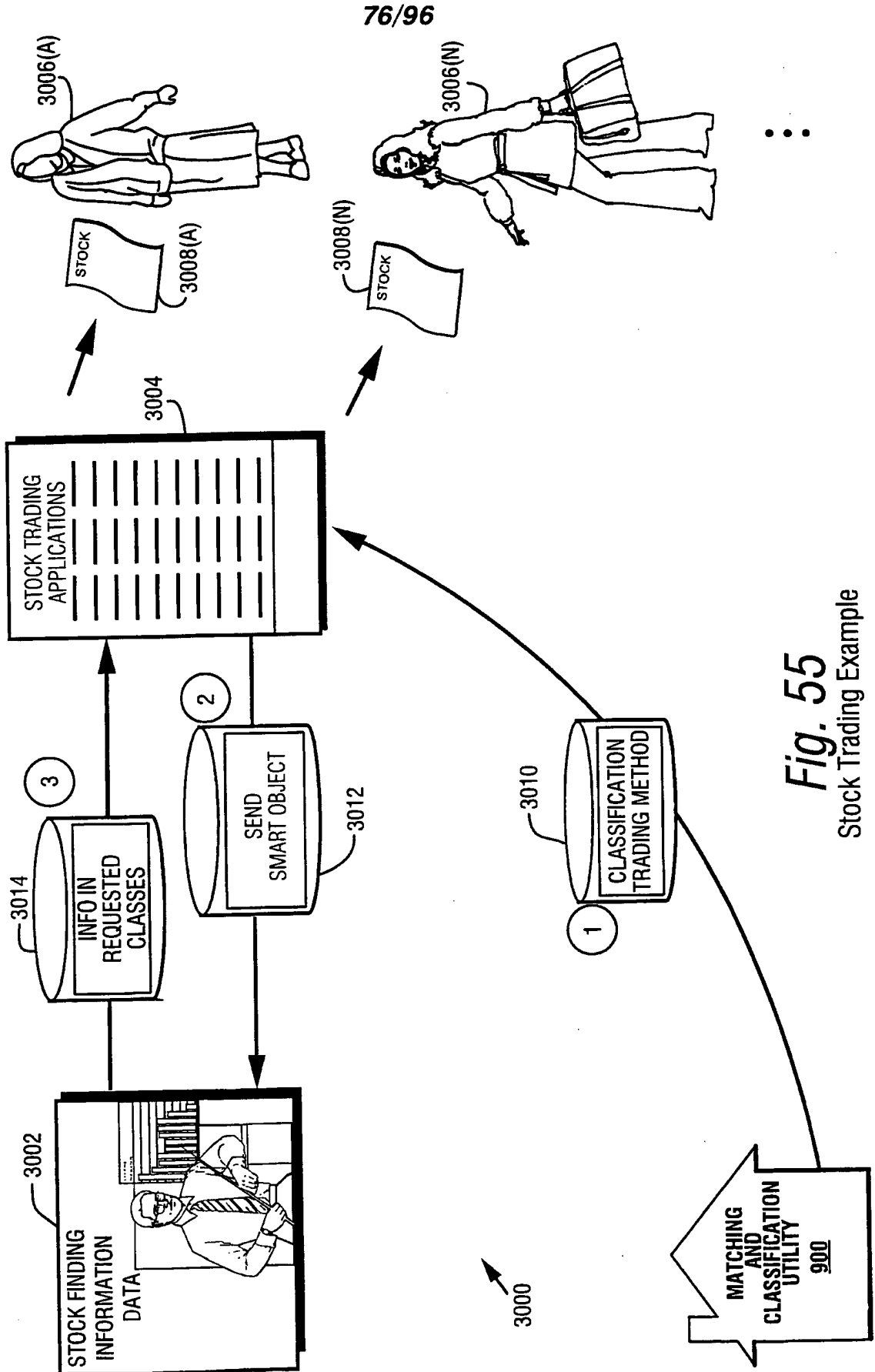


Fig. 55
Stock Trading Example

77/96

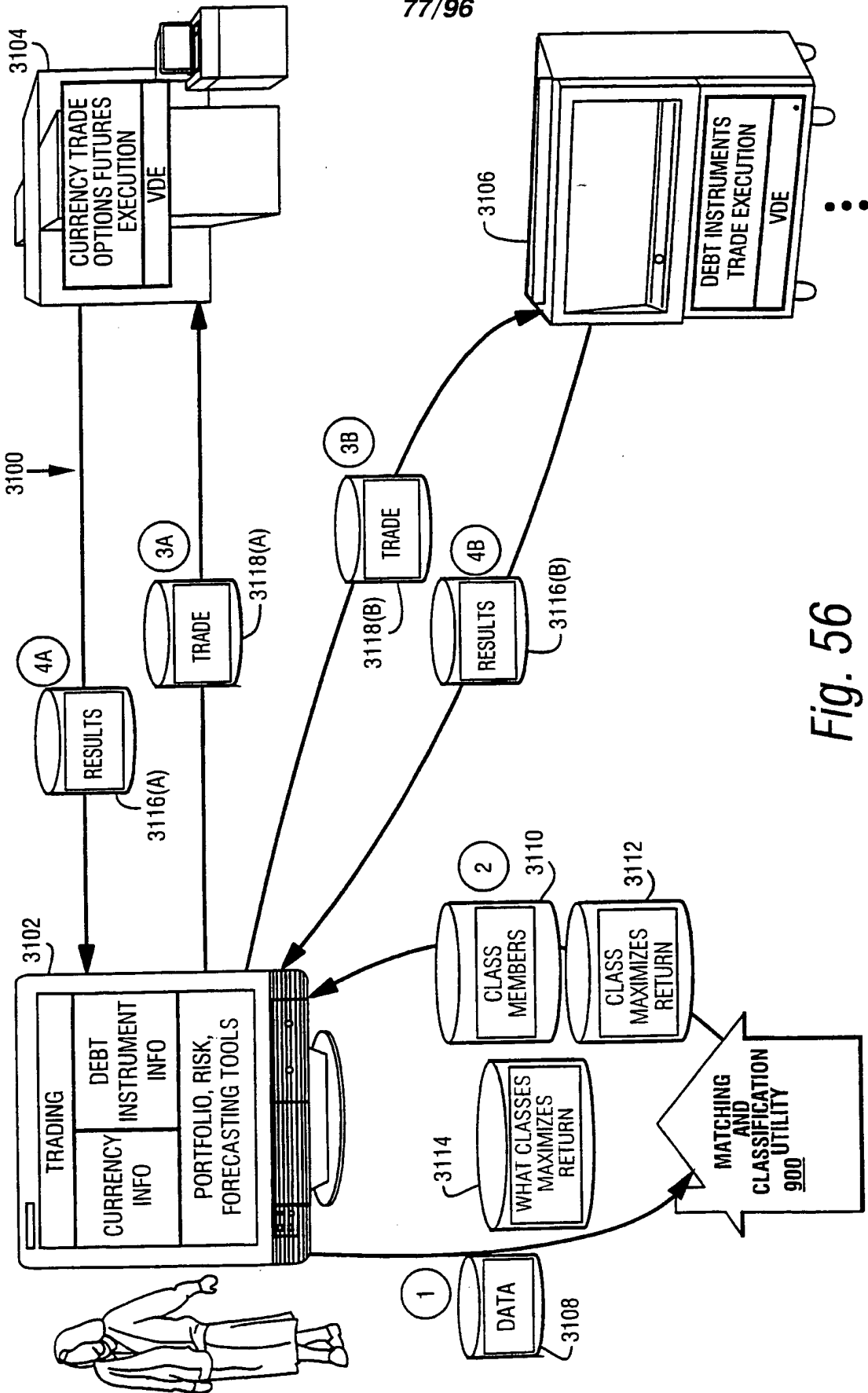


Fig. 56
Currency Trading Example

78/96

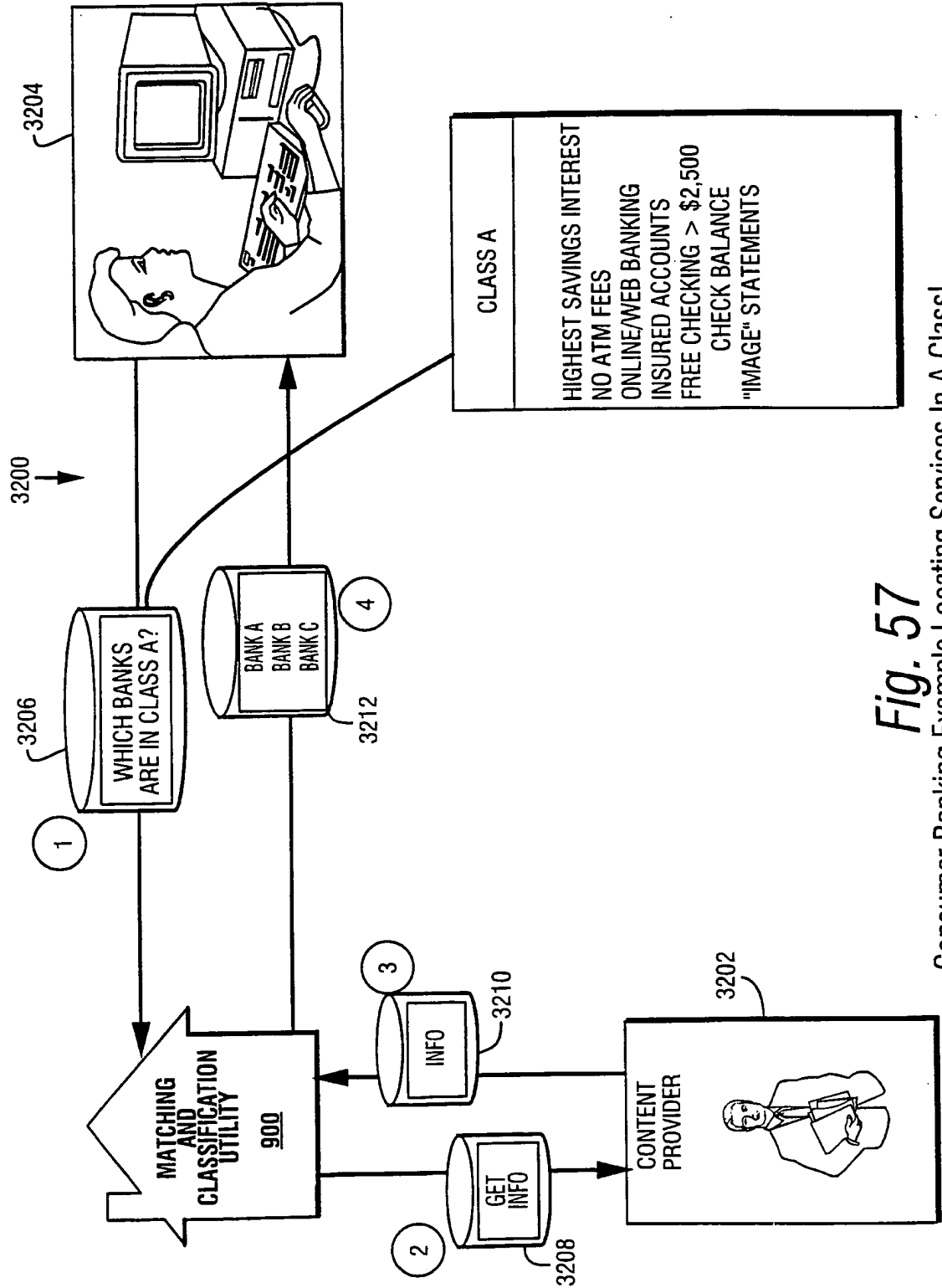


Fig. 57

Consumer Banking Example Locating Services In A Class!

79/96

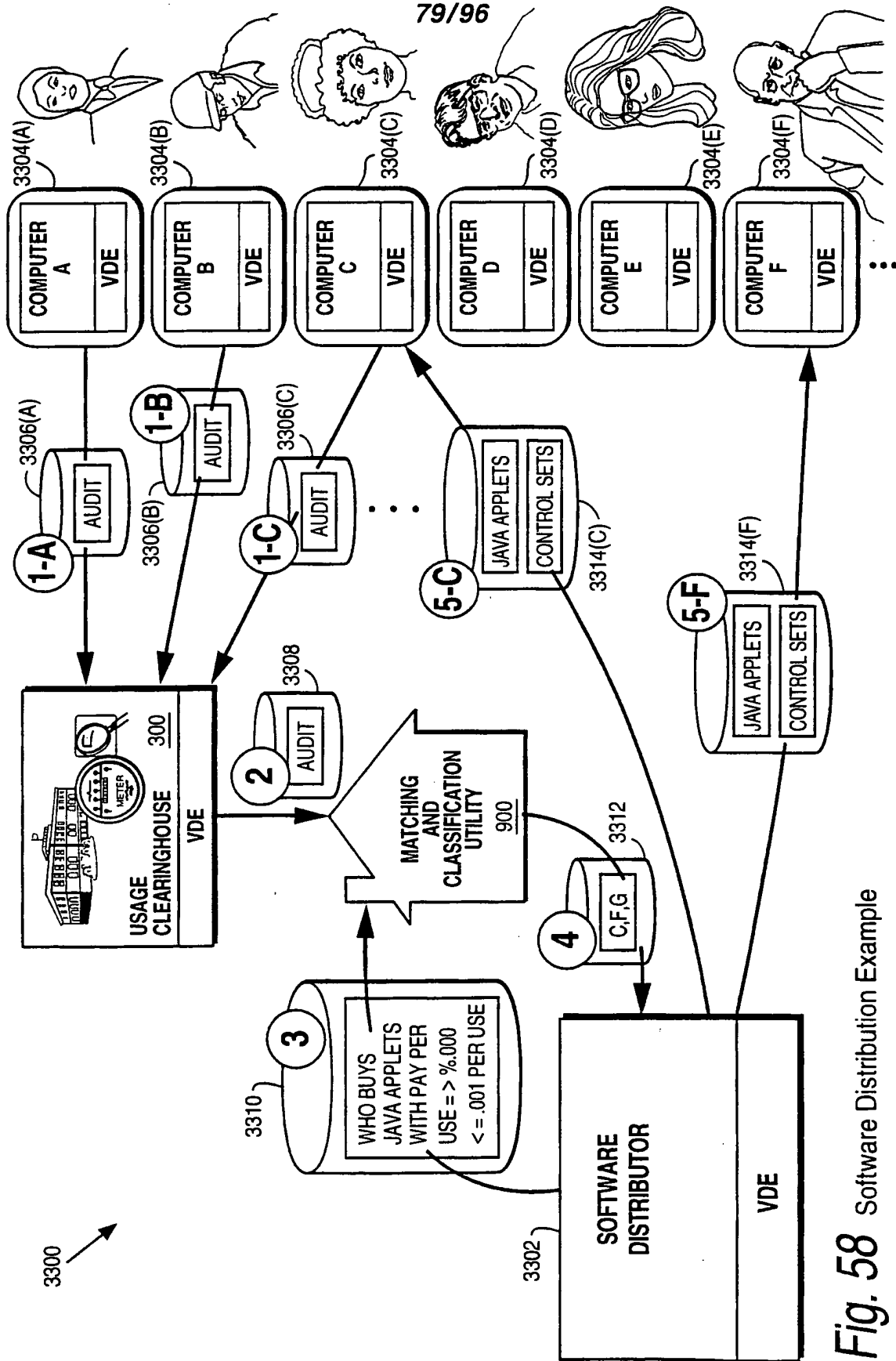


Fig. 58 Software Distribution Example

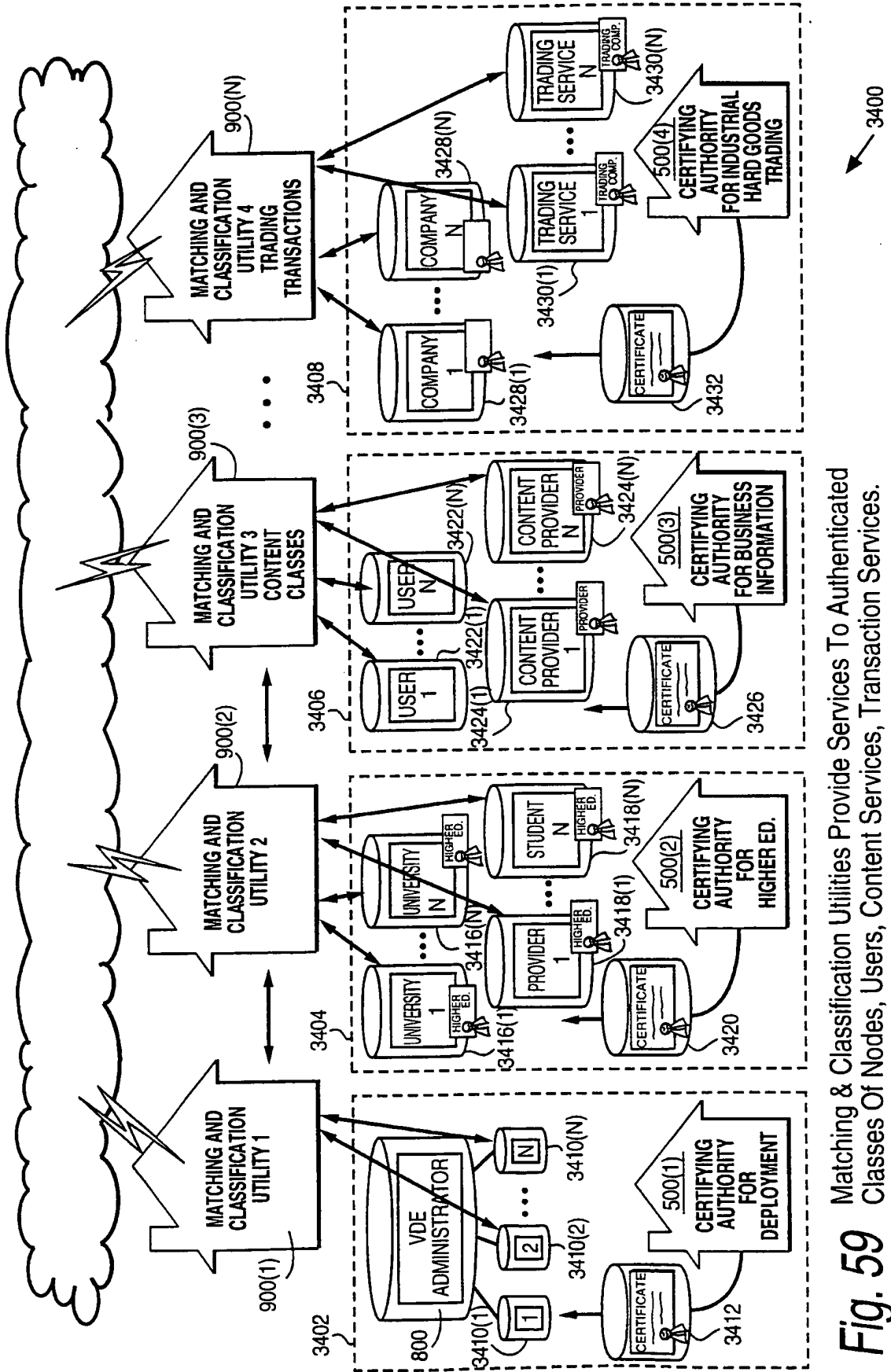
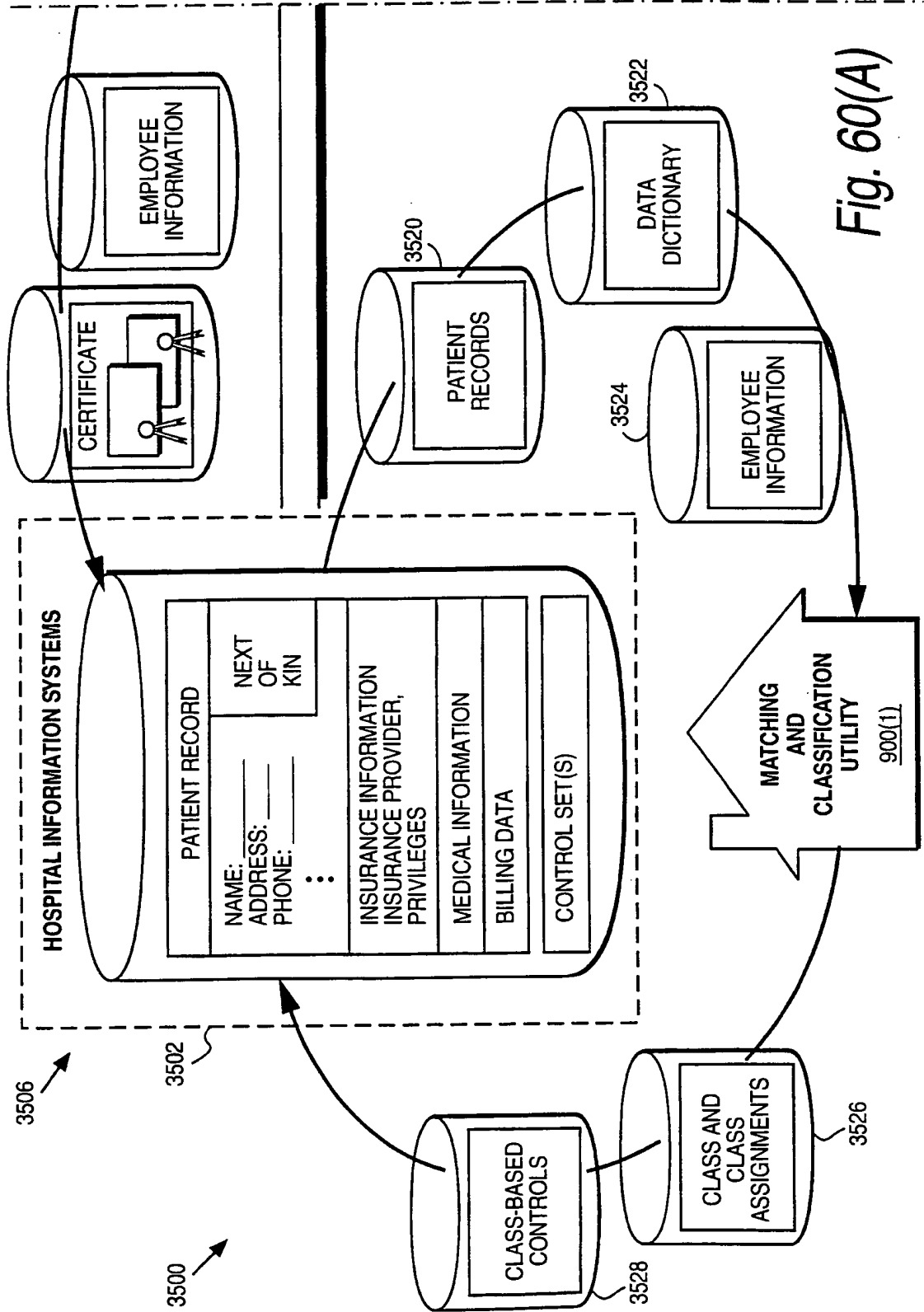


Fig. 59 Matching & Classification Utilities Provide Services To Authenticated Classes Of Nodes, Users, Content Services, Transaction Services.

81/96

TO
FIG.60B

TO
FIG.60B



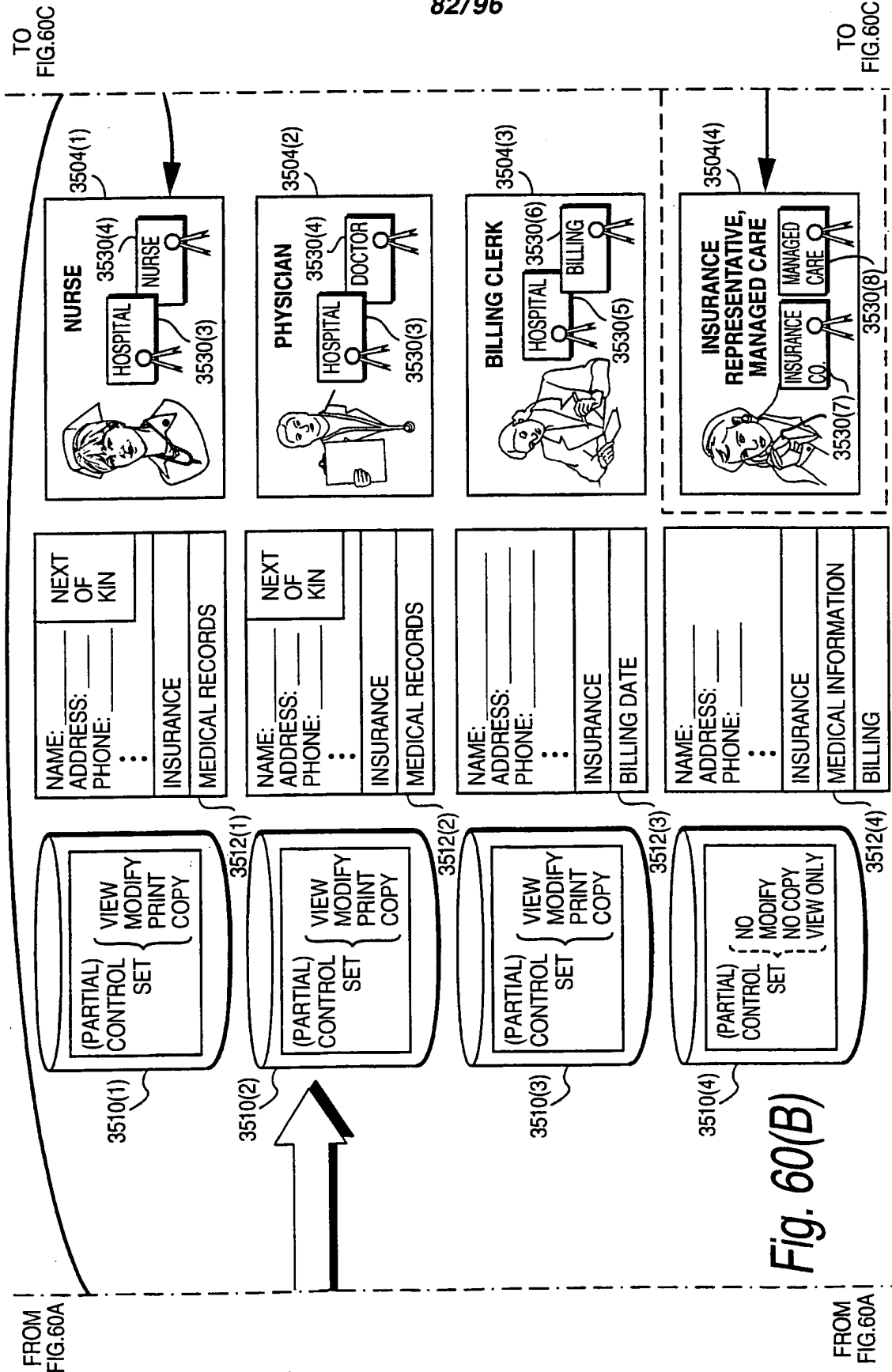


Fig. 60(B)

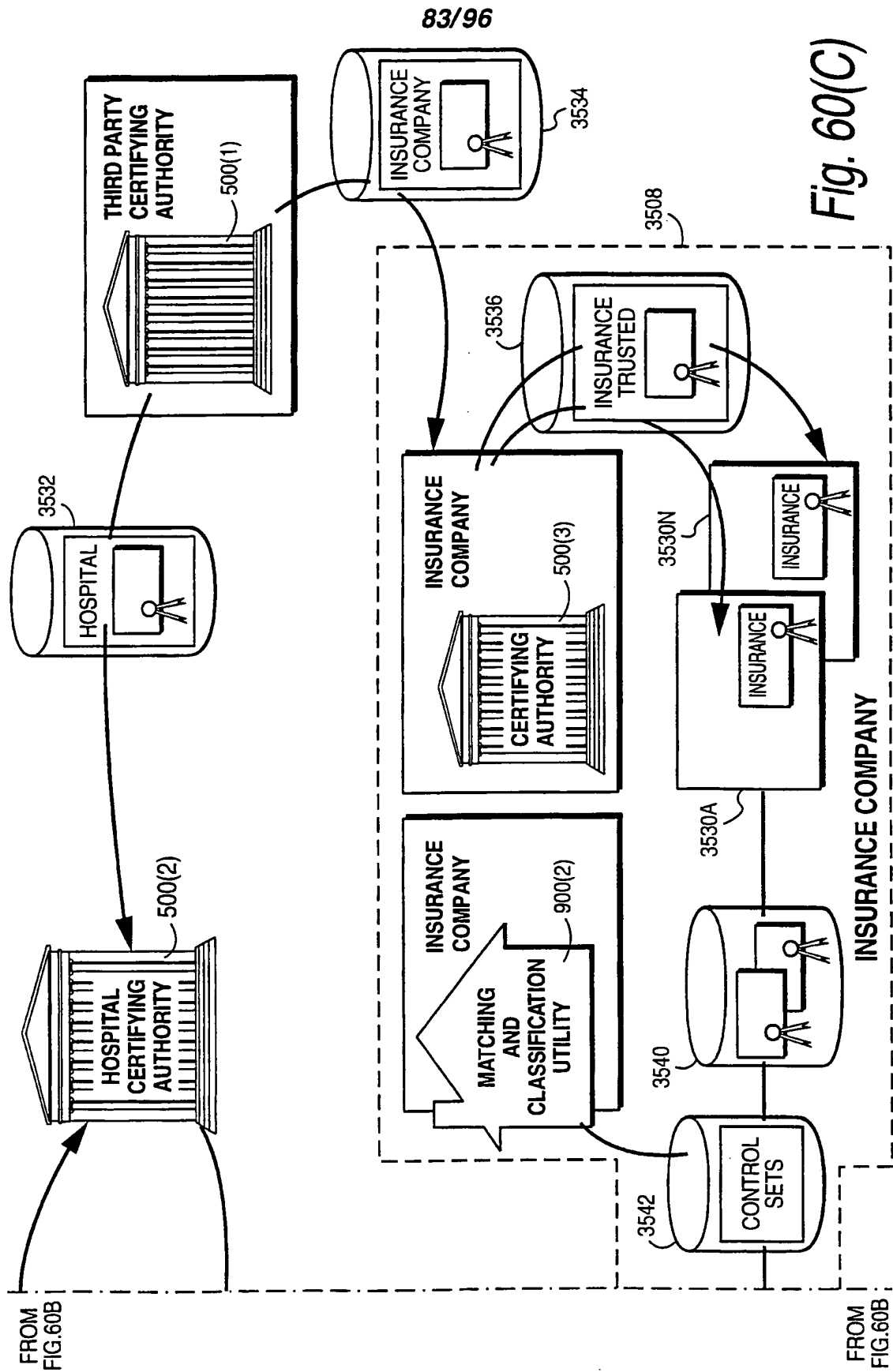


Fig. 60(C)

FROM FIG. 60B

FROM FIG. 60B

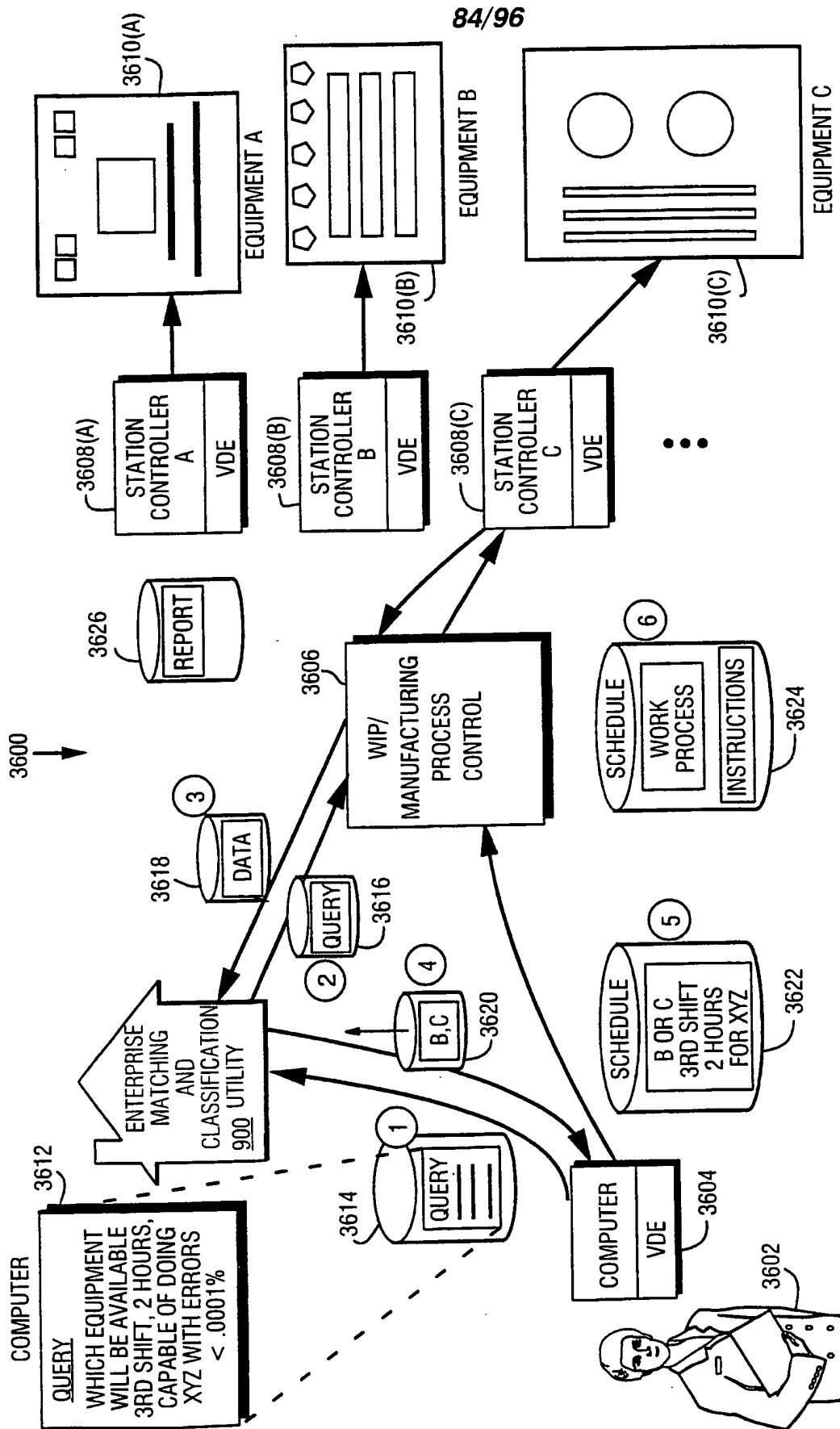


Fig. 61 Workflow Example

85/96

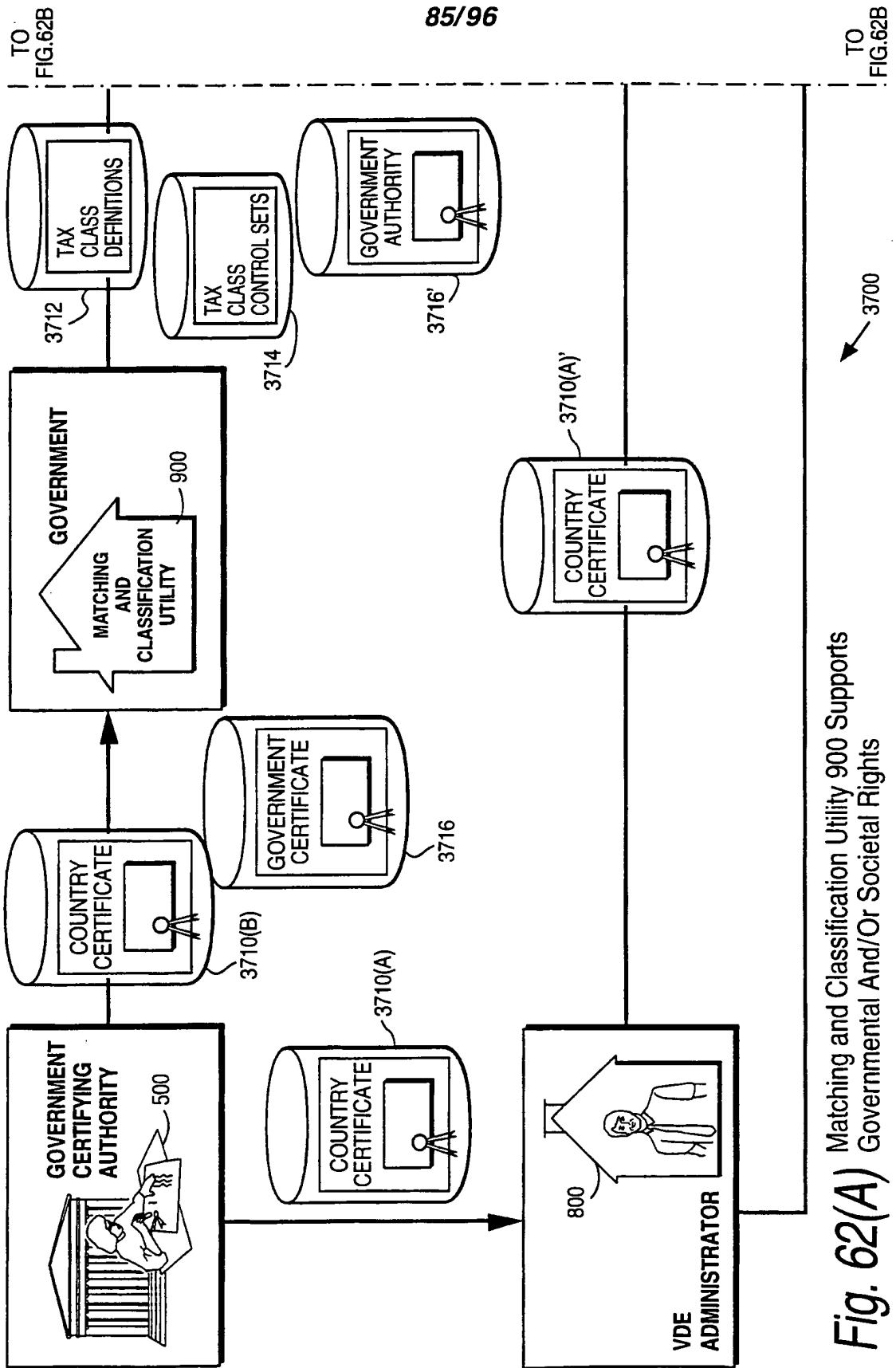


Fig. 62(A) Matching and Classification Utility 900 Supports Governmental And/Or Societal Rights

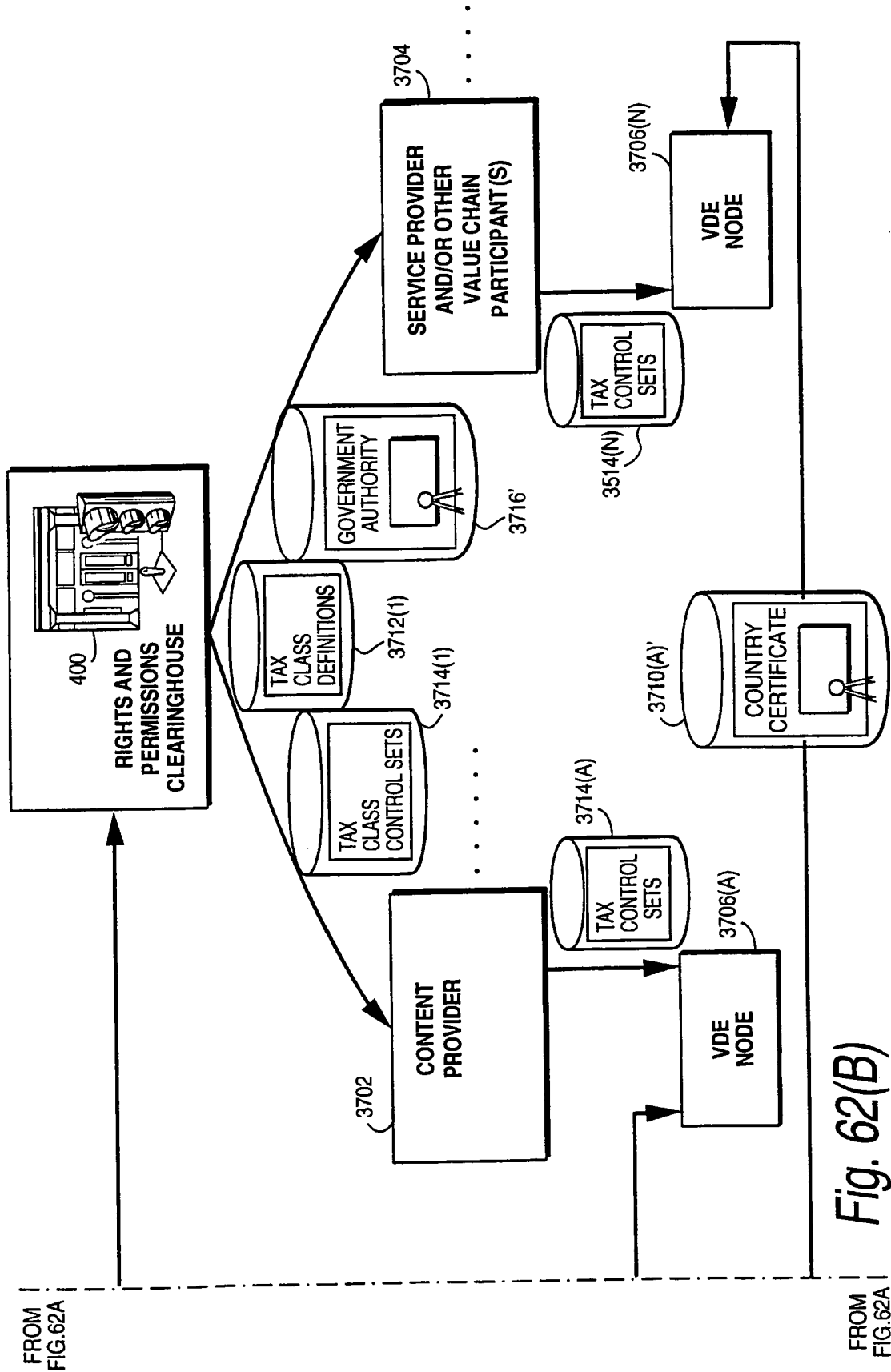


Fig. 62(B)

FROM FIG. 62A

FROM FIG. 62A

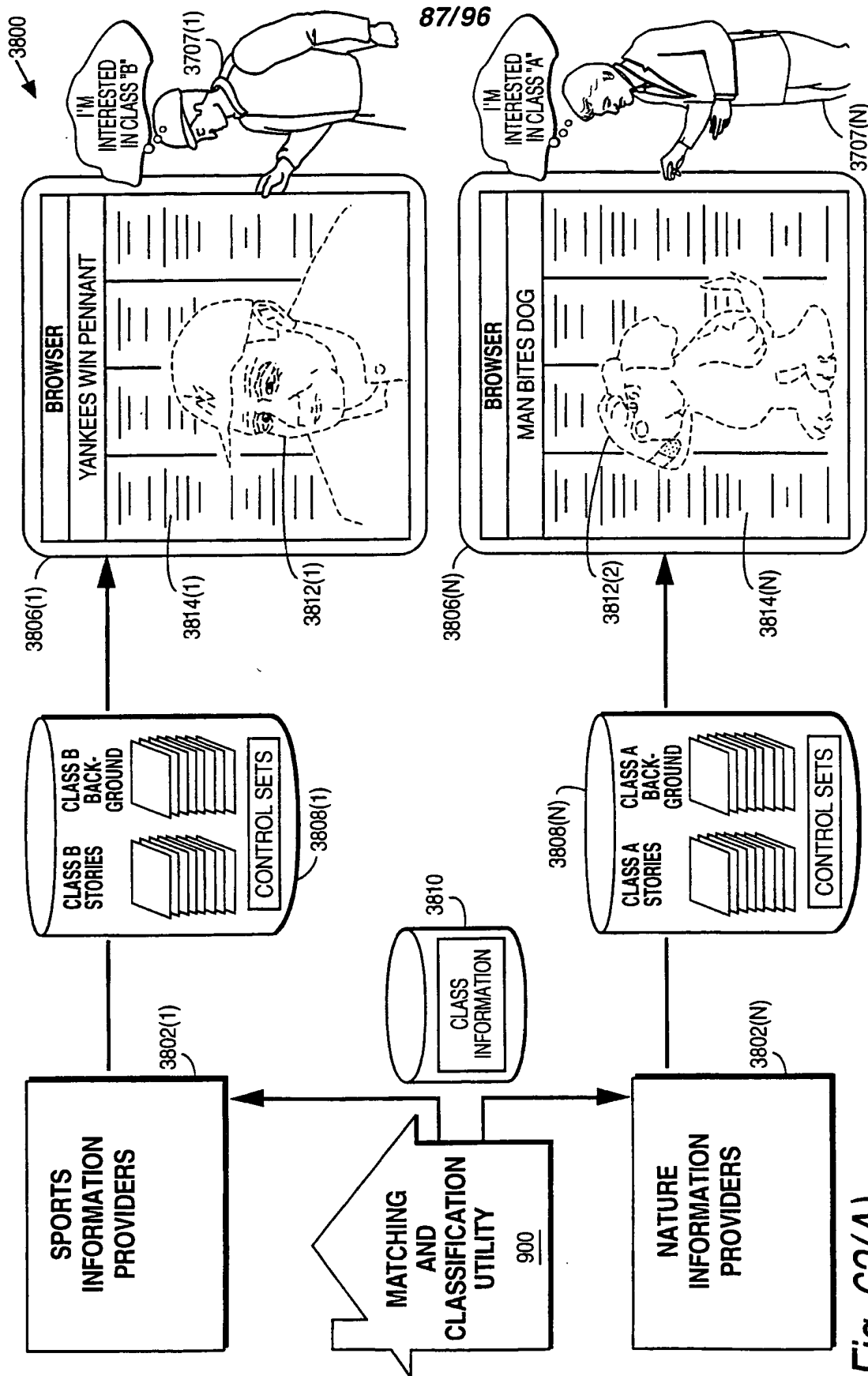


Fig. 63(A) Example Or Use Of Classification To Effect Presentation Of Information.

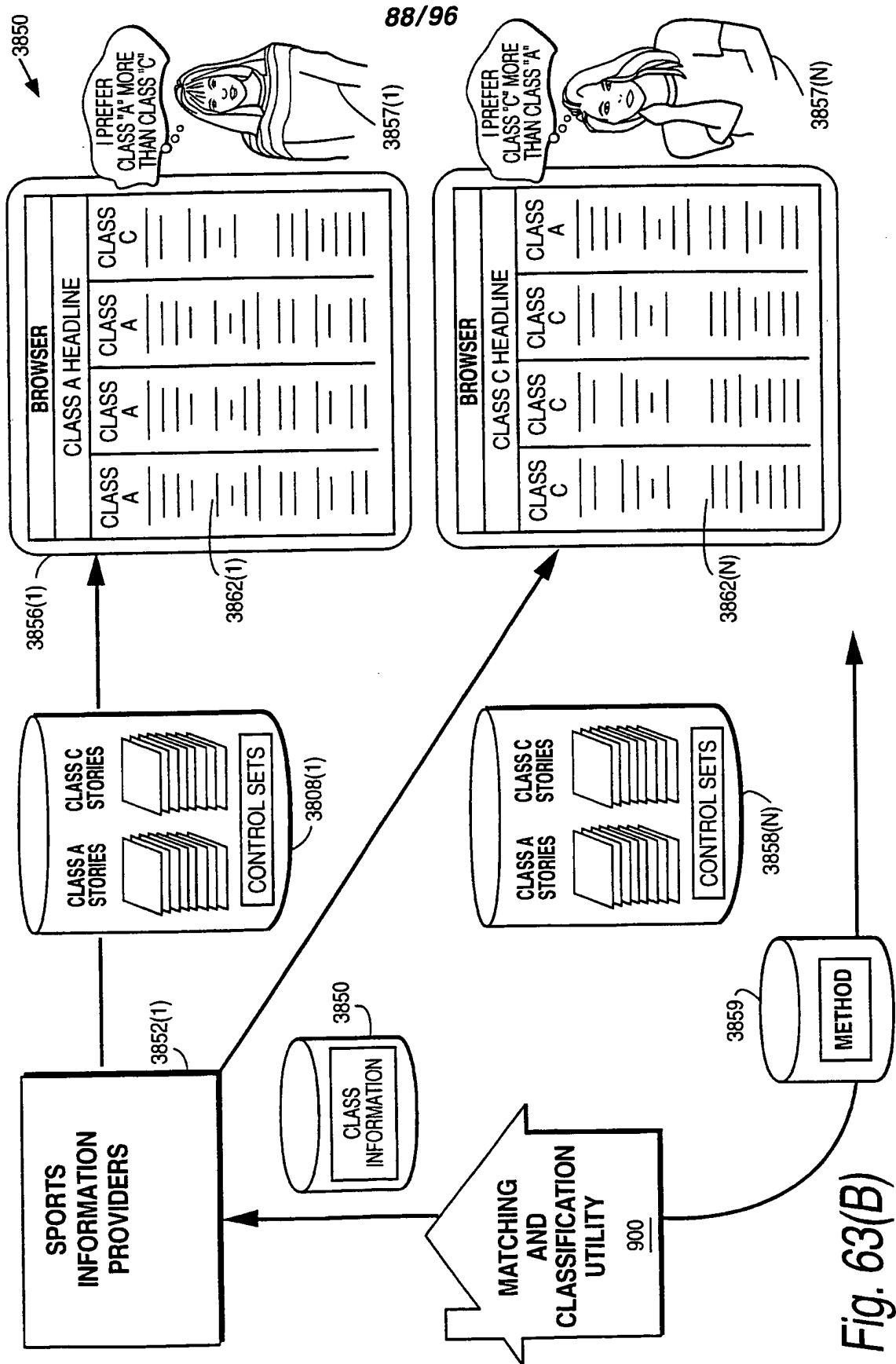


Fig. 63(B)

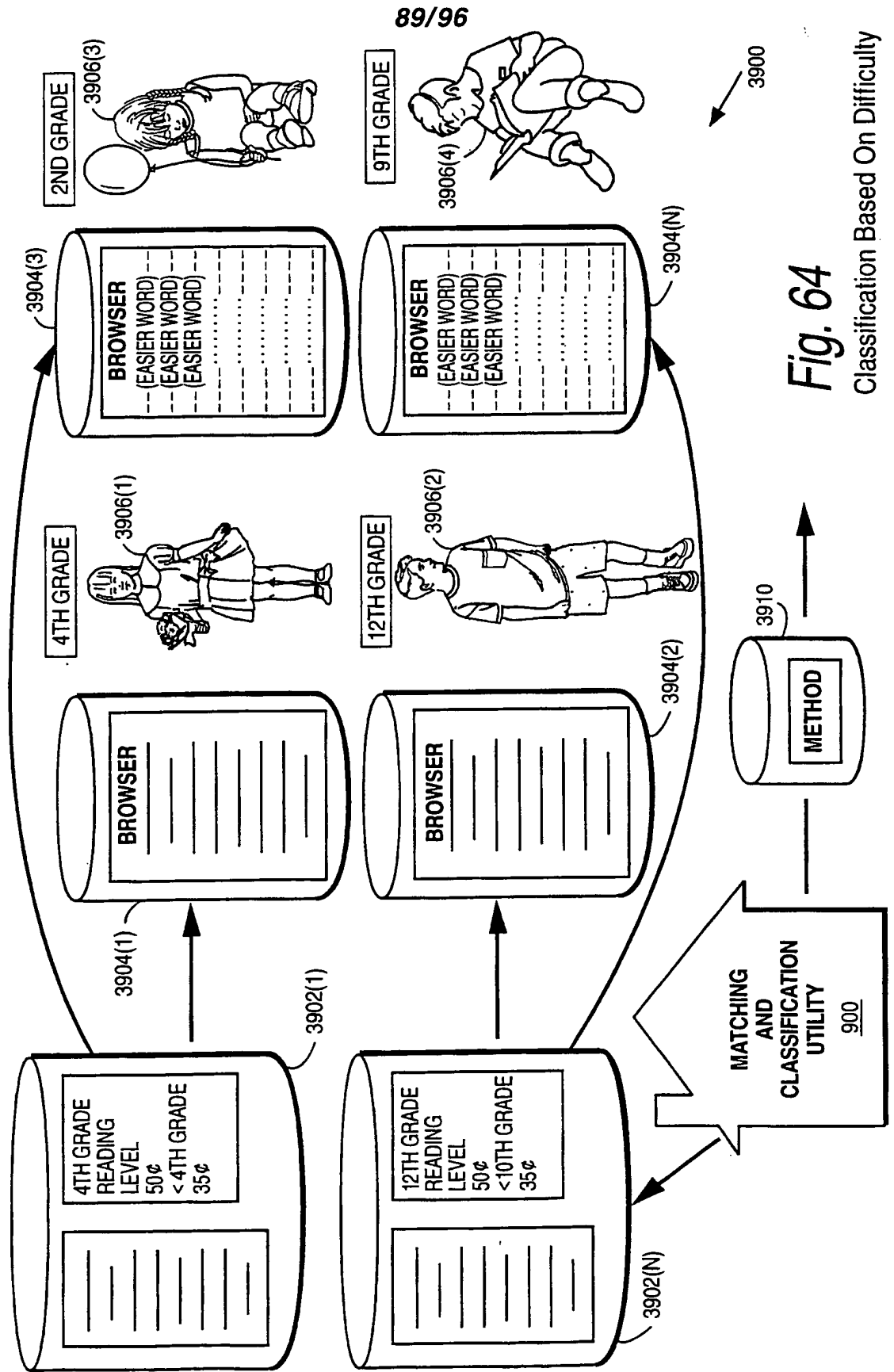


Fig. 64

Classification Based On Difficulty

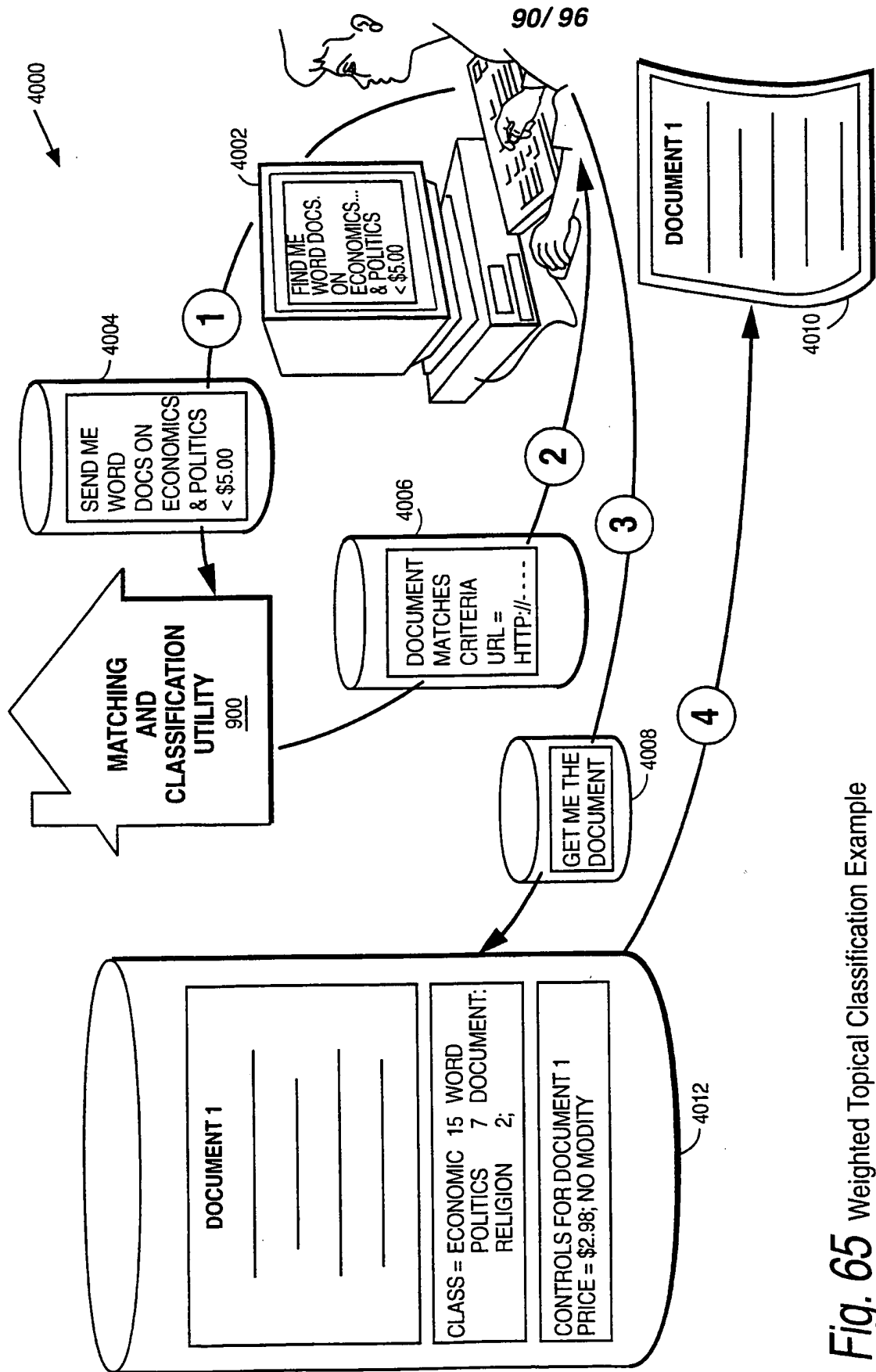


Fig. 65 Weighted Topical Classification Example

91/96

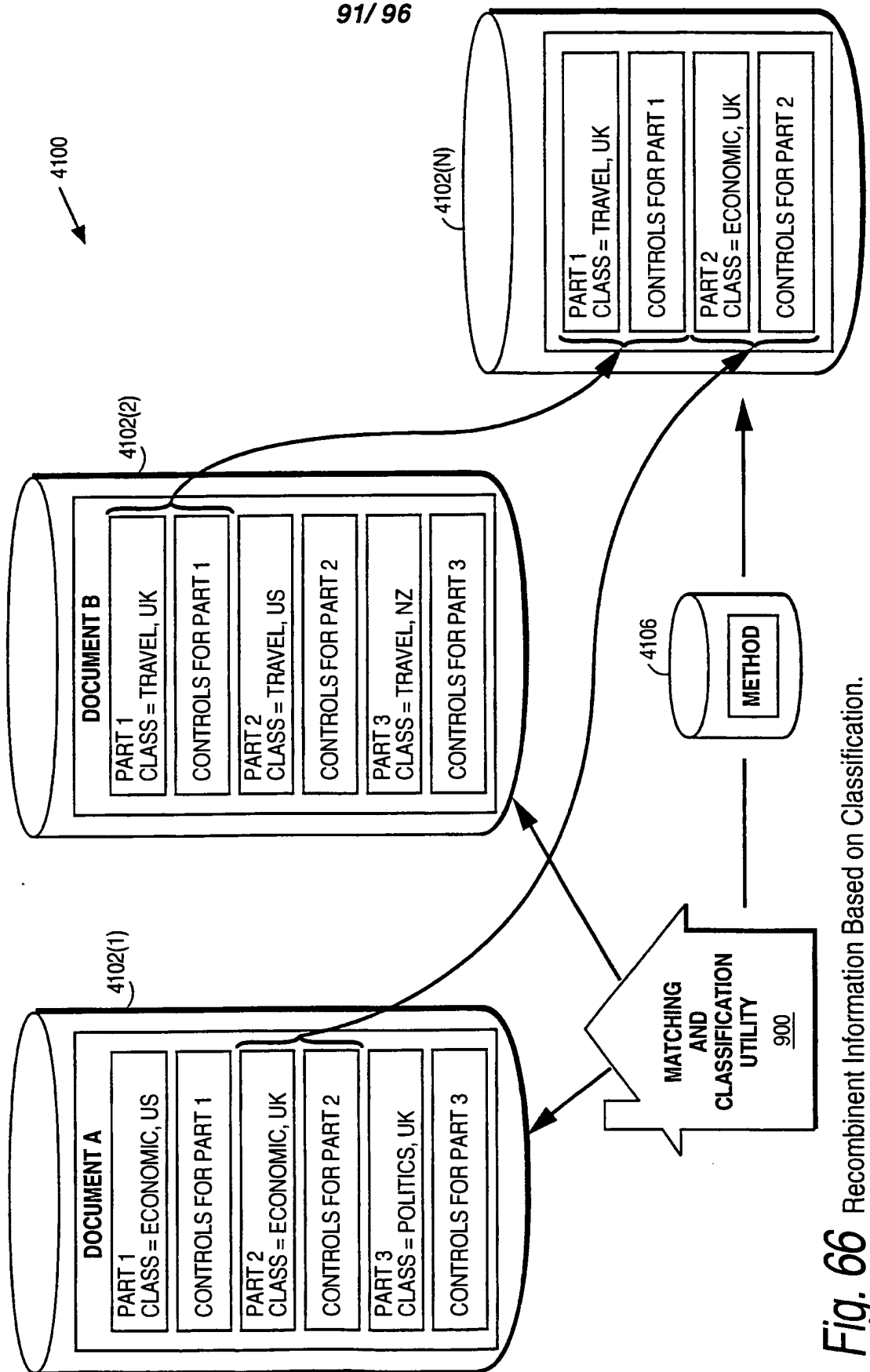


Fig. 66 Recombinant Information Based on Classification.

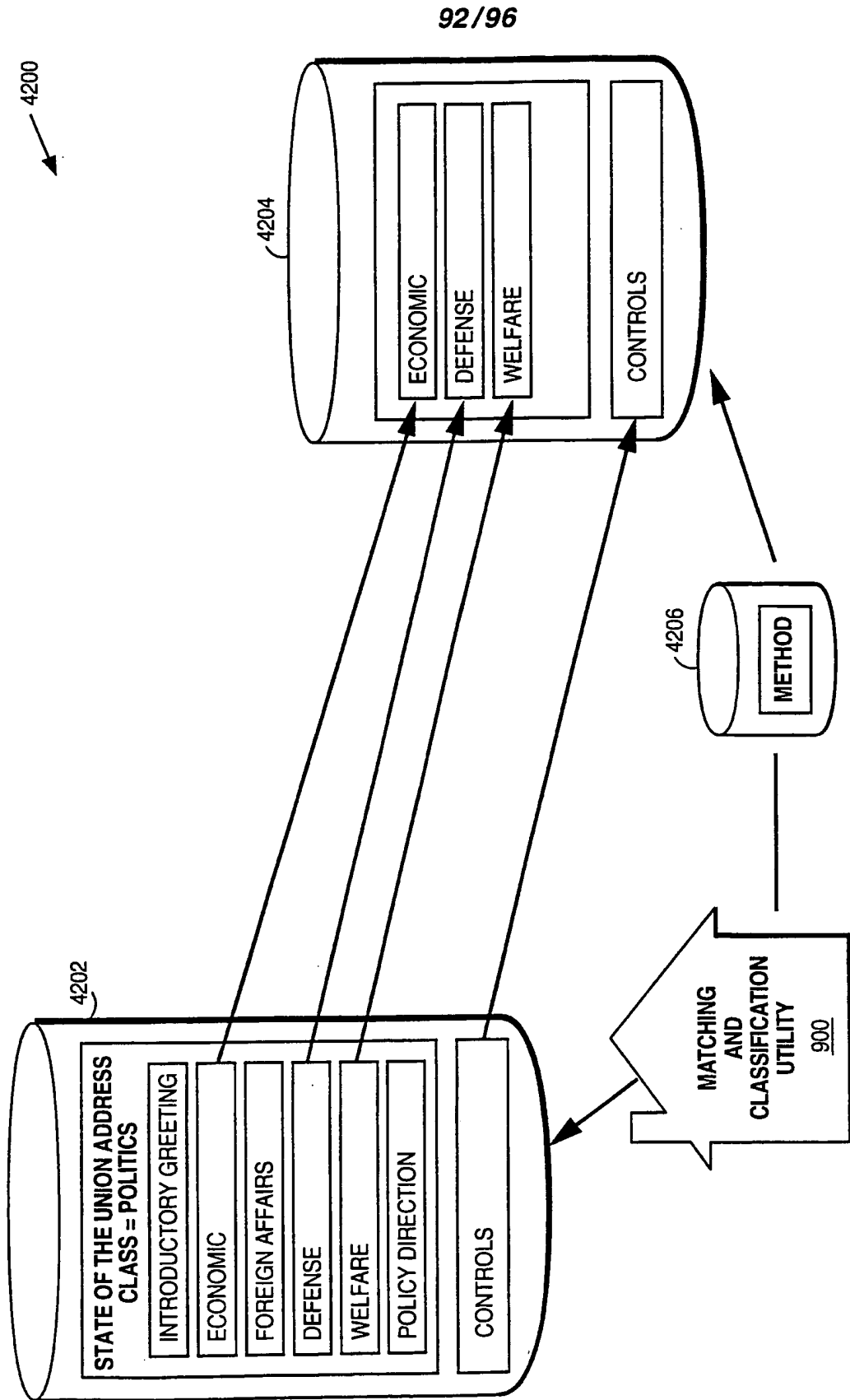
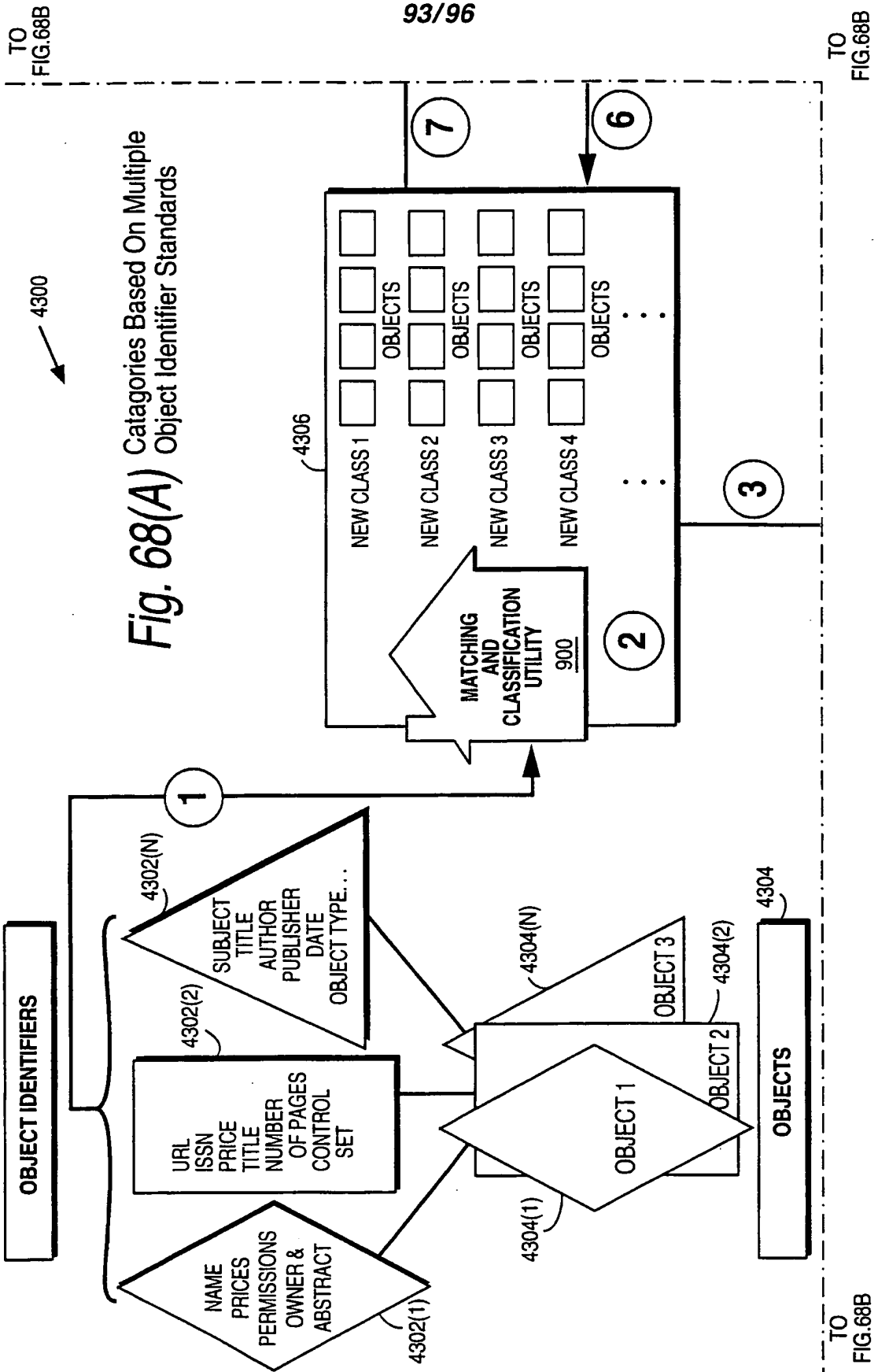


Fig. 67 Nested Classification.



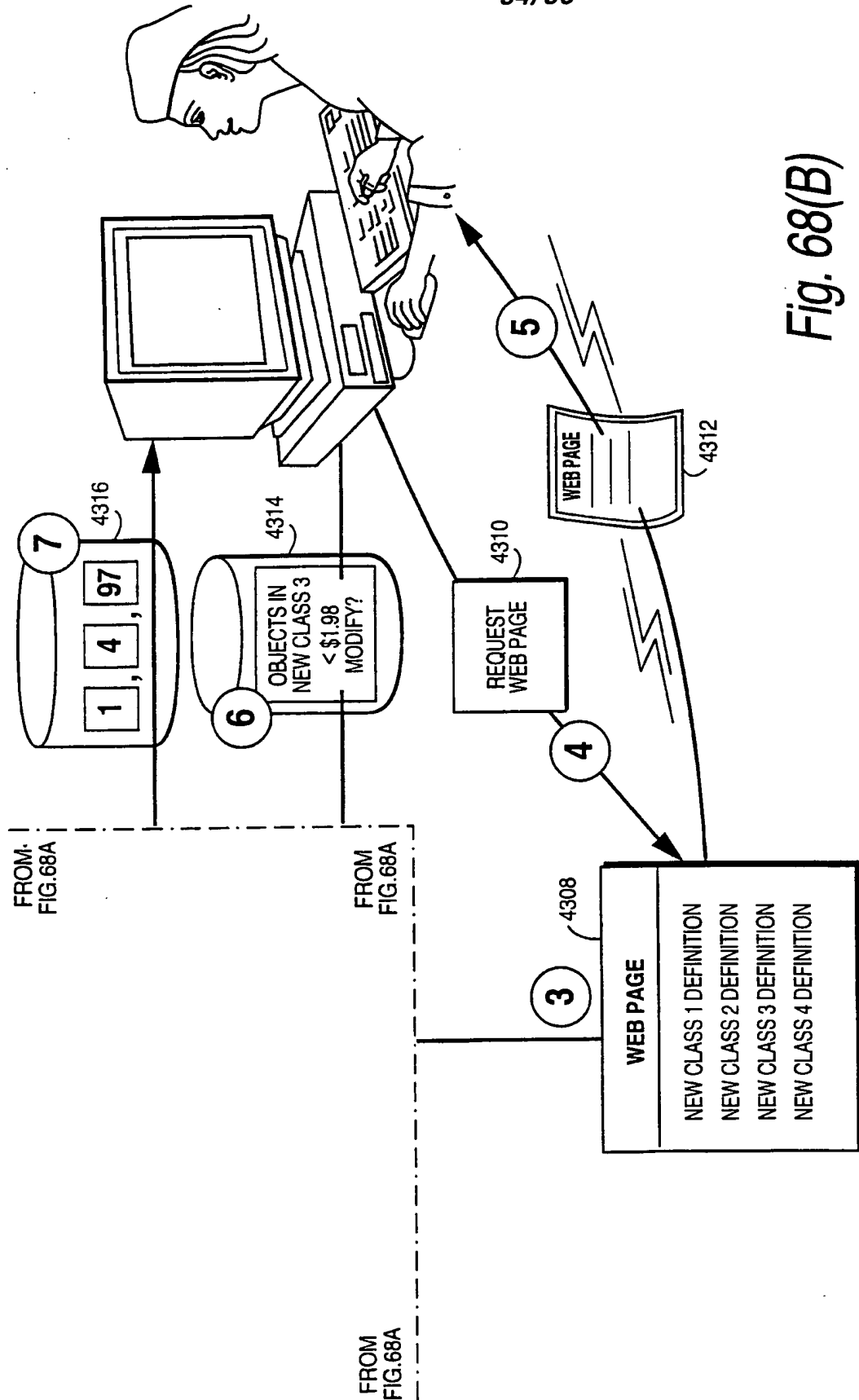


Fig. 68(B)

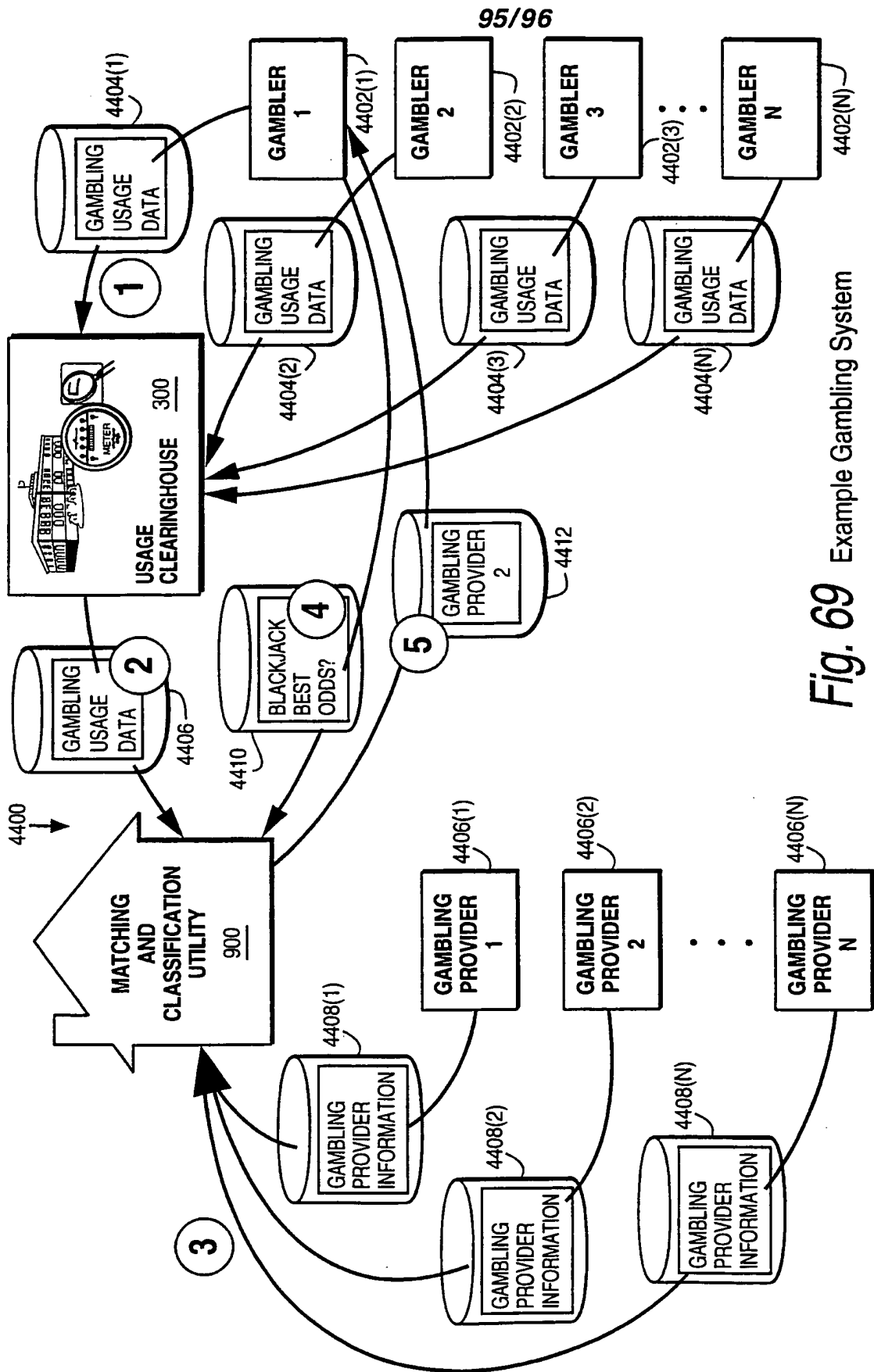


Fig. 69 Example Gambling System

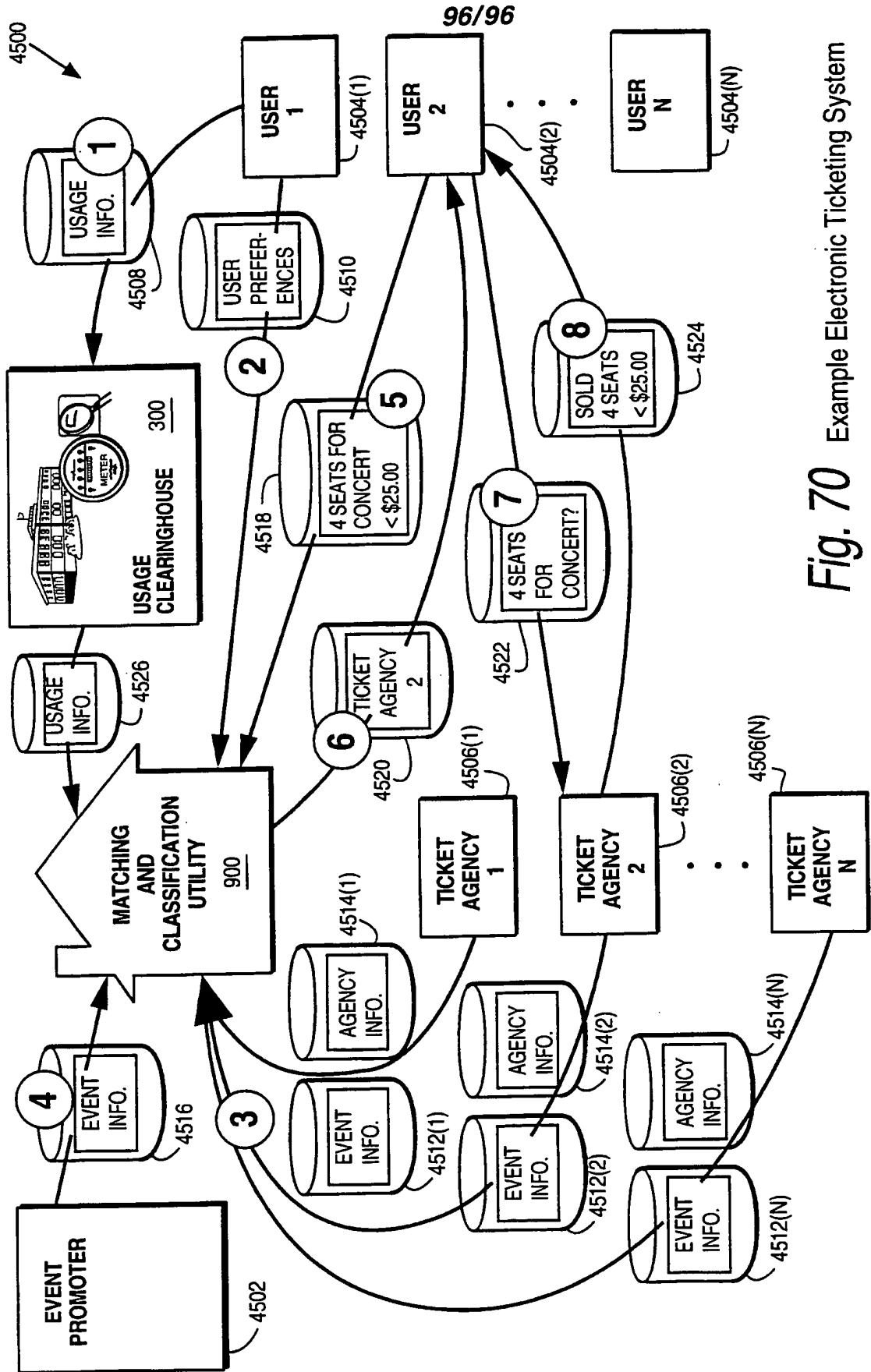


Fig. 70 Example Electronic Ticketing System

(12) **EUROPEAN PATENT SPECIFICATION**

- (13)** Date of publication of patent specification: **18.04.90** **(51)** Int. Cl.⁵: **G 06 F 9/46**
(17) Application number: **82302596.0**
(21) Date of filing: **21.05.82**
(18) Divisional applications **88200917, 88200916, 88200921** filed on **09.05.88**.

(54) **Digital data processing system.**

(19) Priority: **22.05.81 US 266413**
22.05.81 US 266539 **22.05.81 US 266414**
22.05.81 US 266521 **22.05.81 US 266532**
22.05.81 US 266415 **22.05.81 US 266403**
22.05.81 US 266409 **22.05.81 US 266408**
22.05.81 US 266424 **22.05.81 US 266401**
22.05.81 US 266421 **22.05.81 US 266524**
22.05.81 US 266404

(20) Date of publication of application:
22.12.82 Bulletin 82/51
(22) Publication of the grant of the patent:
18.04.90 Bulletin 90/16

(23) Designated Contracting States:
AT BE CH DE FR GB IT LI LU NL SE

(24) References cited:
Hasselmeier, Spruth: "Rechnerstrukturen",
1974, pp 75-103
Klar, Wichmann: "Mikroprogrammierung", June
1975, pp 159-163, 176-179, 185-187, 195-205,
214-215

(71) Proprietor: **DATA GENERAL CORPORATION**
Route 9
Westboro Massachusetts 01581 (US)

(72) Inventor: Ahlstrom, John K.
1309 San Damar
Mountain View California 94043 (US)
Inventor: Bachman, Brett L.
214 W. Canton Street Suite 4
Boston Massachusetts 02116 (US)
Inventor: Belgard, Richard A.
21250 Glenmont Drive
Saratoga California 95070 (US)
Inventor: Bernstein, David H.
41 Bay Colony Drive
Ashland Massachusetts 01721 (US)
Inventor: Bratt, Richard Glenn
9 Brook Trail Road
Wayland Massachusetts 01778 (US)
Inventor: Clancy, Gerald F.
13069 Jaccaranda Center
Saratoga California 95070 (US)
Inventor: Farber, David A.
1700 Lakewood Avenue
Durham North Carolina 27707 (US)
Inventor: Gavrin, Edward S.
Beaver Pond Road RFD 4
Lincoln Massachusetts 01773 (US)

Note: Within nine months from the publication of the mention of the grant of the European patent, any person may give notice to the European Patent Office of opposition to the European patent granted. Notice of opposition shall be filed in a written reasoned statement. It shall not be deemed to have been filed until the opposition fee has been paid. (Art. 99(1) European patent convention).

Courier Press, Leamington Spa, England.

EP 0 067 556 B1

⑤ References cited:

IBM TECHNICAL DISCLOSURE BULLETIN, vol. 22, no. 3, august 1979, pages 1286-1289, New York, US D.B. LOMET: "Regions for controlling the propagation of addressability in capability systems"

IBM TECHNICAL DISCLOSURE BULLETIN, vol. 15, no. 9, February 1973, pages 2721-2722 W.L. DUNNE: "Common Compiler/Interpreter for a programming language"

ADVANCES IN COMPUTERS, vol. 14, 1976, Academic Press, New York, US pages 231-272 D.K. HSIAO et al.: "Information Secure Systems"

IEEE DIGEST OF PAPERS FROM COMPCON FALL MEETING, September 10-12, 1974, Washington, Micros and Minis, pages 15-17 Long Beach, US C.J. NEUHAUSER et al.: "Description of an emulation laboratory"

7th Annual Symposium on COMPUTER ARCHITECTURE, May 6-8, 1980, Conference Proceedings, pages 245-252 New York, US V. BERSTIS: "Security and protection of data in the IBM System/38"

VERY LARGE DATA BASES, vol.9, no. 2C, October 1977, Third International Conference Proceedings, pages 507-514 Tokyo, JP D. DOWNS et al.: "A kernel design for a secure data base management system"

Inventor: Gruner, Ronald Hans
112 Dublin Wood Drive
Cary North Carolina 27514 (US)
Inventor: Houseman, David L.
1213 Selwyn Lane
Cary North Carolina 27511 (US)
Inventor: Jones, Thomas M. Jones
300 Reade Road
Chapel Hill North Carolina 27514 (US)
Inventor: Katz, Lawrence H.
10943 S. Forest Ridge Road
Oregon City Oregon 97045 (US)
Inventor: Mundie, Craig James
136 Castewood Drive
Cary North Carolina (US)
Inventor: Pilat, John F.
1308 Ravenhurst Drive
Raleigh North Carolina 27609 (US)
Inventor: Richmond, Michael S.
Fearrington Post Box 51
Pittsboro North Carolina 27312 (US)
Inventor: Schleimer, Stephen I.
1208 Ellen Place
Chapel Hill North Carolina 27514 (US)
Inventor: Wallach, Steven J.
12436 Green Meadow Lane
Saratoga California 95070 (US)
Inventor: Wallach, Walter A., Jr.
1336 Medfield Road
Raleigh North Carolina 27607 (US)
Inventor: Wells, Douglas M.
106 Robin Road
Chapel Hill North Carolina 27514 (US)

⑭ Representative: Pears, David Ashley et al
REDDIE & GROSE 16 Theobalds Road
London WC1X 8PL (GB)

Description

The present invention relates to a digital data processing system and, more particularly, to a multiprocessor digital data processing system suitable for use in a data processing network and having a simplified, flexible user interface and flexible, multileveled internal mechanism.

A general trend in the development of data processing systems has been towards systems suitable for use in interconnected data processing networks. Another trend has been towards data processing systems wherein the internal structure of the system is flexible, protected from users, and effectively invisible to the user and wherein the user is presented with a flexible and simplified interface to the system.

Certain problems and shortcomings affecting the realization of such a data processing system have appeared repeatedly in the prior art and must be overcome to create a data processing system having the above attributes. These prior art problems and limitations include the following topics.

First, the data processing systems of the prior art have not provided a system wide addressing system suitable for use in common by a large number of data processing systems interconnected into a network. Addressing systems of the prior art have not provided sufficiently large address spaces and have not allowed information to be permanently and uniquely identified. Prior addressing systems have not made provisions for information to be located and identified as to type or format, and have not provided sufficient granularity. In addition, prior addressing systems have reflected the physical structure of particular data processing systems. That is, the addressing systems have been dependent upon whether a particular computer was, for example, an 8, 16, 32, 64 or 128 bit machine. Since prior data processing systems have incorporated addressing mechanisms wherein the actual physical structure of the processing system is apparent to the user, the operations a user could perform have been limited by the addressing mechanisms. In addition, prior processor systems have operated as fixed word length machines, further limiting user operations.

Prior data processing systems have not provided effective protection mechanisms preventing one user from effecting another user's data and programs without permission. Such protection mechanisms have not allowed unique, positive identification of users requesting access to information, or of information, nor have such mechanisms been sufficiently flexible in operation. In addition, access rights have pertained to the users rather than to the information, so that control of access rights has been difficult. Finally, prior art protection mechanisms have allowed the use of "Trojan Horse arguments". That is, users not having access rights to certain information have been able to gain access to that information through another user or procedure having such access rights.

Yet another problem of the prior art is that of providing a simple and flexible user's interface to a data processing system. The character of user's interface to a data processing system is determined, in part, by the means by which a user refers to and identifies operands and procedures of the user's programs and by the instruction structure of the system. Operands and procedures are customarily referred to and identified by some form of logical address having points of reference, and validity, only within a user's program. These addresses must be translated into logical and physical addresses within a data processing system each time a program is executed, and must then be frequently retranslated or generated during execution of a program. In addition, a user must provide specific instructions as to data format and handling. As such reference to operands or procedures typically comprise a major portion of the instruction stream of the user's program and requires numerous machine translations and operations to implement. A user's interface to a conventional system is thereby complicated, and the speed of execution of programs reduced, because of the complexity of the program references to operands and procedures.

A data processing system's instruction structure includes both the instructions for controlling system operations and the means by which these instructions are executed. Conventional data processing systems are designed to efficiently execute instructions in one or two user languages, for example, FORTRAN or COBOL. Programs written in any other language are not efficiently executable. In addition, a user is often faced with difficult programming problems when using any high level language other than the particular one or two languages that a particular conventional system is designed to utilize.

Yet another problem in conventional data processing systems is that of protecting the system's internal mechanisms, for example, stack mechanisms and internal control mechanisms, from accidental or malicious interference by a user.

Finally, the internal structure and operation of prior art data processing systems have not been flexible, or adaptive, in structure and operation. That is, the internal structure and operation of prior systems have not allowed the systems to be easily modified or adapted to meet particular data processing requirements. Such modifications may include changes in internal memory capacity, such as the addition or deletion of special purpose subsystems, for example, floating point or array processors. In addition, such modifications have significantly affected the users interface with the system. Ideally, the actual physical structure and operation of the data processing system should not be apparent at the user interface.

It has already been proposed (IBM Technical Disclosure Bulletin Vol. 22 No. 3 Aug. 1979 pp 1286—1289) to maintain such a large address space that every object which is ever created can have a unique identifier. This requires a very large identifier field, e.g. 40 to 50 bits.

The object of the present invention is to implement such a concept so that it may be applied across many computers geographically distributed and without requiring all computers to use the same

programming language.

The system according to the invention is defined in the appended claims.

It is known in virtual address machines to use name tables to provide logical addresses for translation to physical addresses (Klar, Wichmann, "Mikroprogrammierung" June 1975, especially pp. 159—163, 176—179, 185—187, 195—205, 214—215) and this reference also discusses the emulation in software of different target machines.

More specifically, the embodiment of the invention described in detail below provides a data processing system suitable for use in interconnected data processing networks, which internal structure is flexible, protected from users, effectively invisible to users, and provides a flexible and simplified interface to users. The data processing system provides an addressing mechanism allowing permanent and unique identification of all information generated for use in or by operation of the system, and an extremely large address space which is accessible to and common to all such data processing systems. The addressing mechanism provides addresses which are independent of the physical configuration of the system and allow information to be completely identified, with a single address, to the bit granular level and with regard to information type or format. The present invention further provides a protection mechanism wherein variable access rights are associated with individual bodies of information. Information, and users requesting access to information, are uniquely identified through the system addressing mechanism. The protection mechanism also prevents use of Trojan Horse arguments. And, the present invention provides an instruction structure wherein high level user language instructions are transformed into dialect coded, uniform, intermediate level instructions to provide equal facility of execution for a plurality of user languages. Another feature is the provision of an operand reference mechanism wherein operands are referred to in user's programs by uniform format names which are transformed, by an internal mechanism transparent to the user, into addresses. The present invention additionally provides multilevel control and stack mechanisms protecting the system's internal mechanism from interference by users. Yet another feature is a data processing system having a flexible internal structure capable of performing multiple, concurrent operations and comprised of a plurality of separate, independent processors. Each such independent processor has a separate microinstruction control and at least one separate and independent port to a central communications and memory node. The communications and memory node is also an independent processor having separate and independent microinstruction control. The memory processor is internally comprised of a plurality of independently operating, microinstruction controlled processors capable of performing multiple, concurrent memory and communications operations. The present invention also provides further data processing system structural and operational features for implementing the above features.

It is thus advantageous to incorporate the present invention into a data processing system because the present invention provides addressing mechanisms suitable for use in large interconnected data processing networks. Additionally, the present invention is advantageous in that it provides an information protection mechanism suitable for use in large, interconnected data processing networks. The present invention is further advantageous in that it provides a simplified, flexible, and more efficient interface to a data processing system. The present invention is yet further advantageous in that it provides a data processing system which is equally efficient with any user level language by providing a mechanism for referring to operands in user programs by uniform format names and instruction structure incorporating dialect coded, uniform format intermediate level instructions. Additionally, the present invention protects data processing system internal mechanisms from user interference by providing multilevel control and stack mechanisms. The present invention is yet further advantageous in providing a flexible internal system structure capable of performing multiple, concurrent operations, comprising a plurality of separate, independent processors, each having a separate microinstruction control and at least one separate and independent port to a central, independent communications and memory processor comprised of a plurality of independent processors capable of performing multiple, concurrent memory and communications operations.

Other advantages and features of the present invention will be understood by those of ordinary skill in the art, after referring to the following detailed description of the preferred embodiments and drawings wherein.

BRIEF DESCRIPTION OF DRAWINGS

- Fig. 1 is a partial block diagram of a computer system incorporating the present invention;
 Fig. 2 is a diagram illustrating computer system addressing structure of the present invention;
 Fig. 3 is a diagram illustrating the computer system instruction stream of the present invention;
 Fig. 4 is a diagram illustrating the control structure of a conventional computer system;
 Fig. 4A is a diagram illustrating the control structure of a computer system incorporating the present invention;
 Fig. 5 — Fig. A1 inclusive are diagrams all relating to the present invention;
 Fig. 5 is a diagram illustrating a stack mechanism;
 Fig. 6 is a diagram illustrating procedures, procedure objects, processes, and virtual processors;
 Fig. 7 is a diagram illustrating operating levels and mechanisms of the present computer;
 Fig. 8 is a diagram illustrating a physical implementation of processes and virtual processors;

EP 0 067 556 B1

- Fig. 9 is a diagram illustrating a process and process stack objects;
Fig. 10 is a diagram illustrating operation of macrostacks and secure stacks;
Fig. 11 is a diagram illustrating detailed structure of a stack;
Fig. 12 is a diagram illustrating a physical descriptor;
5 Fig. 13 is a diagram illustrating the relationship between logical pages and frames in a memory storage space;
Fig. 14 is a diagram illustrating access control to objects;
Fig. 15 is a diagram illustrating virtual processors and virtual processor swapping;
Fig. 16 is a partial block diagram of an I/O system of the present computer system;
10 Fig. 17 is a diagram illustrating operation of a ring grant generator;
Fig. 18 is a partial block diagram of a memory system;
Fig. 19 is a partial block diagram of a fetch unit of the present computer system;
Fig. 20 is a partial block diagram of an execute unit of the present computer system;
Fig. 101 is a more detailed partial block diagram of the present computer system;
15 Fig. 102 is a diagram illustrating certain information structures and mechanisms of the present computer system;
Fig. 103 is a diagram illustrating process structures;
Fig. 104 is a diagram illustrating a macrostack structure;
Fig. 105 is a diagram illustrating a secure stack structure;
20 Figs. 106 A, B, and C are diagrams illustrating the addressing structure of the present computer system;
Fig. 107 is a diagram illustrating addressing mechanisms of the present computer system;
Fig. 108 is a diagram illustrating a name table entry;
Fig. 109 is a diagram illustrating protection mechanisms of the present computer system;
25 Fig. 110 is a diagram illustrating instruction and microinstruction mechanism of the present computer system;
Fig. 201 is a detailed block diagram of a memory system;
Fig. 202 is a detailed block diagram of a fetch unit;
Fig. 203 is a detailed block diagram of an execute unit;
30 Fig. 204 is a detailed block diagram of an I/O system;
Fig. 205 is a partial block diagram of a diagnostic processor system;
Fig. 206 is a diagram illustrating assembly of Figs. 201—205 to form a detailed block diagram of the present computer system;
Fig. 207 is a detailed block diagram of a memory interface controller;
35 Fig. 209 is a diagram of a memory to I/O system port interface;
Fig. 210 is a diagram of a memory operand port interface;
Fig. 211 is a diagram of a memory instruction port interface;
Fig. 230 is a detailed block diagram of memory field interface unit logic;
Fig. 231 is a diagram illustrating memory format manipulation operations;
40 Fig. 238 is a detailed block diagram of fetch unit offset multiplexer;
Fig. 239 is a detailed block diagram of fetch unit bias logic;
Fig. 240 is a detailed block diagram of a generalized four way, set associative cache representing name cache, protection cache, and address translation unit;
Fig. 241 is a detailed block diagram of portions of computer system instruction and microinstruction
45 control logic;
Fig. 242 is a detailed block diagram of portions of computer system microinstruction control logic;
Fig. 243 is a detailed block diagram of further portions of computer system microinstruction control logic;
Fig. 244 is a diagram illustrating computer system states of operation;
50 Fig. 245 is a diagram illustrating computer system states of operation for a trace trap request;
Fig. 246 is a diagram illustrating computer system states of operation for a memory repeat interrupt;
Fig. 247 is a diagram illustrating priority level and masking of computer system events;
Fig. 248 is a detailed block diagram of event logic.
Fig. 249 is a detailed block diagram of microinstruction control store logic;
55 Fig. 251 is a diagram illustrating a return control word stack word;
Fig. 252 is a diagram illustrating machine control words;
Fig. 253 is a detailed block diagram of a register address generator;
Fig. 254 is a block diagram of interval and egg timers;
Fig. 255 is a detailed block diagram of execute unit control logic;
60 Fig. 257 is a detailed block diagram of execute unit multiplier data paths and memory;
Fig. 260 is a diagram illustrating operation of an execute unit command queue load and interface to a fetch unit;
Fig. 261 is a diagram illustrating operation of an execute unit operand buffer load and interface to a fetch unit;
65 Fig. 262 is a diagram illustrating operation of an execute unit storeback or transfer of results and

interface to a fetch unit;

Fig. 263 is a diagram illustrating operation of an execute unit check test condition and interface to a fetch unit;

6 Fig. 264 is a diagram illustrating operation of an execute unit exception test and interface to a fetch unit;

Fig. 265 is a block diagram of an execute unit arithmetic operation stack mechanism;

Fig. 266 is a diagram illustrating execute unit and fetch unit interrupt handshaking and interface;

Fig. 267 is a diagram illustrating execute unit and fetch unit interface and operation for nested interrupts;

10 Fig. 268 is a diagram illustrating execute unit and fetch unit interface and operation for loading an execute unit control store;

Fig. 269 is a detailed block diagram and illustration of operation of an I/O system ring grant generator;

Fig. 270 is a detailed block diagram of a fetch unit micromachine of the present computer system;

Fig. 271 is a diagram illustrating a logical descriptor;

15 Fig. 272 is a diagram illustrating use of fetch unit stack registers;

Fig. 273 is a diagram illustrating structures controlling event invocations;

Fig. 301 is a diagram illustrating pointer formats;

Fig. 302 is a diagram illustrating an associated address table;

Fig. 303 is a diagram illustrating a namespace overview of a procedure object;

20 Fig. 304 is a diagram illustrating name table entries;

Fig. 305 is a diagram illustrating an example of name resolution;

Fig. 306 is a diagram illustrating name cache entries;

Fig. 307 is a diagram illustrating translation of S-interpreter universal identifiers to dialect numbers;

Fig. 401 is a diagram illustrating operating systems and system resources;

25 Fig. 402 is a diagram illustrating multiprocess operating systems;

Fig. 403 is a diagram illustrating an extended operating system and a kernel operating system;

Fig. 404 is a diagram illustrating an EOS view of objects;

Fig. 405 is a diagram illustrating pathnames to universal identifier translation;

Fig. 406 is a diagram illustrating universal identifier detail;

30 Fig. 407 is a diagram illustrating address translation with an address translation unit, a memory hash table, and a memory;

Fig. 408 is a diagram illustrating hashing in an active subject table;

Fig. 409 is a diagram illustrating logical allocation units and objects;

Fig. 410 is a diagram illustrating an active logical allocation unit table and active allocation units;

35 Fig. 411 is a diagram illustrating a conceptual logical allocation unit directory structure;

Fig. 412 is a diagram illustrating detail of a logical allocation unit directory entry;

Fig. 413 is a diagram illustrating universal identifiers and active object numbers;

Fig. 416 is a diagram illustrating subject templates, primitive access control list entries, and extended access control list entries;

40 Fig. 421 is a diagram illustrating an active primitive access matrix and an active primitive access matrix entry;

Fig. 422 is a diagram illustrating primitive data access checking;

Fig. 448 is a diagram illustrating event counters and await entries;

Fig. 449 is a diagram illustrating an await table overview;

45 Fig. 453 is a diagram illustrating an overview of a virtual processor;

Fig. 454 is a diagram illustrating virtual processor synchronization;

Fig. 467 is a diagram illustrating an overview of a macrostack object;

Fig. 468 is a diagram illustrating details of a macrostack object base;

Fig. 469 is a diagram illustrating details of a macrostack frame;

50 Fig. 470 is a diagram illustrating an overview of a secure stack;

Fig. 471 is a diagram illustrating details of a secure stack frame; and,

Fig. 472 is a diagram illustrating an overview of procedure object.

DESCRIPTION OF THE PREFERRED EMBODIMENTS

55 The following description presents the structure and operation of a computer system incorporating a presently preferred embodiment of the present invention. As indicated in the following Table of Contents, certain features of computer system structure and operation will first be described in an Introductory Overview. Next, these and other features will be described in further detail in a more detailed Introduction to the detailed descriptions of the computer system. Following the Introduction, the structure and

60 operation of the computer system will be described in detail. The detailed descriptions will present descriptions of the structure and operation of each of the major subsystems, or elements, of the computer system, of the interfaces between these major subsystems, and of overall computer system operation. Next, certain features of the operation of the individual subsystems will be presented in further detail.

65 Certain conventions are used throughout the following descriptions to enhance clarity of presentation. First, and with exception of the Introductory Overview, each figure referred to in the following descriptions

will be referred to by a three digit number. The most significant digit represents the number of the chapter in the following descriptions in which a particular figure is first referred to. The two least significant digits represent the sequential number of appearance of a figure in a particular chapter. For example, Figure 319 would be the nineteenth figure appearing in the third chapter. Figures appearing in the Introductory Overview are referred to by a one or two digit number representing the order in which they are referred to in the Introductory Overview. It should be noted that certain figure numbers, for example, Figure 208, do not appear in the following figures and descriptions; the subject matter of these figures has been incorporated into other figures and these figures deleted, during drafting of the following descriptions, to enhance clarity of presentation.

Second, reference numerals comprise a two digit number (00—99) preceded by the number of the figure in which the corresponding elements first appear. For example, reference numerals 31901 to 31999 would refer to elements 1 through 99 appearing in Fig. 319.

Finally, interconnections between related circuitry is represented in two ways. First, to enhance clarity of presentation, interconnections between circuitry may be represented by common signal names or references, rather than by drawn representations of wires or buses. Second, where related circuitry is shown in two or more figures, the figures may share a common figure number and will be distinguished by a letter designation, for example, Figs. 319, 319A, and 319B. Common electrical points between such circuitry may be indicated by a bracket enclosing a lead to such a point and a designation of the form "A—b". "A" indicates other figures having the same common point for example, 319A, and "b" designates the particular common electrical point. In cases of related circuitry shown in this manner in two or more figures, reference numerals to elements will be assigned in sequence through the group of figures; the figure number portion of such reference numerals will be that of the first figure of the group of figures.

INTRODUCTORY OVERVIEW

- 25 A. Hardware Overview (Fig. 1)
- B. Individual Operating Features (Figs. 2, 3, 4, 5, 6)
 - 1. Addressing (Fig. 2)
 - 2. S-Language Instructions and Namespace Addressing (Fig. 3)
 - 3. Architectural Base Pointer Addressing
 - 30 4. Stack Mechanisms (Figs. 4-5)
- C. Procedure Processes and Virtual Processors (Fig. 6)
- D. CS 101 Overall Structure and Operation (Figs. 7, 8, 9, 10, 11, 12, 13, 14, 15)
 - 1. Introduction (Fig. 7)
 - 2. Compilers 702 (Fig. 7)
 - 35 3. Binder 703 (Fig. 7)
 - 4. EOS 704 (Fig. 7)
 - 5. KOS and Architectural Interface 708 (Fig. 7)
 - 6. Processes 610 and Virtual Processors 612 (Fig. 8)
 - 7. Processes 610 and Stacks (Fig. 9)
 - 40 8. Processes 610 and Calls (Figs. 10, 11)
 - 9. Memory References and the Virtual Memory Management System (Fig. 12, 13)
 - 10. Access Control (Fig. 14)
 - 11. Virtual Processors and Virtual Processor Swapping (Fig. 15)
- E. CS 101 Structural Implementation (Figs. 16, 17, 18, 19, 20)
 - 45 1. (IOS) 116 (Figs. 16, 17)
 - 2. Memory (MEM) 112 (Fig. 18)
 - 3. Fetch Unit (FU) 120 (Fig. 19)
 - 4. Execute Unit (EU) 122 (Fig. 20)
- 50 1. Introduction (Figs. 101—110)
 - A. General Structure and Operation (Fig. 101)
 - a. General Structure
 - b. General Operation
 - c. Definition of Certain Terms
 - 55 d. Multi-program Operation
 - e. Multi-Language Operation
 - f. Addressing Structure
 - g. Protection Mechanism
 - B. Computer System 10110 Information Structure and Mechanisms (Figs. 102, 103, 104, 105)
 - 60 a. Introduction (Fig. 102)
 - b. Process Structures 10210 (Figs. 103, 104, 105)
 - 1. Procedure Objects (Fig. 103)
 - 2. Stack Mechanisms (Figs. 104, 105)
 - 3. FURSM 10214 (Fig. 103)
 - 65 C. Virtual Processor State Blocks and Virtual Process Creation (Fig. 102)

EP 0 067 556 B1

- D. Addressing Structures 10220 (Figs. 103, 106, 107, 108)
 - 1. Objects, UID's, AON's, Names, and Physical Addresses (Fig. 106)
 - 2. Addressing Mechanisms 10220 (Fig. 107)
 - 3. Name Resolution (Figs. 103, 108)
 - 5 4. Evaluation of AON Addresses to Physical Addresses (Fig. 107)
 - E. CS 10110 Protection Mechanisms (Fig. 109)
 - F. CS 10110 Micro-Instruction Mechanisms (Fig. 110)
 - G. Summary of Certain CS 10110 Features and Alternate Embodiments.
10. 2. Detailed Description of CS 10110 Major Subsystems Figs. 201—206, 207—274
- A. MEM 10110 (Figs. 201, 206, 207-237)
 - a. Terminology
 - b. MEM 10112 physical Structure (Fig. 201)
 - c. MEM 10112 General Operation
 - 15 d. MEM 10112 Port Structure
 - 1. IO Port Characteristics
 - 2. JO Port Characteristics
 - 3. JI Port Characteristics
 - e. MEM 10112 Control Structure and Operation (Fig. 207)
 - 20 1. MEM 10112 Control Structure
 - 2. MEM 10112 Control Operation
 - f. MEM 10112 Operations
 - g. MEM 10112 Interfaces to JP 10114 and IOS 10116 (Figs. 209, 210, 211, 204)
 - 25 1. IO Port 20910 Operating Characteristics (Figs 209, 204)
 - 2. JO Port 21010 Operating Characteristics (Fig. 210)
 - 3. JI Port 21110 Operating Characteristics (Fig. 211)
 - h. FIU 20120 (Figs. 201, 230, 231)
 - B. Fetch Unit 10120 (Figs. 202, 206, 101, 103, 104, 238)
 - 1. Descriptor Processor 20210 (Figs. 202, 101, 103, 104, 238, 239).
 - 30 a. Offset Processor 20218 Structure
 - b. AON Processor 20216 Structure
 - c. Length Processor 20220 Structure
 - d. Descriptor Processor 20218 Operation
 - a.a. Offset Selector 20238
 - 35 b.b. Offset Multiplexer 20240 Detailed Structure (Fig. 238)
 - c.c. Offset Multiplexer 20240 Detailed Operation
 - aaa. Internal Operation
 - bbb. Operation Relative to DESP 20210
 - e. Length Processor 20220 (Fig. 239)
 - 40 a.a. Length ALU 20252
 - b.b. BIAS 20246 (Fig. 239)
 - f. AON Processor 20216
 - a.a. AONGRF 20232
 - b.b. AON Selector 20248
 - 45 2. Memory Interface 20212 (Figs. 106, 240)
 - a.a. Descriptor Trap 20256 and Data Trap 20258
 - b.b. Name Cache 10226, Address Translation Unit 10228, and Protection Cache 10234 (Fig. 106)
 - c.c. Structure and Operation of Generalized Cache and NC 10226 (Fig. 240)
 - d.d. ATU 10228 and PC 10234
 - 50 3. Fetch Unit Control Logic 20214 (Fig. 202)
 - a.a. Fetch Unit Control Logic 20214 Overall Structure
 - b.b. Fetch Unit Control Logic 20214 Operation
 - a.a.a. Prefetcher 20264, Instruction Buffer 20262, Parser 20264, Operation Code Register 20268, CPC 20270, IPC 20272, and EPC 20274 (Fig. 241)
 - 55 b.b.b. Fetch Unit Dispatch Table 11010, Execute Unit Dispatch Table 20266, and Operation Code Register 20268 (Fig. 242)
 - c.c.c. Next Address Generator 24310 (Fig. 243)
 - cc. FUCTL 20214 Circuitry for CS 10110 Internal Mechanisms (Figs. 244-250)
 - a.a.a. State Logic 20294 (Figs. 244A—244Z)
 - 60 b.b.b. Event Logic 20284 (Figs. 245, 246, 247, 248)
 - c.c.c. Fetch Unit S-Interpreter Table 11012 (Fig. 249)
 - d.d. CS 10110 Internal Mechanism Control
 - a.a.a. Return Control Word Stack 10358 (Fig. 251)
 - b.b.b. Machine Control Block (Fig. 252)
 - 65 c.c.c. Register Address Generator 20288 (Fig. 253)

EP 0 067 556 B1

- d.d.d. Timers 20296 (Fig. 254)
- e.e.e. Fetch Unit 10120 Interface to Execute Unit 10122
- C. Execute Unit 10122 (Figs. 203, 255-268)
 - a. General Structure of Execute Unit 10122
 - 1. Execute Unit I/O 20312
 - 2. Execute Unit Control Logic 20310
 - 3. Multiplier Logic 20314
 - 4. Exponent Logic 20316
 - 5. Multiplier Control 20318
 - 6. Test and Interface Logic 20320
 - b. Execute Unit 10122 Operation (Fig. 255)
 - 1. Execute Unit Control Logic 20310 (Fig. 255)
 - a.a. Command Queue 20342
 - b.b. Command Queue Event Control Store 25514 and Command Queue Event Address Control Store 25516
 - c.c. Execute Unit S-Interpreter Table 20344
 - d.d. Microcode Control Decode Register 20346
 - e.e. Next Address Generator 20340
 - 2. Operand Buffer 20322 (Fig. 256)
 - 3. Multiplier 20314 (Figs. 257, 258)
 - a.a. Multiplier 20314 I/O Data Paths and Memory (Fig. 257)
 - a.a.a. Container Size Check
 - b.b.b. Final Result Output Multiplexer 20324
 - 4. Test and Interface Logic 20320 (Figs. 260-268)
 - a.a. FU 10120/EU 10122 Interface
 - a.a.a. Loading of Command Queue 20342 (Fig. 260)
 - b.b.b. Loading of Operand Buffer 20320 (Fig. 261)
 - c.c.c. Storeback (Fig. 262)
 - d.d.d. Test Conditions (Fig. 263)
 - e.e.e. Exception Checking (Fig. 264)
 - f.f.f. Idle Routine
 - g.g.g. EU 10122 Stack Mechanisms (Figs. 265, 266, 267)
 - h.h.h. Loading of Execute Unit S-Interpreter Table 20344 (Fig. 268)
- D. I/O System 10116 (Figs. 204, 206, 269)
 - a. I/O System 10116 Structure (Fig. 204)
 - b. I/O System 10116 Operation (Fig. 269)
 - 1. Data Channel Devices
 - 2. I/O Control Processor 20412
 - 3. Data Mover 20410 (Fig. 269)
 - a.a. Input Data Buffer 20440 and Output Data Buffer 20442
 - b.b. Priority Resolution and Control 20444 (Fig. 269)
- E. Diagnostic Processor 10118 (Fig. 101, 205)
- F. CS 10110 Micromachine Structure and Operation (Figs. 270—274)
 - a. Introduction
 - b. Overview of Devices Comprising FU Micromachine (Fig. 270)
 - 1. Devices Used By Most Microcode
 - a.a. MOD Bus 10144, JPD Bus 10142, and DB Bus 27021
 - b.b. Microcode Addressing
 - c.c. Descriptor Processor 20218 (Fig. 271)
 - d.d. EU 10122 Interface
 - 2. Specialized Micromachine Devices
 - a.a. Instruction Stream Reader 27001
 - b.b. SOP Decoder 27003
 - c.c. Name Translation Unit 27015
 - d.d. Memory Reference Unit 27017
 - e.e. Protection Unit 27019
 - f.f. KOS Micromachine Devices
 - c. Micromachine Stacks and Microroutine Calls and Returns (Figs. 272, 273)
 - 1. Micromachine Stacks (Fig. 272)
 - 2. Micromachine Invocations and Returns
 - 3. Means of Invoking Microroutines
 - 4. Occurrence of Event Invocations (Fig. 273)
 - d. Virtual Micromachines and Monitor Micromachine
 - 1. Virtual Mode
 - 2. Monitor Micromachine

- e. Interrupt and Fault Handling
 - 1. General Principles
 - 2. Hardware Interrupt and Fault Handling in CS 10110
 - 3. Monitor Mode: Differential Masking and Hardware Interrupt Handling
 - 5 g. FU Micromachine and CS 10110 Subsystems
 - 3. Namespace, S-Interpreters and Pointers (Figs. 301—307, 274)
 - A. Pointers and Pointer Resolution (Figs. 301, 302)
 - 10 a. Pointer Formats (Fig. 301)
 - b. Pointers in FU 10120 (Fig. 302)
 - c. Descriptor to Pointer Conversion
 - B. Namespace and the S-Interpreters (Figs. 303—307)
 - 15 a. Procedure Object 606 Overview (Fig. 303)
 - b. Namespace
 - 1. Name Resolution and Evaluation
 - 2. The Name Table (Fig. 304)
 - 3. Architectural Base Pointers (Figs. 305, 306)
 - 20 a.a. Resolving and Evaluating Names (Fig. 305)
 - b.b. Implementation of Name Evaluation and Name Resolve in CS 10110
 - c.c. Name Cache 10226 Entries (Fig. 306)
 - d.d. Name Cache 10226 Hits
 - e.e. Name Cache 10226 Misses
 - f.f. Flushing Name Cache 10226
 - g.g. Fetching the Instruction Stream
 - h.h. Parsing the Instruction Stream
 - 25 c. The S-Interpreters (Fig. 307)
 - 1. Translating SIP into a Dialect Number (Fig. 307)
 - 2. Dispatching
- 30 4. The Kernel Operation System
- A. Introduction
 - a. Operating Systems (Fig. 401)
 - 1. Resources Controlled by Operating Systems (Fig. 402)
 - b. The Operating System in CS 10110
 - 35 c. Extended Operating System and the Kernel Operating System (Fig. 403)
- B. Objects and Object Management (Fig. 404)
 - a. Objects and User Programs (Fig. 405)
 - b. UIDs 40401 (Fig. 406)
 - c. Object Attributes
 - 40 d. Attributes and Access Control
 - e. Implementation of Objects
 - 1. Introduction (Figs 407, 408)
 - 2. Objects in Secondary Storage 10124 (Figs. 409, 410).
 - 45 a.a. Representation of an Object's Contents on Secondary Storage 10124
 - b.b. LAUD 40903 (Figs. 411, 412)
 - 3. Active Objects (Fig. 413)
 - a.a. UID 40401 to AON 41304 Translation
- C. The Access Control System
 - 50 a. Subjects
 - b. Domains
 - c. Access Control Lists
 - 1. Subject Templates (Fig. 416)
 - 2. Primitive Access Control Lists (PACLs)
 - 3. APAM 10918 and Protection Cache 10234 (Fig. 421)
 - 55 4. Protection Cache 10234 and Protection Checking (Fig. 422)
- D. Processes
 - 1. Synchronization of Processes 610 and Virtual Processors 612
 - a. Event Counters 44801, Await Entries 44804, and Await Tables (Fig. 448, 449)
 - b. Synchronization with Event Counters 44801 and Await Entries 44804
- 60 E. Virtual processors 612 (Fig. 453)
 - a. Virtual Processor Management (Fig. 453)
 - b. Virtual Processors 612 and Synchronization (Fig. 454)
- F. Process 610 Stack Manipulation
 - 1. Introduction to Call and Return
 - 65 2. Macrostacks (MAS) 502 (Fig. 467)

- a.a. MAS Base 10410 (Fig. 468)
- b.b. Per-domain Data Area 46853 (Fig. 468)
- c.c. MAS Frame 46709 Detail (Fig. 469)
- 3. SS 504 (Fig. 470)
 - a.a. SS Base 47001 (Fig. 471)
 - b.b. SS Frames 47003 (Fig. 471)
 - a.a.a. Ordinary SS Frame Headers 10514 (Fig. 471)
 - b.b.b. Detailed Structure of Macrostate 10516 (Fig. 471)
 - c.c.c. Cross-domain SS Frames 47039 (Fig. 471)
- 4. Portions of Procedure Object 608 Relevant to Call and Return (Fig. 472)
- 5. Execution of Mediated Calls
 - a.a. Mediated Call SInS
 - b.b. Simple Mediated Calls (Figs. 270, 468, 469, 470, 471, 472)
 - c.c. Invocations of Procedures 602 Requiring SEBs 46864 (Figs. 270, 468, 469, 470, 471, 472)
 - d.d. Cross-Procedure Object Calls (Figs. 270, 468, 469, 470, 471, 472)
 - e.e. Cross-Domain Calls (Figs. 270, 408, 418, 468, 469, 470, 471, 472)
 - f.f. Failed Cross-Domain Calls (Figs. 270, 468, 469, 470, 471, 472)
- 6. Neighborhood Calls (Figs. 468, 469, 472)

INTRODUCTORY OVERVIEW

The following overview will first briefly describe the overall physical structure and operation of a presently preferred embodiment of a digital computer system incorporating the present invention. Then certain operating features of that computer system will be individually described. Next, overall operation of the computer system will be described in terms of those individual features.

A. Hardware Overview (Fig. 1)

Referring to Fig. 1, a block diagram of Computer System (CS) 101 incorporating the present invention is shown. Major elements of CS 101 are I/O System (IOS) 116, Memory (MEM) 112, and Job Processor (JP) 114. JP 114 is comprised of a Fetch Unit (FU) 120 and an Execute Unit (EU) 122. CS 101 may also include a Diagnostic Processor (DP), not shown or described in the instant description.

Referring first to IOS 116, a primary function of IOS 116 is control of transfer of information between MEM 112 and the outside world. Information is transferred from MEM 112 to IOS 116 through IOM Bus 130, and from IOS 116 to MEM 112 through MIO Bus 129. IOMC Bus 131 is comprised of bi-directional control signals coordinating operation of MEM 112 and IOS 116. IOS 116 also has an interface to FU 120 through IOJP Bus 132. IOJP Bus 132 is a bi-directional control bus comprised essentially of two interrupt lines. These interrupt lines allow FU 120 to indicate to IOS 116 that a request for information by FU 120 has been placed in MEM 112, and allows IOS 116 to inform FU 120 that information requested by FU 120 has been transferred into a location in MEM 112. MEM 112 is CS 101's main memory and serves as the path for information transfer between the outside world and JP 114. MEM 112 provides instructions and data to FU 120 and EU 122 through Memory Output Data (MOD) Bus 140 and receives information from FU 120 and EU 122 through Job Processor Data (JPD) Bus 142. FU 120 submits read and write requests to MEM 112 through Physical Descriptor (PD) Bus 146.

JP 114 is CS 101's CPU and, as described above, is comprised of FU 120 and EU 122. A primary function of FU 120 is executing operations of user's programs. As part of this function, FU 120 controls transfer of instructions and data from MEM 112 and transfer of results of JP 114 operations back to MEM 112. FU 120 also performs operating system type functions, and is capable of operating as a complete, general purpose CPU. EU 122 is primarily an arithmetic and logic unit provided to relieve FU 120 of certain arithmetic operations. FU 120, however, is capable of performing EU 122 operations. In alternate embodiments of CS 101, EU 122 may be provided only as an option for users having particular arithmetic requirements. Coordination of FU 120 and EU 122 operations is accomplished through FU/EU (FUEU) Bus 148, which includes bi-directional control signals and mutual interrupt lines. As described further below, both FU 120 and EU 122 contain register file arrays referred to respectively as CRF and ERF, in addition to registers associated with, for example, ALUs.

A primary feature of CS 101 is that IOS 116, MEM 112, FU 120 and EU 122 each contain separate and independent microinstruction control, so that IOS 116, MEM 112, and EU 122 operate asynchronously under the general control of FU 120. EU 122, for example, may execute a complex arithmetic operation upon receipt of data and a single, initial command from FU 120.

Having briefly described the overall structure and operation of CS 101, certain features of CS 101 will be individually further described next below.

B. Individual Operating Features (Figs. 2, 3, 4, 5, 6)

1. Addressing (Fig. 2)

Referring to Fig. 2, a diagrammatic representation of portions of CS 101's addressing structure is shown. CS 101's addressing structure is based upon the concept of Objects. An Object may be regarded as a

EP 0 067 556 B1

container for holding a particular type of information. For example, one type of Object may contain data while another type of Object may contain instructions or procedures, such as a user program. Still another type of Object may contain microcode. In general, a particular Object may contain only one type or class of information. An Object may, for example, contain up to 232 bits of information, but the actual size of a particular Object is flexible. That is, the actual size of a particular Object will increase as information is written into that Object and will decrease as information is taken from that Object. In general, information in Objects is stored sequentially, that is without gaps.

Each Object which can ever exist in any CS 101 system is uniquely identified by a serial number referred to as a Unique Identifier (UID). A UID is a 128 bit value comprised of a serial number dependent upon, for example, the particular CS 101 system and user, and a time code indicating time of creation of that Object. UIDs are permanently assigned to Objects, no two Objects may have the same UID, and UIDs may not be reused. UIDs provide an addressing base common to all CS 101 systems which may ever exist, through which any Object ever created may be permanently and uniquely identified.

As described above, UIDs are 128 bit values and are thus larger than may be conveniently handled in present embodiments of CS 101. In each CS 101, therefore, those Objects which are active (currently being used) in that system are assigned 14 bit Active Object Numbers (AONs). Each Object active in that system will have a unique AON. Unlike UIDs, AONs are only temporarily assigned to particular Objects. AONs are valid only within a particular CS 101 and are not unique between systems. An Object need not physically reside in a system to be assigned an AON, but can be active in that system only if it has been assigned an AON.

A particular bit within a particular Object may be identified by means of a UID address or an AON address. In CS 101, AONs and AON addresses are valid only within JP 114 while UIDs and UID addresses are used in MEM 112 and elsewhere. UID and AON addresses are formed by appending a 32 bit Offset (O) field to that Object's UID or AON. O fields indicate offset, or location, of a particular bit relative to the start of a particular Object.

Segments of information (sequences of information bits) within particular Objects may be identified by means of descriptors. A UID descriptor is formed by appending a 32 bit Length (L) field of a UID address. An AON, or logical descriptor is formed by appending a 32 bit L field to an AON address. L fields identify length of a segment of information bits within an Object, starting from the information bit identified by the UID or AON address. In addition to length information, UID and logical descriptors also contain Type fields containing information regarding certain characteristics of the information in the information segment. Again, AON based descriptors are used within JP 114, while UID based descriptors are used in MEM 112.

Referring to Figs. 1 and 2 together, translation between UID addresses and descriptors and AON addresses and descriptors is performed at the interface between MEM 112 and JP 114. That is, addresses and descriptors within JP 114 are in AON form while addresses and descriptors in MEM 112, IOS 116, and the external world are in UID form. In other embodiments of CS 101 using AONs, transformation from UID to AON addressing may occur at other interfaces, for example at the IOS 116 to MEM 112 interface, or at the IOS 116 to external world interface. Other embodiments of CS 101 may use UIDs throughout, that is not use AONs even in JP 114.

Finally, information within MEM 112 is located through MEM 112 Physical Addresses identifying particular physical locations within MEM 112's memory space. Both IOS 116 and JP 114 address information within MEM 112 by providing physical addresses to MEM 112. In the case of physical addresses provided by JP 114, these addresses are referred to as Physical Descriptors (PDs). As described below, JP 114 contains circuitry to translate logical descriptors into physical descriptors.

2. S-Language Instructions and Namespace Addressing (Fig. 3)

CS 101 is both an S-Language machine and a Namespace machine. That is, operations to be executed by CS 101 are expressed as S-Language Operations (SOPs) while operands are identified by Names. SOPs are of a lower, more detailed, level than user language instructions, for example FORTRAN and COBOL, but of a higher level than conventional machine language instructions. SOPs are specific to particular user languages rather than a particular embodiment of CS 101, while conventional machine language instructions are specific to particular machines. SOPs are in turn interpreted and executed by microcode. There will be an S-Language Dialect, a set of SOPs, for each user language. CS 101, for example, may have SOP Dialects for COBOL, FORTRAN, and SPL. A particular distinction of CS 101 is that all SOPs are of a uniform, fixed length, for example 16 bits. CS 101 may generally contain one or more sets of microcode for each S-Language Dialect. These microcode Dialect Sets may be completely distinct, or may overlap where more than one SOP utilizes the same microcode.

As stated above, in CS 101 all operands are identified by Names, which are 8, 12, or 16 bit numbers. CS 101 includes one or more "Name Tables" containing an Entry for each operand Name appearing in programs currently being executed. Each Name Table Entry contains information describing the operand referred to by a particular Name, and the directions necessary for CS 101 to translate that information into a corresponding logical descriptor. As previously described, logical descriptors may then be transformed into physical descriptors to read and write operands from or to MEM 112. As described above, UIDs are unique for all CS 101 systems and AONs are unique within individual CS 101 systems. Names, however, are unique only within the context of a user's program. That is, a particular Name may appear in two different

EP 0 067 556 B1

user's programs and, within each program, will have different Name Table Entries and will refer to different operands.

CS 101 may thereby be considered as utilizing two sets of instructions. A first set is comprised of SOPs, that is instructions selecting algorithms to be executed. The second set of instructions are comprised of Names, which may be regarded as entry points into tables of instructions for making references regarding operands.

Referring to Fig. 3, a diagramic representation of CS 101 instruction stream is shown. A typical SIN is comprised of an SOP and may include one or more Names referring to operands. SOPs and Names allow user's programs to be expressed in very compact code. Fewer SOPs than machine language instructions are required to express a user's program. Also, use of SOPs allows easier and simpler construction of compilers, and facilitates adaption of CS 101 systems to new user languages. In addition, use of Names to refer to operands means that SOPs are independent of the form of the operands upon which they operate. This in turn allows for more compact code in expressing user programs in that SOPs specifying operations dependent upon operand form are not required.

3. Architectural Base Pointer Addressing

As will be described further below, a user's program residing in CS 101 will include one or more Objects. First, a Procedure Object contains at least the SINs of the user's programs and a Name Table containing entries for operand Names of the program. The SINs may include references, or calls, to other Procedure Objects containing, for example, procedures available in common to many users. Second, a Static Data Area may contain static data, that is data having an existence for at least a single execution of the program. And third, a Macro-stack, described below, may contain local data, that is data generated during execution of a program. Each Procedure Object, the Static Data Area and the Macro-stack are individual Objects identified by UIDs and AONs and addressable through UID and AON addresses and descriptors.

Locations of information within a user's Procedure Objects, Static Data Area, and Macro-stack are expressed as offsets from one of three values, or base addresses, referred to as Architectural Base Pointers (ABPs). For example, location information in Name Tables is expressed as offsets from one of the ABPs. ABPs may be expressed as previously described.

The three ABPs are the Frame Pointer (FP), the Procedure Base Pointer (PBP), and the Static Data Pointer (SDP). Locations of data local to a procedure, for example in the procedure's Macrostack, are described as offsets from FP. Locations of non-local data, that is Static Data, are described as offsets from SDP. Locations of SINs in Procedure Objects are expressed as offsets from PBP; these offsets are determined as a Program Counter (PC) value. Values of the ABPs vary during program execution and are therefore not provided by the compiler converting a user's high level language program into a program to be executed in a CS 101 system: When the program is executed, CS 101 provides the proper values for the ABPs. When a program is actually being executed, the ABP's values are stored in FU 120's GRF.

Other pointers are used, for example, to identify the top frame of CS 101's Secure Stack (a microcode level stack described below) or to identify the microcode Dialect currently being used in execute the SINs of a procedure. These pointers are similar to FP, SDP, and PBP.

4. Stack Mechanisms (Fig. 4—5)

Referring to Fig. 4 and 4A, diagramic representations of various control levels and stack mechanisms of, respectively, conventional machines and CS 101, are shown. Referring first to Fig. 4, top level of control is provided by User Language Instructions 402, for example in FORTRAN or COBOL. User Language Instructions 402 are converted into a greater number of more detailed Machine Language Instructions 404, used within a machine to execute user's programs. Within the machine, Machine Language Instructions 404 are interpreted and executed by Microcode Instructions 406, that is sequences of microinstructions which in turn directly control Machine Hardware 408. Some conventional machines may include a Stack Mechanism 410 used to save current machine state, that is current microinstruction and contents of various machine registers, if a current Machine Language Instruction 404 cannot be executed or is interrupted. In general, machine state on the microcode and hardware level is not saved. Execution of a current Machine Language Instruction 404 is later resumed at start of the microinstruction sequence for executing that Machine Language Instruction 404.

Referring to Fig. 4A, top level control in CS 101 is by User Language Instructions 412 as in a conventional machine. In CS 101, however, User Language Instructions 412 are translated into SOPs 414 which are of a higher level than conventional machine language instructions. In general, a single User Language Instruction 412 is transformed into at most two or three SOPs 414, as opposed to an entire sequence of conventional Machine Language Instructions 404. SOPs 414 are interpreted and executed by Microcode Instructions 416 (sequences of microinstructions) which directly control CS 101 Hardware 418. CS 101 includes a Macro-stack Mechanism (MAS) 420, at SOPs 414 level, which is comparable to but different in construction and operation from a conventional Machine Language Stack Mechanism 410. CS 101 also includes Micro-code Stack Mechanisms 422 operating at Microcode 416 level, so that execution of an interrupted microinstruction of a microinstruction sequence may be later resumed with the particular microinstruction which was active at the time of the interrupt. CS 101 is therefore more efficient in handling

interrupts in that execution of microinstruction sequences is resumed from the particular point that a microinstruction sequence was interrupted, rather from the beginning of that sequence. As will be described further below, CS 101's Micro-code Stack Mechanisms 422 on microcode level is effectively comprised of two stack mechanisms. The first stack is Micro-instruction Stack (MIS) 424 while the second stack is referred to as Monitor Stack (MOS) 426. CS 101 SIN Microcode 428 and MIS 424 are primarily concerned with execution of SOPs of user's programs. Monitor Microcode 430 and MOS 426 are concerned with operation of certain CS 101 internal functions.

Division of CS 101's microcode stacks into an MIS 424 and a MOS 426 illustrates a further feature of CS 101. In conventional machines, monitor functions may be performed by a separate CPU operating in conjunction with the machine's primary CPU. In CS 101, a single hardware CPU is used to perform both functions with actual execution of both functions performed by separate groups of microcode. Monitor microcode operations may be initiated either by certain SInS 414 or by control signals generated directly by CS 101's Hardware 418. Invocation of Monitor Microcode 430 by Hardware 418 generated signals insures that CS 101's monitor functions may always be invoked.

Referring to Fig. 5, a diagramic representation of CS 101's stack mechanisms for a single user's program, or procedure, is shown. Basically, and with exception of MOS 426, CS 101's stacks reside in MEM 112 with certain portions of those stacks accelerated into FU 120 and EU 122 to enhance speed of operation.

Certain areas of MEM 112 storage space are set aside to contain Macro-Stacks (MASs) 502, stack mechanisms operating on the SInS level, as described above. Other areas of MEM 112 are set aside to contain Secure Stack (SS) 504, operating on the microcode level, as described above and of which MIS 424 is a part.

As described further below, both FU 120 and EU 122 contain register file arrays, referred to respectively as GRF and ERF, in addition to registers associated with, for example, ALUs. Referring to FU 120, shown therein is FU 120's GRF 506. GRF 506 is horizontally divided into three areas. A first area, referred to as General Registers (GRs) 508 may in general be used in the same manner as registers in a conventional machine. A second area of GRF 506 is Micro-Stack (MIS) 424, and is set aside to contain a portion of a Process's SS 504. A third portion of GRF 506 is set aside to contain MOS 426. Also indicated in FU 120 is a block referred to as Microcode Control State (mCS) 510. mCS 510 represents registers and other FU 120 hardware containing current operating state of FU 120 on the microinstruction and hardware level. mCS 510 may include, for example, the current microinstruction controlling operation of FU 120.

Referring to EU 122, indicated therein is a first block referred to as Execute Unit State (EUS) 512 and a second block referred to as SOP Stack 514. EUS 512 is similar to mCS 510 in FU 120 and includes all registers and other EU 122 hardware containing information reflecting EU 122's current operating state. SOP Stack 518 is a portion of EU 122's ERF 516 which has been set aside as a stack mechanism to contain a portion of a process's SS 504 pertaining to EU 122 operations.

Considering first MASs 502, as stated above MASs 502 operate generally upon the SInS level. MASs 502 are used in general to store current state of a process's (defined below) execution of a user's program.

Referring next to MIS 424, in a present embodiment of CS 101 that portion of GRF 506 set aside to contain MIS 424 may have a capacity of eight stack frames. That is, up to 8 microinstruction level interrupts or calls pertaining to execution of a user's program may be stacked within MIS 424. Information stored in MIS 424 stack frames is generally information from GR 508 and MCS 510. MIS 424 stack frames are transferred between MIS 424 and SS 504 such that at least one frame, and no more than 8 frames, of SS 504 reside in GRF 506. This insures that at least the top-most frames of a process's SS 504 are present in FU 120, thereby enhancing speed of operation of FU 120 by providing rapid access to those top frames. SS 504, residing in MEM 112, may contain, for all practical purposes, an unlimited number of frames so that MIS 424 and SS 504 appear to a user to be effectively an infinitely deep stack.

MOS 426 resides entirely in FU 120 and, in a present embodiment of CS 101, may have a capacity of 8 stack frames. A feature of CS 101 operation is that CS 101 mechanisms for handling certain events or interrupts should not rely in its operation upon those portions of CS 101 whose operation has resulted in those faults or interrupts. Among events handled by CS 101 monitor microcode, for example, are MEM 112 page faults. An MEM 112 page fault occurs whenever FU 120 makes a reference to data in MEM 112 and that data is not in MEM 112. Due to this and similar operations, MOS 426 resides entirely in FU 120 and thus does not rely upon information in MEM 112.

As described above, GRs 508, MIS 424, and MOS 426 each reside in certain assigned portions of GRF 506. This allows flexibility in modifying the capacity of GRs 508, MIS 424, and MOS 426 as indicated by experience, or to modify an individual CS 101 for particular purposes.

Referring finally to EU 122, EUS 512 is functionally a part of a process's SS 504. Also as previously described, EU 122 performs arithmetic operations in response to SInS and may be interrupted by FU 120 to aid certain FU 120 operations. EUS 512 allows stacking of interrupts. For example, FU 120 may first interrupt an arithmetic SOP to request EU 122 to aid in evaluation of a Name Table Entry. Before that first interrupt is completed, FU 120 may interrupt again, and so on.

SOP Stack 514, is a single frame stack for storing current state of EU 122 when an interrupt interrupts execution of an arithmetic SOP. An interrupted SOP's state is transferred into SOP Stack 514 and the interrupt begins execution in EUS 512. Upon occurrence of a second interrupt (before the first interrupt is completed) EU's first interrupt state is transferred from EUS 512 to a stack frame in SS 504, and execution

of the second interrupt begins in EUS 512. If a third interrupt occurs before completion of second interrupt, EU's second interrupt state is transferred from EUS 512 to another stack frame in SS 504 and execution of the third interrupt is begun in EUS 512; and so on. EUS 512 and SS 504 thus provide an apparently infinitely deep microstack for EU 122. Assuming that the third interrupt is completed, state of second interrupt is transferred from SS 504 to EUS 512 and execution of second interrupt resumed. Upon completion of second interrupt, state of first interrupt is transferred from SS 504 to EUS 512 and completed. After completion of first interrupt, state of the original SOP is transferred from SOP Stack 514 to EUS 512 and execution of that SOP resumed.

10 C. Procedure Processes, and Virtual Processors (Fig. 6)

Referring to Fig. 6, a diagrammatic representation of procedures, processes, and virtual processes is shown. As described above, a user's program to be executed is compiled to result in a Procedure 602. A Procedure 602 includes a User's Procedure Object 604 containing the SOPs of the user's program and a Name Table containing Entries for operand Names of the user's program, and a Static Data Area 606. A Procedure 602 may also include other Procedure Objects 608, for example utility programs available in common to many users. In effect, a Procedure 602 contains the instructions (procedures) and data of a user's program.

A Process 610 includes, as described above, a Macro Stack (MAS) 502 storing state of execution of a user's Procedure 602 at the SOP level, and a Secure Stack (SS) 504 storing state of execution of a user's Procedure 602 at the microcode level. A Process 610 is associated with a user's Procedure 602 through the ABPs described above and which are stored in the MAS 502 of the Process 610. Similarly, the MAS 502 and SS 504 of a Process 610 are associated through non-architectural pointers, described above. A Process 602 is effectively a body of information linking the resources, hardware, microcode, and software, of CS 101 to a user's Procedure 602. In effect, a Process 610 makes the resources of CS 101 available to a user's Procedure 602 for executing of that Procedure 602. CS 101 is a multi-program machine capable of accommodating up to, for example, 128 processes 610 concurrently. The number of Processes 610 which may be executed concurrently is determined by the number of Virtual Processors 612 of CS 101. There may be, for example, up to 16 Virtual Processors 612.

As indicated in Fig. 6, a Virtual Processor 612 is comprised of a Virtual Processor State Block (VPSB) 614 associated with the SS 504 of a Process 612. A VPSB 614 is, in effect, a body of information accessible to CS 101's operating system and through which CS 101's operating system is informed of, and provided with access to, a Process 610 through that process 610's SS 504. A VPSB 614 is associated with a particular Process 610 by writing information regarding that Process 610 into that VPSB 614. CS 101's operating system may, by gaining access to a Process 610 through an associated PSB 614, read information, such as ABPs, from that Process 610 to FU 120, thereby swapping that Process 610 onto FU 120 for execution. It is said that a Virtual Processor 612 thereby executes a process 610; a Virtual Processor 612 may be regarded therefor, as a processor having "Virtual", or potential, existence which becomes "real" when its associated Process 610 is swapped into FU 120. In CS 101, as indicated in Fig. 6, only one Virtual Processor 612 may execute on FU 120 at a time and the operating system selects which Virtual Processor 612 will execute on FU 120 at any given time. In addition, CS 101's operating system selects which Processes 610 will be associated with the available Virtual Processors 612.

Having briefly described certain individual structural and operating features of CS 101, the overall operation of CS 101 will be described in further detail next below in terms of these individual features.

45 D. CS 101 Overall Structure and Operation (Figs. 7, 8, 9, 10, 11, 12, 13, 14, 15)

1. Introduction (Fig. 7)

As indicated in Fig. 7, CS 101 is a multiple level system wherein operations in one level are generally transparent to higher levels. User 701 does not see the S-Language, addressing, and protection mechanisms defined at Architectural Level 708. Instead, he sees User Interface 709, which is defined by Compilers 702, Binder 703, and Extended (high level) Operating System (EOS) 704. Compilers 702 translate high-level language code into S-Names and Binder 703 translates symbolic Names in programs into UID-offset addresses.

As Fig. 7 shows, Architectural Level 708 is not defined by FU 120 Interface 711. Instead, the architectural resources level are created by S-Language Interpreted S-Names when a program is executed; Name Interpreter 715 operates under control of S-Language Interpreters 705 and translates Names into logical descriptors. In CS 101, both S-Language Interpreters 705 and Name Interpreter 715 are implemented as microcode which executes on FU 120. S-Language Interpreters 705 may also use EU 122 to perform calculations. A Kernel Operating System (KOS) provides CS 101 with UID-offset addressing, objects, access checking, processes, and virtual processors, described further below. KOS has three kinds of components: KOS Microcode 710, KOS Software 706, and KOS Tables in MEM 112. KOS 710 components are microcode routines which assist FU 120 in performing certain required operations. Like other high-level language routines, KOS 706 components contain S-Names which are interpreted by S-Interpreter Microcode 705. Many KOS High-Level Language Routines 706 are executed by special KOS processes; others may be executed by any process. Both KOS High-Level Language Routines 706 and KOS Microcode 710 manipulate KOS Tables in MEM 112.

EP 0 067 556 B1

FU 120 Interface 711 is visible only to KOS and to S-Interpreter Microcode 705. For the purposes of this discussion, FU 120 may be seen as a processor which contains the following main elements:

A Control Mechanism 725 which executes microcode stored in Writable Control Store 713 and manipulates FU 120 devices as directed by this microcode.

A GRF 506 containing registers in which data may be stored.

A Processing Unit 715.

All microcode which executes on FU 120 uses these devices; there is in addition a group of devices for performing special functions; these devices are used only by microcode connected with those functions. The microcode, the specialized devices, and sometimes tables in MEM 112 make up logical machines for performing certain functions. These machines will be described in detail below.

In the following, each of the levels illustrated in Fig. 7 will be discussed in turn. First, the components at User Interface 709 will be examined to see how they translate user programs and requests into forms usable by CS 101. Then the components below the User Interface 709 will be examined to see how they create logical machines for performing CS 101 operations.

2. Compilers 702 (Fig. 7)

Compilers 702 translate files containing the highlevel language code written by User 701 into Procedure Objects 608. Two components of a Procedure Object 608 are code (SOPs) and Names, previously described. SOPs represent operations, and the Names represent data. A single SIN thus specifies an operation to be performed on the data represented by the Names.

3. Binder 703 (Fig. 7)

In some cases, Compiler 702 cannot define locations as offsets from an ABP. For example, if a procedure calls a procedure contained in another procedure object, the location to which the call transfers control cannot be defined as an offset from the PBP used by the calling procedure. In these cases, the compiler uses symbolic Names to define the locations. Binder 703 is a utility which translates symbolic Names into UID-offset addresses. It does so in two ways: by combining separate Procedure Objects 608 into a single large Procedure Object 608, and then redefining symbolic Names as offsets from that Procedure Object 608's ABPs, or by translating symbolic Names when the program is executed. In the second case, Binder 703 requires assistance from EOS 704.

4. EOS 704 (Fig. 7)

EOS 704 manages the resources that User 701 requires to execute his programs. From User 701's point of view, the most important of these resources are files and processes. EOS 704 creates files by requesting KOS to create an object and then mapping the file onto the object. When a User 701 performs an operation on a file, EOS 704 translates the file operation into an operation on an object. KOS creates them at EOS 704's request and makes them available to EOS 704, which in turn makes them available to User 701. EOS 704 causes a process to execute by associating it a Virtual Processor 612. In logical terms, a Virtual Processor 612 is the means which KOS provides EOS 704 for executing processes 610. As many Processes 610 may apparently execute simultaneously in CS 101 as there are Virtual Processors 612. The illusion of simultaneous execution is created by multiplexing JP 114 among the Virtual processors; the manner in which Processes 610 and Virtual Processors 610 are implemented will be explained in detail below.

5. KOS and Architectural Interface 708 (Fig. 7)

S-Interpreter Microcode 710 and Name Interpreter Microcode 715 require an environment provided by KOS Microcode 710 and KOS Software 706 to execute SINs. For example, as previously explained, Names and program locations are defined in terms of ABPs whose values vary during execution of the program. The KOS environment provides values for the ABPs, and therefore makes it possible to interpret Names and program locations as locations in MEM 112. Similarly, KOS help is required to transform logical descriptors into references to MEM 112 and to perform protection checks.

The environment provided by KOS has the following elements:

A Process 610 which contains the state of an execution of the program for a given User 701.

A Virtual Processor 612 which gives the Process 610 access to JP 114.

An Object Management System which translates UIDs into values that are usable inside JP 114.

A Protection System which checks whether a Process 610 has the right to perform an operation on an Object.

A Virtual Memory Management System which moves those portions of Objects which a Process 610 actually references from the outside world into MEM 112 and translates logical descriptors into physical descriptors.

In the following, the logical properties of this environment and the manner in which a program is executed in it will be explained.

6. Processes 610 and Virtual Processors 612 (Fig. 8)

Processes 610 and Virtual Processors 612 have already been described in logical terms; Fig. 8 gives a high-level view of their physical implementation.

Fig. 8 illustrates the relationship between Processes 610, Virtual Processors 612, and JP 114. In physical terms, a Process 610 is an area of MEM 112 which contains the current state of a user's execution of a program. One example of such state is the current values of the ABPs and a program Counter (PC). Given the current value of the PBP and the PC, the next SOP in the program can be executed; similarly, given the current values of SDP and FP, the program's Names can be correctly resolved. Since the Process 610 contains the current state of a program's execution, the program's physical execution can be stopped and resumed at any point. It is thus possible to control program execution by means of the Process 610.

As already mentioned, a process 610's execution proceeds only when KOS has bound it to a Virtual Processor 612, that is, an area of MEM 112 containing the state required to execute microinstructions on JP 114 hardware. The operation of binding is simply a transfer of Process 610 state from the Process 610's area of MEM 112 to a Virtual Processor 612's area of MEM 112. Since binding and unbinding may take place at any time, EOS 704 may multiplex Processes 610 among Virtual Processors 612. In Fig. 8, there are more Processes 610 than there are Virtual Processors 612. The physical execution of a Process 610 on JP 114 takes place only while the Process 610's Virtual Processor 612 is bound to JP 114, i.e., when state is transferred from Virtual Processor 612's area of MEM 112 to JP 114's registers. Just as EOS 704 multiplexes Virtual Processors 612 among Processes 610, KOS multiplexes JP 114 among Virtual Processors 612. In Fig. 8, only one Process 610 is being physically executed. The means by which JP 114 is multiplexed among Virtual Processors 612 will be described in further detail below.

7. Processes 610 and Stacks (Fig. 9)

In CS 101 systems, a Process 610 is made up of six Objects: one Process Object 901 and Five Stack Objects 902 to 906. Fig. 9 illustrates a Process 610. Process Object 901 contains the information which EOS 704 requires to manage the Process 610. EOS 704 has no direct access to Process Object 901, but instead obtains the information it needs by means of functions provided to it by KOS 706, 710. Included in the information are the UIDs of Stack Objects 902 through 906. Stack Objects 902 to 906 contain the Process 610's state.

Stack Objects 902 through 905, are required by CS 101's domain protection method and comprise Process 610's MAS 502. Briefly, a domain is determined in part by operations performed when a system is operating in that domain. For example, the system is in EOS 704 domain when executing EOS 704 operations and in KOS 706, 710 domain when executing KOS 706, 710 operations. A Process 610 must have one stack for each domain it enters. In the present embodiment, the number of domains is fixed at four, but alternate embodiments may allow any number of domains, and correspondingly, any number of Stack Objects. Stack Object 906 comprises Process 610's Secure Stack 504 and is required to store state which may be manipulated only by KOS 706, 710.

Each invocation made by a Process 610 results in the addition of frames to Secure Stack 504 and to Macro-Stack 502. The state stored in the Secure Stack 504 frame includes the macrostate for the invocation, the state required to bind Process 610 to a Virtual Processor 612. The frame added to Macro-Stack 502 is placed in one of Stack Objects 902 through 905. Which Stack Objects 902 to 905 gets the frame is determined by the invoked procedure's domain of execution.

Fig. 9 shows the condition of a Process 610's MAS 502 and Secure Stack 504 after the Process 610 has executed four invocations. Secure Stack 504 has one frame for each invocation; the frames of Process 610's MAS 502 are found in Stack Objects 902, 904, and 905. As revealed by their locations, Frame 1 is for an invocation of a routine with KOS 706, 710 domain of execution, Frame 2 for an invocation of a routine with the EOS 704 domain of execution, and Frames 3 and 4 for invocations of routines with the User domain of execution. Process 610 has not yet invoked a routine with the Data Base Management System (DBMS) domain of execution. The frames in Stack Objects 902 through 905 are linked together, and a frame is added to or removed from Secure Stack 504 every time a frame is added to Stack Objects 902 through 905. MAS 502 and Secure Stack 504 thereby function as a single logical stack even though logically contained in five separate Objects.

8. Processes 610 and Calls (Figs. 10, 11)

In the CS 101, calls and returns are executed by KOS 706, 710. When KOS 706, 710 performs a call for a process, it does the following:

It saves the calling invocation's macrostate in the top frame of Secure Stack 504 (Fig. 9).

It locates the procedure whose Name is contained in the call. The location of the first SIN in the procedure becomes the new PBP.

Using information contained in the called procedure, KOS 706, 710 creates a new MAS 502 frame in the proper Stack Object 902 through 905 and a new Secure Stack 504 frame in Secure Stack 504. FP is updated to point to the new MAS 502. If necessary, SDP is also updated.

Once the values of the ABPs have been updated, the PC is defined, Names can be resolved, and execution of the invoked routine can commence. On a return from the invocation to the invoking routine, the stack frames are deleted and the ABPs are set to the values saved in the invoking routine's macrostate. The invoking routine then continues execution at the point following the invocation.

A Process 610 may be illustrated in detail by putting the FORTRAN statement A + B into a FORTRAN routine called EXAMPLE and invoking it from another FORTRAN routine named CALLER. To simplify the

EP 0 067 556 B1

example, it is assumed that CALLER and EXAMPLE both have the same domain of execution. The parts of EXAMPLE which are of interest look like this:

```
5      SUBROUTINE EXAMPLE (C)
      INTEGER X,C
      INTEGER A,B
10     ...
      A = B
      ...
15     RETURN
      END
```

20 The new elements are a formal argument, C, and a new local variable, X. A formal argument is a data item which receives its value from a data item used in the invoking routine. The formal argument's value thus varies from invocation to invocation. The portions of INVOKER which are of interest look like this:

```
      SUBROUTINE INVOKER
25     INTEGER Z
      ...
30     CALL EXAMPLE (Z)
      ...
      END
```

35 The CALL statement in INVOKER specifies the Name of the subroutine being invoked and the actual arguments for the subroutine's formal arguments. During the invocation, the subroutine's formal arguments take on the values of the actual arguments. Thus, during the invocation specified by this CALL statement, the formal argument C will have the value represented by the variable Z in INVOKER.

40 When INVOKER is compiled, the compiler produces a CALL SIN corresponding to the CALL statement. The CALL SIN contains a Name representing a pointer to the beginning of the called routine's location in a procedure object and a list of Names representing the call's actual arguments. When CALL is executed, the Names are interpreted to resolve the SIN's Names as previously described, and KOS 710 microcode to perform MAS 502 and Secure Stack 504 operations.

45 Fig. 10 illustrates the manner in which the KOS 710 call microcode manipulates MAS 502 and Secure Stack 504.

Fig. 10 includes the following elements:

Call Microcode 1001, contained in FU 120 Writable Control Store 1014.

PC Device 1002, which contains part of macrostate belonging to the invocation of INVOKER which is executing the CALL statement.

50 Registers in FU Registers 1014. Registers 1004 contents include the remainder of macrostate and the descriptors corresponding to Names for EXAMPLE's location and the actual argument Z.

Procedure Object 1006 contains the entries for INVOKER and EXAMPLE, their Name Tables, and their code.

55 Macro-Stack Object 1008 (MAS 502) and Secure Stack Object 1010 (Secure Stack 504) contain the stack frames for the invocations of INVOKER and EXAMPLE being discussed here. EXAMPLE's frame is in the same Macro-Stack object as INVOKER's frame because both routines are contained in the same Procedure Object 1006, and therefore have the same domain of execution.

60 KOS Call Microcode 1001 first saves the macrostate of INVOKER's invocation on Secure Stack 504. As will be discussed later, when the state is saved, KOS 706 Call Microcode 1001 uses other KOS 706 microcode to translate the location information contained in the macrostate into the kind of pointers used in MEM 112. Then Microcode 1001 uses the descriptor for the routine Name to locate the pointer to EXAMPLE's entry in Procedure Object 1006. From the entry, it locates pointers to EXAMPLE's Name Table and the beginning of EXAMPLE's code. Microcode 1001 takes these pointers, uses other KOS 706 microcode to translate them into descriptors, and places the descriptors in the locations in Registers 1004 reserved for the values of the PBP and NTP. It then updates the values contained in PC Device 1002 so that

when the call is finished, the next SIN to be executed will be the first SIN in EXAMPLE.

CALL Microcode 1001 next constructs the frames for EXAMPLE on Secure Stack 504 and Macro-Stack 502. This discussion concerns itself only with Frame 1102 on Macro-Stack 502. Fig. 11 illustrates EXAMPLE's Frame 1102. The size of Frame 1102 is determined by EXAMPLE's local variables (X, A, and B) and formal arguments (C). At the bottom of Frame 1102 is Header 1104. Header 1104 contains information used by KOS 706, 710 to manage the stack. Next comes Pointer 1106 to the location which contains the value represented by the argument C. In the invocation, the actual for C is the local variable Z in INVOKER. As is the case with all local variables, the storage represented by Z is contained in the stack frame belonging to INVOKER's invocation. When a name interpreter resolved C's name, it placed the descriptor in a register. Call Microcode 1001 takes this descriptor, converts it to a pointer, and stores the pointer above Header 1104.

Since the FP ABP points to the location following the last pointer to an actual argument, Call Microcode 1001 can now calculate that location, convert it into a descriptor, and place it in a FU Register 1004 reserved for FP. The next step is providing storage for EXAMPLE's local variables. EXAMPLE's procedure Object 1006 contains the size of the storage required for the local variables, so Call Microcode 1001 obtains this information from procedure Object 1006 and adds that much storage to Frame 1102. Using the new value of FP and the information contained in the Name Table Entries for the local data, Name Interpreter 715 can now construct descriptors for the local data. For example, A's entry in Name Table specified that it was offset 32 bits from FP, and was 32 bits long. Thus, its storage falls between the storage for X and B in Figure 11.

9. Memory References and the Virtual Memory Management System (Fig. 12, 13)

As already explained, a logical descriptor contains an AON field, an offset field, and a length field. Fig. 12 illustrates a Physical Descriptor. Physical Descriptor 1202 contains a Frame Number (FN) field, a Displacement (D) field, and a Length (L) field. Together, the Frame Number field and the Displacement field specify the location in MEM 112 containing the data, and the Length field specifies the length of the data.

As is clear from the above, the virtual memory management system must translate the AON-offset location contained in a logical descriptor 1204 into a Frame Number-Displacement location. It does so by associating logical pages with MEM 112 frames. (N.B: MEM 112 frames are not to be confused with stack frames). Fig. 13, illustrates how Macrostack 502 Object 1302 is divided into Logical Pages 1304 in secondary memory and how Logical Pages 1304 are moved onto Frames 1306 in MEM 112. A Frame 1306 is a fixed-size, contiguous area of MEM 112. When the virtual memory management system brings data into MEM 112, it does so in frame-sized chunks called Logical Pages 1308. Thus, from the virtual memory system's point of view, each object is divided into Logical Pages 1308 and the address of data on a page consists of the AON of the data's Object, the number of pages in the object, and its displacement on the page. In Fig. 13, the location of the local variable B of EXAMPLE is shown as it is defined by the virtual memory system. B's location is a UID and an offset, or, inside JP 114, an AON and an offset. As defined by the virtual memory system, B's location is the AON, the page number 1308, and a displacement within the page. When a process references the variable B, the virtual memory management system moves all of Logical Page 1308 into a MEM 112 Frame 1306. B's displacement remains the same, and the virtual memory system translates its Logical Page Number 1308 into the number of Frame 1306 in MEM 112 which contains the page.

The virtual memory management system must therefore perform two kinds of translations: (1) AON-offset addresses into AON-page number-displacement addresses, and (2) AON-page number into a frame number.

10. Access Control (Fig. 14)

Each time a reference is made to an Object, KOS 706, 710 checks whether the reference is legal. The following discussion will first present the logical structure of access control in CS 101, and then discuss the microcode and devices which implement it.

CS 101 defines access in terms of subjects, modes of access, and Object size. A process may reference a data item located in an Object if three conditions hold:

- 1) If the process's subject has access to the Object.
- 2) If the modes of access specified for the subject include those required to perform the intended operation.
- 3) If the data item is completely contained in the Object, i.e., if the data item's length added to the data item's offset do not exceed the number of bits in the Object.

The subjects which have access to an Object and the kinds of access they have to the Object are specified by a data structure associated with the Object called the Access Control List (ACL). An Object's size is one of its attributes. Neither an Object's size nor its ACL is contained in the Object. Both are contained in system tables, and are accessible by means of the Object's UID.

Fig. 14 shows the logical structure of access control in CS 101. Subject 1408 has four components: Principal 1404, Process 1405, Domain 1406, and Tag 1407. Tag 1407 is not implemented in a present embodiment of CS 101, so the following description will deal only with principal 1404, Process 1405, and Domain 1406.

EP 0 067 556 B1

Principal 1404 specifies a user for which the process which is making the reference was created; Process 1405 specifies the process which is making the reference; and, Domain 1406 specifies the domain of execution of the procedure which the process is executing when it makes the reference.

5 Each component of the Subject 1408 is represented by a UID. If the UID is a null UID, that component of the subject does not affect access checking. Non-null UIDs are the UIDs of Objects that contain information about the subject components. Principal Object 1404 contains identification and accounting information regarding system users, Process Object 1405 contains process management information, and Domain Object 1406 contains information about per-domain error handlers.

10 There may be three modes of accessing an Object 1410: read, write, and execute. Read and write are self-explanatory; execute is access which allows a subject to execute instructions contained in the Object.

Access Control Lists (ACLs) 1412 are made up of Entries 1414. Each entry two components: Subject Template 1416 and Mode Specifier 1418. Subject Template 1416 specifies a group of subjects that may reference the Object and Mode Specifier 1418 specifies the kinds of access these subjects may have to the Object. Logically speaking, ACL 1412 is checked each time a process references an Object 1410. The reference may succeed only if the process's current Subject 1408 is one of those on Object 1410's ACL 1412 and if the modes in the ACL Entry 1414 for the Subject 1408 allow the kind of access the process wishes to make.

20 11. Virtual Processors and Virtual Processor Swapping (Fig. 15)

As previously mentioned, the execution of a program by a Process 610 cannot take place unless EOS 704 has bound the Process 610 to a Virtual Processor 612. Physical execution of the Process 610 takes place only while the process's Virtual Processor 612 is bound to JP 114. The following discussion deals with the data bases belonging to a Virtual Processor 612 and the means by which a Virtual Processor 612 is bound to and removed from JP 114.

25 Fig. 15 illustrates the devices and tables which KOS 706, 710 uses to implement Virtual Processors 612. FU 120 WCS contains KOS Microcode 706 for binding Virtual Processors 612 to JP 114 and removing them from JP 114. Timers 1502 and Interrupt Line 1504 are hardware devices which produce signals that cause the invocation of KOS Microcode 706. Timers 1502 contains two timing devices: Interval Timer 1506, which may be set by KOS 706, 710 to signal when a certain time is reached, and Egg Timer 1508, which guarantees that there is a maximum time interval for which a Virtual processor 612 can be bound to JP 114 before it invokes KOS Microcode 706. Interrupt Line 1504 becomes active when JP 114 receives a message from IOS 116, for example when IOS 116 has finished loading a logical page into MEM 112.

30 FU 120 Registers 508 contain state belonging to the Virtual Processor 612 currently bound to JP 114. Here, this Virtual Processor 612 is called Virtual Processor A. In addition, Registers 508 contain registers reserved for the execution of VP Swapping Microcode 1510. ALU 1942 (part of FU 120) is used for the descriptor-to-pointer and pointer-to-descriptor transformations required when one Virtual Processor 612 is unbound from JP 114 and another bound to JP 114. MEM 112 contains data bases for Virtual Processors 612 and data bases used by KOS 706, 710 to manage Virtual Processors 612. KOS 706, 710 provides a fixed number of Virtual Processors 612 for CS 101. Each Virtual Processor 612 is represented by a Virtual Processor State Block (VPSB) 614. Each VPSB 614 contains information used by KOS 706, 710 to manage the Virtual Processor 612, and in addition contains information associating the Virtual Processor 612 with a process. Fig. 15 shows two VPSBs 614, one belonging to Virtual Processor 612A, and another belonging to Virtual Processor 612B, which will replace Virtual Processor 612A on JP 114. The VPSBs 614 are contained in VPSB Array 1512. The index of a VPSB 614 in VPSB Array 1512 is Virtual Processor Number 1514 belonging to the Virtual Processor 612 represented by a VPSB 614. Virtual Processor Lists 1516 are lists which KOS 706, 710 uses to manage Virtual Processors 612. If a Virtual Processor 612 is able to execute, its Virtual Processor Number 1514 is on a list called the Runnable List; Virtual Processors 612 which cannot run are on other lists, depending on the reason why they cannot run. It is assumed that Virtual Processor 612B's Virtual Processor Number 1514 is the first one on the Runnable List.

35 When a process is bound to a Virtual Processor 612, the Virtual Processor Number 1514 is copied into the process's Process Object 901 and the AONs of the process's Process Object 901 and stacks are copied into the Virtual Processor 612's VPSB 614. (AONs are used because a process's stacks are wired active as long as the process is bound to a Virtual Processor 612). Binding is carried out by KOS 706, 710 at the request of EOS 704. In Fig. 15, two Secure Stack Objects 906 are shown, one belonging to the process to which Virtual Processor 612A is bound, and one belonging to that to which Virtual Processor 612B is bound.

40 Having described certain overall operating features of CS 101, a present implementation of CS 101's structure will be described further next below.

60 E. CS 101 Structural Implementation (Figs. 16, 17, 18, 19, 20)

1. (IOS) 116 (Figs. 16, 17)

Referring to Fig. 16, a partial block diagram of IOS 116 is shown. Major elements of IOS 116 include an ECLIPSE® Burst Multiplexer Channel (BMC) 1614 and a NOVA® Data Channel (NDC) 1616, an IO Controller (IOC) 1618 and a Data Mover (DM) 1610. IOS 116's data channel devices, for example BMC 1614 and NDC 65 1616, comprise IOS 116's interface to the outside world. Information and addresses are received from

external devices, such as disk drives, communications modes, or other computer systems, by IOS 116's data channel devices and are transferred to DM 1610 (described below) to be written into MEM 112. Similarly, information read from MEM 112 is provided through DM 1610 to IOS 116's data channel devices and thus to the above described external devices. These external devices are a part of CS 101's addressable memory space and may be addressed through UID addresses.

IOC 1618 is a general purpose CPU, for example an ECLIPSE® computer available from Data General Corporation. A primary function of IOC 1618 is control of data transfer through IOS116. In addition, IOC 1618 generates individual Maps for each data channel device for translating external device addresses into physical addresses within MEM 112. As indicated in Fig. 16, each data channel device contains an individual Address Translation Map (MAP) 1632 and 1636. This allows IOS 116 to assign individual areas of MEM 112's physical address space to each data channel device. This feature provides protection against one data channel device writing into or reading from information belonging to another data channel device. In addition, IOC 1618 may generate overlapping address translation Maps for two or more data channel devices to allow these data channel devices to share a common area of MEM 112 physical address space.

Data transfer between IOS 116's data channel devices and MEM 112 is through DM 1610, which includes a Buffer memory (BUF) 1641. BUF 1641 allows MEM 112 and IOS 116 to operate asynchronously. DM 1610 also includes a Ring Grant Generator (RGG) 1644 which controls access of various data channel devices to MEM 112. RGG 1644 is designed to be flexible in apportioning access to MEM 112 among IOS 116's data channel devices as loads carried by various data channel devices varies. In addition, RGG 1644 insures that no one, or group, of data channel devices may monopolize access to MEM 112.

Referring to Fig. 17, a diagramic representation of RGG 1644's operation is shown. As described further in a following description, RGG 1644 may be regarded as a commutator scanning a number of ports which are assigned to various channel devices. For example, ports A, C, E, and G may be assigned to a BMC 1614, ports B and F to a NDC 1616, and ports D and H to another data channel device. RGG 1644 will scan each of these ports in turn and, if the data channel device associated with a particular port is requesting access to MEM 112, will grant access to MEM 112 to that data channel device. If no request is present at a given port, RGG 1644 will continue immediately to the next port. Each data channel device assigned one or more ports is thereby insured opportunity of access to MEM 112. Unused ports, for example indicating data channel devices which are not presently engaged in information transfer, are effectively skipped over so that access to MEM 112 is dynamically modified according to the information transfer loads of the various data channel devices. RGG 1644's ports may be reassigned among IOS 116's various data channel devices as required to suit the needs of a particular CS 101 system. If, for example, a particular CS 101 utilizes NDC 1616 more than a BMC 1614, that CS 101's NDC 1616 may be assigned more ports while that CS 101's BMC 1614 is assigned fewer ports.

2. Memory (MEM) 112 (Fig. 18)

Referring to Fig. 18, a partial block diagram of MEM 112 is shown. Major elements of MEM 112 are Main Store Bank (MSB) 1810, a Bank Controller (BC) 1814, a Memory Cache (MC) 1816, a Field Interface Unit (FIU) 1820, and Memory Interface Controller (MIC) 1822. Interconnections of these elements with input and output buses of MEM 112 to IOS 116 and JP 114 are indicated.

MEM 112 is an intelligent, prioritizing memory having a single port to IOS 116, comprised of IOM Bus 130, MIO Bus 129, and IOMC Bus 131, and dual ports to JP 114. A first JP 114 port is comprised of MOD Bus 140 and PD Bus 146, and a second port is comprised of JPD Bus 142 and PD Bus 146. In general, all data transfers from and to MEM 112 by IOS 116 and JP 114 are of single, 32 bit words; IOM Bus 130, MIO Bus 129, MOD Bus 140, and JPD Bus 142 are each 32 bits wide. CS 101, however, is a variable word length machine wherein the actual physical width of data buses are not apparent to a user. For example, a Name in a user's program may refer to an operand containing 97 bits of data. To the user, that 97 bit data item will appear to be read from MEM 112 to JP 114 in a single operation. In actuality, JP 114 will read that operand from MEM 112 in a series of read operations referred to as a string transfer. In this example, the string transfer will comprise three 32 bit read transfers and one single bit read transfer. The final single bit transfer, containing a single data bit, will be of a 32 bit word wherein one bit is data and 31 bits are fill. Write operations to MEM 112 may be performed in the same manner. If a single read or write request to MEM 112 specifies a data item of less than 32 bits of data, that transfer will be accomplished in the same manner as the final transfer described above. That is, a single 32 bit word will be transferred wherein non-data bits are fill bits.

Bulk data storage in MEM 112 is provided in MSB 1810, which is comprised of one or more Memory Array cards (MAs) 1812. The data path into and out of MA 1812 is through BC 1814, which performs all control and timing functions for MAs 1812. BC 1814's functions include addressing, transfer of data, controlling whether a read or write operation is performed, refresh, sniffing, and error correction code operations. All read and write operations from and to MAs 1812 through BC 1814 are in blocks of four 32 bit words.

The various MAs 1812 comprising MSB 1810 need not be of the same data storage capacity. For example, certain MAs 1812 may have a capacity of 256 kilobytes while other MAs 1812 may have a capacity of 512 kilobytes. Addressing of the MAs 1812 in MSB 1810 is automatically adapted to various MA 1812

configurations. As indicated in Fig. 18, each MA 1812 contains an address circuit (A) which receives an input from the next lower MA 1812 indicating the highest address in that next lower MA 1812. The A circuit on an MA 1812 also receives an input from that MA 1812 indicating the total address space of that MA 1812. The A circuit of that MA 1812 adds the highest address input from next lower MA 1812 to its own input representing its own capacity and generates an output to the next MA 1812 indicating its own highest address. All MAs 1812 of MSB 1810 are addressed in parallel by BC 1814. Each MA 1812 compares such addresses to its input from the next lower MA 1812, representing highest address of that next lower MA 1812, and its own output, representing its own highest address, to determine whether a particular address provided by BC 1814 lies within the range of addresses contained within that particular MA 1812. The particular MA 1812 whose address space includes that address will then respond by accepting the read or write request from BC 1814.

MC 1816 is the data path for transfer of data between BC 1814 and IOS 116 and JP 114. MC 1816 contains a high speed cache storing data from MSB 1810 which is currently being utilized by either IOS 116 or JP 114. MSB 1810 thereby provides MEM 112 with a large storage capacity while MC 1816 provides the appearance of a high speed memory. In addition to operating as a cache, MC 1816 includes a bypass write path which allows IOS 116 to write blocks of four 32 bit words directly into MSB 1810 through BC 1814. In addition, MC 1816 includes a cache write-back path which allows data to be transferred out of MC 1816's cache and stored while further data is transferred into MC 1816's cache. Displaced data from MC 1816's cache may then be written back into MSB 1810 at a later, more convenient time. This write-back path enhances speed of operation of MC 1816 by avoiding delays incurred by transferring data from MC 1816 to MSB 1810 before new data may be written into MC 1816.

MEM 112's FIU 1820 allows manipulation of data formats in writes to and reads from MEM 112 by both JP 114 and IOS 116. For example, FIU 1820 may convert unpacked decimal data to packed decimal data, and vice versa. In addition, FIU 1820 allows MEM 112 to operate as a bit addressable memory. For example, as described all data transfers to and from MEM 112 are of 32 bit words. If a data transfer of less than 32 bits is required, the 32 bit word containing those data bits may be read from MC 1816 to FIU 1820 and therein manipulated to extract the required data bits. FIU 1820 then generates a 32 bit word containing those required data bits, plus fill bits, and provides that new 32 bit word to JP 114 or IOS 116. When writing into MEM 112 from IOS 116 through FIU 1820, data is transferred onto IOM Bus 130, read into FIU 1820, operated upon, transferred onto MOD Bus 140, and transferred from MOD Bus 140 to MC 1816. In read operations from MEM 112 to IOS 116, data is transferred from MC 1816 to MOD Bus 140, written into FIU 1820 and operated upon, and transferred onto MIO Bus 129 to IOS 116. In a data read from MEM 112 to JP 114, data is transferred from MC 1816 onto MOD Bus 140, transferred into FIU 1820 and operated upon, and transferred again onto MOD Bus 140 to JP 114. In write operations from JP 114 to MEM 112, data on JPD Bus 142 is transferred into FIU 1820 and operated upon, and is then transferred onto MOD Bus 140 to MC 1816. MOD Bus 140 is thereby utilized as an MEM 112 internal bus for FIU 1820 operations.

Finally, MIC 1822 provides primary control of BC 1814, MC 1816, and FIU 1820. MIC 1822 receives control inputs from and provides control outputs to PD Bus 146 and IOMC Bus 131. MIC 1822 contains primary microcode control for MEM 112, but BC 1814, MC 1816, and FIU 1820 each include internal microcode control. Independent, internal microcode controls allow BC 1814, MC 1816, and FIU 1820 to operate independently of MIC 1822 after their operations have been initiated by MIC 1822. This allows BC 1814 and MSB 1810, MC 1816, and FIU 1820 to operate independently and asynchronously. Efficiency and speed of operation of MEM 112 are thereby enhanced by allowing pipelining of MEM 112 operations.

3. Fetch Unit (FU) 120 (Fig. 19)

A primary function of FU 120 is to execute SINS. In doing so, FU 120 fetches instructions and data (SOPs and Names) from MEM 112, returns results of operations to MEM 112, directs operation of EU 122, executes instructions of user's programs, and performs the various functions of CS 101's operating systems. As part of these functions, FU 120 generates and manipulates logical addresses and descriptors and is capable of operating as a general purpose CPU.

Referring to Fig. 19, a major element of FU 120 is the Descriptor Processor (DESP) 1910. DESP 1910 includes General Register File (GRF) 506. GRF 506 is a large register array divided vertically into three parts which are addressed in parallel. A first part, AONGRF 1932, stores AON fields of logical addresses and descriptors. A second part, OFFGRF 1934, stores offset fields of logical addresses and descriptors and is utilized as a 32 bit wide general register array. A third portion GRF 506, LENGRF 1936, is a 32 bit wide register array for storing length fields of logical descriptors and as a general register for storing data. Primary data path from MEM 112 to FU 120 is through MOD Bus 140, which provides inputs to OFFGRF 1934. As indicated in Fig. 19, data may be transferred from OFFGRF 1934 to inputs of AONGRF 1932 and LENGRF 1936 through various interconnections. Similarly, outputs from LENGRF 1936 and AONGRF 1932 may be transferred to inputs of AONGRF 1932, OFFGRF 1934, and LENGRF 1936.

Output of OFFGRF 1934 is connected to inputs of DESP 1910's Arithmetic and Logic Unit (ALU) 1942. ALU 1942 is a general purpose 32 bit ALU which may be used in generating and manipulating logical addresses and descriptors, as distinct from general purpose arithmetic and logic operands performed by MUX 1940. Output of ALU 1942 is connected to JPD Bus 142 to allow results of arithmetic and logic operations to be transferred to MEM 112 or EU 122.

EP 0 067 556 B1

Also connected from output of OFFGRF 1934 is Descriptor Multiplexer (MUX) 1940. An output of MUX 1940 is provided to an input of ALU 1942. MUX 1940 is a 32 bit ALU, including an accumulator, for data manipulation operations. MUX 1940, together with ALU 1942, allows DESP 1910 to perform 32 bit arithmetic and logic operations. MUX 1940 and ALU 1942 may allow arithmetic and logic operations upon operands of greater than 32 bits by performing successive operations upon successive 32 bit words of larger operands.

Logical descriptors or addresses generated or provided by DESP 1910, are provided to Logical Descriptor (LD) Bus 1902. LD Bus 1902 in turn is connected to an input of Address Translation Unit (ATU) 1928. ATU 1928 is a cache mechanism for converting logical descriptors to MEM 112 physical descriptors.

LD Bus 1902 is also connected to write input of Name Cache (NC) 1926. NC 1926 is a cache mechanism for storing logical descriptors corresponding to operand Names currently being used in user's programs. As previously described, Name Table Entries corresponding to operands currently being used in user's programs are stored in MEM 112. Certain Name Table Entries for operands of a user's program currently being executed are transferred from those Name Tables in MEM 112 to FU 120 and are therein evaluated to generate corresponding logical descriptors. These logical descriptors are then stored in NC 1926. As will be described further below, the instruction stream of a user's program is provided to FU 120's Instruction Buffer (IB) 1962 through MOD Bus 140. FU 120's Parser (P) 1964 separates out, or parses, Names from IB 1962 and provides those Names as address inputs to NC 1924. NC 1924 in turn provides logical descriptor outputs to LD Bus 1902, and thus to input of ATU 1928. NC 1926 input from LD Bus 1902 allows logical descriptors resulting from evaluation of Name Table Entries to be written into NC 1926. FU 120's Protection Cache (PC) 1934 is a cache mechanism having an input connected from LD Bus 1902 and providing information, as described further below, regarding protection aspects of references to data in MEM 112 by user's programs. NC 1926, ATU 1928, and PC 1934 are thereby acceleration mechanisms of, respectively, CS 101's Namespace addressing, logical to physical address structure, and protection mechanism.

Referring again to DESP 1910, DESP 1910 includes BIAS 1952, connected from output of LENGRF 1936. As previously described, operands containing more than 32 data bits are transferred between MEM 112 and JP 114 by means of string transfers. In order to perform string transfers, it is necessary for FU 120 to generate a corresponding succession of logical descriptors wherein length fields of those logical descriptors is no greater than 5 bits, that is, specify lengths of no greater than 32 data bits.

A logical descriptor describing a data item to be transferred by means of a string transfer will be stored in GRF 506. AON field of the logical descriptor will reside in AONGRF 1932, O field in OFFGRF 1934, and L field in LENGRF 1936. At each successive transfer of a 32 bit word in the string transfer, O field of that original logical descriptor will be incremented by the number of data bits transferred while L field will be accordingly decremented. The logical descriptor residing in GRF 506 will thereby describe, upon each successive transfer of the string transfer, that portion of the data item yet to be transferred. O field in OFFGRF 1934 will indicate increasingly larger offsets into that data item, while L field will indicate successively shorter lengths. AON and O fields of the logical descriptor in GRF 506 may be utilized directly as AON and O fields of the successive logical descriptors of the string transfer. L field of the logical descriptor residing in LENGRF 1936, however, may not be so used as L fields of the successive string transfer logical descriptors as this L field indicates remaining length of data item yet to be transferred. Instead, BIAS 1952 generates the 5 bit L fields of successive string transfer logical descriptors while correspondingly decrementing L field of the logical descriptor in LENGRF 1936. During each transfer, BIAS 1952 generates L field of the *next* string transfer logical descriptor while concurrently providing L field of the *current* string transfer logical descriptor. By doing so, BIAS 1952 thereby increases speed of execution of string transfers by performing pipelined L field operations. BIAS 1952 thereby allows CS 101 to appear to the user to be a variable word length machine by automatically performing string transfers. This mechanism is used for transfer of any data item greater than 32 bits, for example double precision floating point numbers.

Finally, FU 120 includes microcode circuitry for controlling all FU 120 operations described above. In particular, FU 120 includes a microinstruction sequence control store (mC) 1920 storing sequences of microinstructions for controlling step by step execution of all FU 120 operations. In general, these FU 120 operations fall into two classes. A first class includes those microinstruction sequences directly concerned with executing the SOPs of user's programs. The second class includes microinstruction sequences concerned with CS 101's operating systems, including and certain automatic, internal FU 120 functions such as evaluation of Name Table Entries.

As previously described, CS 101 is a multiple S-Language machine. For example, mC 1920 may contain microinstruction sequences for executing user's SOPs in at least four different Dialects. mC 1920 is comprised of a writable control store and sets of microinstruction sequences for various Dialects may be transferred into and out of mC 1920 as required for execution of various user's programs. By storing sets of microinstruction sequences for more than one Dialect in mC 1920, it is possible for user's programs to be written in a mixture of user languages. For example, a particular user's program may be written primarily in FORTRAN but may call certain COBOL routines. These COBOL routines will be correspondingly translated into COBOL dialect SOPs and executed by COBOL microinstruction sequences stored in mC 1920.

The instruction stream provided to FU 120 from MEM 112 has been previously described with

EP 0 067 556 B1

reference to Fig. 3. SOPs and Names of this instruction stream are transferred from MOD Bus 140 into IB 1962 as they are provided from MEM 112. IB 1962 includes two 32 bit (one word) registers. IB 1962 also includes prefetch circuitry for reading for SOPs and Names of the instruction stream from MEM 112 in such a manner that IB 1962 shall always contain at least one SOPs or Name. FU 120 includes (P) 1964 which reads and separates, or parses, SOPs and Names from IB 1962. As previously described, P 1964 provides those Names to NC 1926, which accordingly provides logical descriptors to ATU 1928 so as to read the corresponding operands from MEM 112.

SOPs parsed by P 1964 are provided as inputs to Fetch Unit Dispatch Table (FU DT) 1904 and Execute Unit Dispatch Table (EU DT) 1966. Referring first to FU DT 1904, FU DT 1904 is effectively a table for translating SOPs to starting addresses in mC 1912 of corresponding microinstruction sequences. This intermediate translation of SOPs to mC 1912 addresses allows efficient packing of microinstruction sequences within mC 1912. That is, certain microinstruction sequences may be common to two or more S-Language Dialects. Such microinstruction sequences may therefore be written into mC 1912 once and may be referred to by different SOPs of different S-Language Dialects.

EU DT 1966 performs a similar function with respect to EU 122. As will be described below, EU 122 contains a mC, similar to mC 1912, which is addressed through EU DT 1966 by SOPs specifying EU 122 operations. In addition, FU 120 may provide such addresses mC 1912 to initiate EU 122 operations as required to assist certain FU 120 operations. Examples of such operations which may be requested by FU 120 include calculations required in evaluating Name Table Entries to provide logical descriptors to be loaded into NC 1926.

Associated with both FU DT 1904 and EU DT 1966 are Dialect (D) registers 1905 and 1967. D registers 1905 and 1967 store information indicating the particular S-Language Dialect currently being utilized in execution of a user's program. Outputs of D registers 1905 and 1967 are utilized as part of the address inputs to mC 1912 and EU 122's mC.

4. Execute Unit (EU) 122 (Fig. 20)

As previously described, EU 122 is an arithmetic and logic unit provided to relieve FU 120 of certain arithmetic operations. EU 122 is capable of performing addition, subtraction, multiplication, and division operations on integer, packed and unpacked decimal, and single and double precision floating operands. EU 122 is an independently operating microcode controlled machine including Microcode Control (mC) 2010 which, as described above, is addressed by EU DT 1966 to initiate EU 122 operations. mC 2010 also includes logic for handling mutual interrupts between FU 120 and EU 122. That is, FU 120 may interrupt current EU 122 operations to call upon EU 122 to assist an FU 120 operation. For example, FU 120 may interrupt an arithmetic operation currently being executed by EU 122 to call upon EU 122 to assist in generating a logical descriptor from a Name Table Entry.

Similarly, EU 122 may interrupt current FU 120 operations when EU 122 requires FU 120 assistance in executing a current arithmetic operation. For example, EU 122 may interrupt a current FU 120 operation if EU 122 receives an instruction and operands requiring EU 122 to perform a divide by zero.

Referring to Fig. 20, a partial block diagram of EU 122 is shown. EU 122 includes two arithmetic and logic units. A first arithmetic and logic unit (MULT) 2014 is utilized to perform addition, subtraction, multiplication, and division operations upon integer and decimal operands, and upon mantissa fields of single and double precision floating point operands. Second ALU (EXP) 2016 is utilized to perform operations upon single and double precision floating point operand exponent fields in parallel with operations performed upon floating point mantissa fields by MULT 2014. Both MULT 2014 and EXP 2016 include an arithmetic and logic unit, respectively MALU 2074 and EXPALU 2084. MULT 2014 and EXP 2016 also include register files, respectively MRF 2050 and ERF 2080, which operate and are addressed in parallel in a manner similar to AONGRF 1932, OFFGRF 1984 and LENGRF 1936.

Operands for EU 122 to operate upon are provided from MEM 112 through MOD Bus 140 and are transferred into Operand Buffer (OPB) 2022. In addition to serving as an input buffer, OPB 2022 performs certain data format manipulation operations to transform input operands into formats most efficiently operated with by EU 122. In particular, EU 122 and MULT 2014 may be designed to operate efficiently with packed decimal operands. OPB 2022 may transform unpacked decimal operands into packed decimal operands. Unpacked decimal operands are in the form of ASCII characters wherein four bits of each character are binary codes specifying a decimal value between zero and nine. Other bits of each character are referred to as zone fields and in general contain information identifying particular ASCII characters. For example, zone field bits may specify whether a particular ASCII character is a number, a letter, or punctuation. Packed decimal operands are comprised of a series of four bit fields wherein each field contains a binary number specifying a decimal value of between zero and nine. OPB 2022 converts unpacked decimal to packed decimal operands by extracting zone field bits and packing the four numeric value bits of each character into the four bit fields of a packed decimal number.

EU 122 is also capable of transforming the results of arithmetic operands, for example in packed decimal format, into unpacked decimal format for transfer back to MEM 112 or FU 120. In this case, a packed decimal result appearing at output of MALU 2074 is written into MRF 2050 through a multiplexer, not shown in Fig. 20, which transforms the four bit numeric code fields of the packed decimal results into corresponding bits of unpacked decimal operand characters, and forces blanks into the zone field bits of

those unpacked decimal characters. The results of this operation are then read from MRF 2050 to MALU 2074 and zone field bits for those unpacked decimal characters are read from Constant Store (CST) 2060 to MALU 2074. These inputs from MRF 2050 and CST 2060 are added by MALU 2074 to generate final result outputs in unpacked decimal format. These final results may then be transferred onto JPD Bus 142 through Output Multiplexer (OM) 2024.

Considering first floating point operations, in addition or subtraction of floating point operands it is necessary to equalize the values of the floating point operand exponent fields. This is referred to as prealignment. In floating point operations, exponent fields of the two operands are transferred into EXPALU 2034 and compared to determine the difference between exponent fields. An output representing difference between exponent fields is provided from EXPALU 2034 to an input of floating point control (FPC) 2002. FPC 2002 in turn provides control outputs to MALU 2074, which has received the mantissa fields of the two operands. MALU 2074, operating under direction of FPC 2002, accordingly right or left shifts one operand's mantissa field to effectively align that operand's exponent field with the other operand's exponent field. Addition or subtraction of the operand's mantissa fields may then proceed.

EXPALU 2034 also performs addition or subtraction of floating point operand exponent fields in multiplication or division operations, while MALU 2074 performs multiplication and division of the operand mantissa fields. Multiplication and division of floating point operand mantissa fields by MALU 2074 is performed by successive shifting of one operand, corresponding generation of partial products of the other operand, and successive addition and subtraction of those partial products.

Finally, EU 122 performs normalization of the results of floating point operand operations by left shifting of a final result's mantissa field to eliminate zeros in the most significant characters of the final result mantissa field, and corresponding shifting of the final result exponent fields. Normalization of floating point operation results is controlled by FPC 2002. FPC 2002 examines an unnormalized floating point result output of MALU 2074 to detect which, if any, of the most significant characters of that result contain zeros. FPC 2002 then accordingly provides control outputs to EXPALU 2034 and MALU 2074 to correspondingly shift the exponent and mantissa fields of those results so as to eliminate leading character zeros from the mantissa field. Normalized mantissa and exponent fields of floating point results may then be transferred from MALU 2074 and EXPALU 2034 to JPD Bus 142 through OM 2024.

As described above, EU 122 also performs addition, subtraction, multiplication, and division operations on operands. In this respect, EU 122 uses a leading zero detector in FPC 2002 in efficiently performing multiplication and division operations. FPC 2002's leading zero detector examines the characters or bits of two operands to be multiplied or divided, starting from the highest, to determine which, if any, contain zeros so as not to require a multiplication or division operation. FPC 2002 accordingly left shifts the operands to effectively eliminate those characters or bits, thus reducing the number of operations to multiply or divide the operands and accordingly reducing the time required to operate upon the operands.

Finally, EU 122 utilizes a unique method, with associated hardware, for performing arithmetic operations on decimal operands by utilizing circuitry which is otherwise conventionally used only to perform operations upon floating point operands. As described above, MULT 2074 is designed to operate with packed decimal operands, that is operands in the form of consecutive blocks of four bits wherein each block of four bits contains a binary code representing numeric values of between zero and nine. Floating point operands are similarly in the form of consecutive blocks of four bits. Each block of four bits in a floating point operand, however, contains a binary number representing a hexadecimal value of between zero and fifteen. As an initial step in operating with packed decimal operands, those operands are loaded, one at a time, into MALU 2074 and, with each such operand, a number comprised of all hexadecimal sixes is loaded into MALU 2074 from CST 2060. This CST 2060 number is added to each packed decimal operand to effectively convert those packed decimal operands into hexadecimal operands wherein the four bit blocks contain numeric values in the range of six to fifteen, rather than in the original range of zero to nine. MULT 2014 then performs arithmetic operation upon those transformed operands, and in doing so detects and saves information regarding which four bit characters of those operands have resulted in generation of carries during the arithmetic operations. In a final step, the intermediate result resulting from completion of those arithmetic operations upon those transformed operands are reconverted to packed decimal format by subtraction of hexadecimal sixes from those characters for which carries have been generated. Effectively, EU 122 converts packed decimal operands into "Excess Six" operands, performs arithmetic operations upon those "Excess Six" operands, and reconverts "Excess Six" results of those operations back into packed decimal format.

Finally, as previously described FU 120 controls transfer of arithmetic results from EU 122 to MEM 112. In doing so, FU 120 generates a logical descriptor describing the size of MEM 112 address space, or "container", that result is to be transferred into. In certain arithmetic operations, for example integer operations, an arithmetic result may be larger than anticipated and may contain more bits than the MEM 112 "container". Container Size Check Circuit (CSC) 2052 compares actual size of arithmetic results and L fields of MEM 112 "container" logical descriptors. CSC 2052 generates an output indicating whether an MEM 112 "container" is smaller than an arithmetic result.

Having briefly described certain features of CS 101 structure and operation in the above overview, these and other features of CS 101 will be described in further detail next below in a more detailed

EP 0 067 556 B1

introduction of CS 101 structure and operation. Then, in further descriptions, these and other features of CS 101 structure and operation will be described in depth.

1. Introduction (Figs. 101—110)

A. General Structure and Operation (Fig 101)

a. General Structure

Referring to Fig. 101, a partial block diagram of Computer System (CS) 10110 is shown. Major elements of CS 10110 are Dual Port Memory (MEM) 10112, Job Processor (JP) 10114, Input/Output System (IOS) 10116, and Diagnostic Processor (DP) 10118. JP 10114 includes Fetch Unit (FU) 10120 and Execute Unit (EU) 10122.

Referring first to IOS 10116, IOS 10116 is interconnected with External Devices (ED) 10124 through Input/Output (I/O) Bus 10126. ED 10124 may include, for example, other computer systems, keyboard/display units, and disc drive memories. IOS 10116 is interconnected with Memory Input/Output (MIO) Port 10128 of MEM 10112 through Input/Output to Memory (IOM) Bus 10130 and Memory to Input/Output (MIO) Bus 10129, and with FU 10120 through I/O Job Processor (IOJP) Bus 10132.

DP 10118 is interconnected with, for example, external keyboard/CRT Display Unit (DU) 10134 through Diagnostic Processor Input/Output (DPIO) Bus 10136. DP 10118 is interconnected with IOS 10116, MEM 10112, FU 10120, and EU 10122 through Diagnostic Processor (DP) Bus 10138.

Memory to Job Processor (MJP) Port 10140 of Memory 10112 is interconnected with FU 10120 and EU 10122 through Job Processor Data (JPD) Bus 10142. An output of MJP 10140 is connected to inputs of FU 10120 and EU 10122 through Memory Output Data (MOD) Bus 10144. An output of FU 10120 is connected to an input of MJP 10140 through Physical Descriptor (PD) Bus 10146. FU 10120 and EU 10122 are interconnected through Fetch/Execute (F/E) Bus 10148.

b. General Operation

As will be discussed further below, IOS 10116 and MEM 10112 operate independently under general control of JP 10114 in executing multiple user's programs. In this regard, MEM 10112 is an intelligent, prioritizing memory having separate and independent ports MIO 10128 and MJP 10140 to IOS 10116 and JP 10114 respectively. MEM 10112 is the primary path for information transfer between External Devices 10124 (through IOS 10116) and JP 10114. MEM 10112 thus operates both as a buffer for receiving and storing various individual user's programs (e.g., data, instructions, and results of program execution) and as a main memory for JP 10114.

A primary function of IOS 10116 is as an input/output buffer between CS 10110 and ED 10124. Data and instructions are transferred from ED 10124 to IOS 10116 through I/O Bus 10126 in a manner and format compatible with ED 10124. IOS 10116 receives and stores this information, and manipulates the information into formats suitable for transfer into MEM 10112. IOS 10116 then indicates to MEM 10112 that new information is available for transfer into MEM 10112. Upon acknowledgement by MEM 10112, this information is transferred into MEM 10112 through IOM Bus 10130 and MIO Port 10128. MEM 10112 stores the information in selected portions of MEM 10112 physical address space. At this time, IOS 10116 notifies JP 10114 that new information is present in MEM 10112 by providing a "semaphore" signal to FU 10120 through IOJP Bus 10132. As will be described further below, CS 10110 manipulates the data and instructions stored in MEM 10112 into certain information structures used in executing user's programs. Among these structures are certain structures, discussed further below, which are used by CS 10110 in organizing and controlling flow and execution of user programs.

FU 10120 and EU 10122 are independently operating microcode controlled "machines" together comprising the CS 10110 micromachine for executing user's programs stored in MEM 10112. Among the principal functions of FU 10120 are: (1) fetching and interpreting instructions and data from MEM 10112 for use by FU 10120 and EU 10122; (2) organizing and controlling flow of user programs; (3) initiating EU 10122 operations; (4) performing arithmetic and logic operations on data; (5) controlling transfer of data from FU 10120 and EU 10122 to MEM 10112; and, (6) maintaining certain "stack" and "register" mechanisms, described below. FU 10120 "cache" mechanisms, also described below, are provided to enhance the speed of operation of JP 10114. These cache mechanisms are acceleration circuitry including, in part, high speed memories for storing copies of selected information stored in MEM 10112. The information stored in this acceleration circuitry is therefore more rapidly available to JP 10114. EU 10122 is an arithmetic unit capable of executing integer, decimal, or floating point arithmetic operations. The primary function of EU 10122 is to relieve FU 10120 from certain extensive arithmetic operations, thus enhancing the efficiency of CS 10110.

In general, operations in JP 10114 are executed on a memory to memory basis; data is read from MEM 10112, operated upon, and the results returned to MEM 10112. In this regard, certain stack and cache mechanisms in JP 10114 (described below) operate as extensions of MEM 10112 address space.

In operation, FU 10120 reads data and instructions from MEM 10112 by providing physical addresses to MEM 10112 by way of PA Bus 10146 and MJP Port 10140. The instructions and data are transferred to FU 10120 and EU 10122 by way of MJP Port 10140 and MOD Bus 10144. Instructions are interpreted by FU 10120 microcode circuitry, not shown in Fig. 101 but described below, and when necessary, microcode instructions are provided to EU 10122 from FU 10120's microcode control by way of F/E Bus 10148, or by way of JPD Bus 10142.

As stated above, FU 10120 and EU 10122 operate asynchronously with respect to each other's functions. A microinstruction from FU 10120 microcode circuitry to EU 10122 may initiate a selected operation of EU 10122. EU 10122 may then proceed to independently execute the selected operation. FU 10120 may proceed to concurrently execute other operations while EU 10122 is completing the selected arithmetic operation. At completion of the selected arithmetic operation, EU 10122 signals FU 10120 that the operation results are available by way of a "handshake" signal through F/E Bus 10148. FU 10120 may then receive the arithmetic operation results for further processing or, as discussed momentarily, may directly transfer the arithmetic operation results to MEM 10112. As described further below, an instruction buffer referred to as a "queue" between FU 10120 and EU 10122 allows FU 10120 to assign a sequence of arithmetic operations to be performed by EU 10122.

Information, such as results of executing an instruction, is written into MEM 10112 from FU 10120 or EU 10122 by way of JPD Bus 10142. FU 10120 provides a "physical write address" signal to MEM 10112 by way of PA Bus 10146 and MJP Port 10140. Concurrently, the information to be written into MEM 10112 is placed on JPD Bus 10142 and is subsequently written into MEM 10112 at the locations selected by the physical write address.

FU 10120 places a semaphore signal on IOJP Bus 10132 to signal to IOS 10116 that information, such as the results of executing a user's program, is available to be read out of CS 10110. IOS 10116 may then transfer the information from MEM 10112 to IOS 10116 by way of MIO Port 10128 and IOM Bus 10130. Information stored in IOS 10116 is then transferred to ED 10124 through I/O Bus 10126.

During execution of a user's program, certain information required by JP 10116 may not be available in MEM 10112. In such cases as further described in a following discussion, JP 10114 may write a request for information into MEM 10112 and notify IOS 10116, by way of IOJP Bus 10132, that such a request has been made. IOS 10116 will then read the request and transfer the desired information from ED 10124 into MEM 10112 through IOS 10116 in the manner described above. In such operations, IOS 10116 and JP 10114 operate together as a memory manager wherein the memory space addressable by JP 10114 is termed virtual memory space, and includes both MEM 10112 memory space and all external devices to which IOS 10116 has access.

As previously described, DP 10118 provides a second interface between Computer System 10110 and the external world by way of DPIO Bus 10136. DP 10118 allows DU 10134, for example a CRT and keyboard unit or a teletype, to perform all functions which are conventionally provided by a hard (i.e., switches and lights) console. For example, DP 10118 allows DU 10134 to exercise control of Computer System 10110 for such purposes as system initialization and start up, execution of diagnostic processes, and fault monitoring and identification. DP 10118 has read and write access to most memory and register portions within each of IOS 10116, MEM 10112, FU 10120, and EU 10122 by way of DP Bus 10138. Memories and registers in CS 10110 can therefore be directly loaded or initialized during system start up, and can be directly read or loaded with test and diagnostic signals for fault monitoring and identification. In addition, as described further below, microinstructions may be loaded into JP 10114's microcode circuitry at system start up or as required.

Having described the general structure and operation of Computer System 10110, certain features of Computer System 10110 will next be briefly described to aid in understanding the following, more detailed descriptions of these and other features of Computer System 10110.

c. Definition of Certain Terms

Certain terms are used relating to the structure and operation of CS 10110 throughout the following discussions. Certain of these terms will be discussed and defined first, to aid in understanding the following descriptions. Other terms will be introduced in the following descriptions as required.

A *procedure* is a sequence of operational steps, or instructions, to be executed to perform some operation. A procedure may include data to be operated upon in performing the operation.

A *program* is a static group of one or more procedures. In general, programs may be classified as user programs, utility programs, and operating system programs. A user program is a group of procedures generated by and private to one particular user of a group of users interfacing with CS 10110. Utility programs are commonly available to all users; for example, a compiler comprises of a set of procedures for compiling a user language program into an S-language program. Operating system programs are groups of procedures internal to CS 10110 for allocation and control of CS 10110 resources. Operating system programs also define interfaces within CS 10110. For example, as will be discussed further below all operands in a program are referred to by "NAME". An operating system program translates operand NAME into the physical locations of the operands in MEM 10112. The NAME translation program thus defines the interface between operand NAME (name space addresses) and MEM 10112 physical addresses.

A *process* is an independent locus of control passing through physical, logical or virtual address spaces, or, more particularly, a path of execution through a series of programs (i.e., procedures). A process will generally include a user program and data plus one or more utility programs (e.g., a compiler) and operating system programs necessary to execute the user program.

An *object* is a uniquely identifiable portion of "data space" accessible to CS 10110. An object may be regarded as a container for information and may contain data or procedure information or both. An object may contain for example, an entire program, or set of procedures, or a single bit of data. Objects need not

be contiguously located in the data space accessible to CS 10110, and the information contained in an object need not be contiguously located in that object.

A *domain* is a state of operation of CS 10110 for the purposes of CS 10110's protection mechanisms. Each domain is defined by a set of procedures having access to objects within that domain for their execution. Each object has a single domain of execution in which it is executed if it is a procedure object, or used, if it is a data object. CS 10110 is said to be operating in a particular domain if it is executing a procedure having that domain of execution. Each object may belong to one or more domains; an object belongs to a domain if a procedure executing in that domain has potential access to the object. CS 10110 may, for example have four domains: User domain, Data Base Management System (DBMS) domain, Extended Operating System (EOS) domain, and Kernel Operating System (KOS) domain. User domain is the domain of execution of all user provided procedures, such as user or utility procedures. DBMS domain is the domain of execution for operating system procedures for storing, retrieving, and handling data. EOS domain is the domain of execution of operating system procedures defining and forming the user level interface with CS 10110, such as procedures for controlling an executing files, processes, and I/O operations. KOS domain is the domain of execution of the low level, secure operating system which manages and controls CS 10110's physical resources. Other embodiments of CS 10110 may have fewer or more domains than those just described. For example, DBMS procedures may be incorporated into the EOS domain or EOS domain may be divided by incorporating the I/O procedures into an I/O domain. There is no hardware enforced limitation on the number of, or boundaries between, domains in CS 10110. Certain CS 10110 hardware functions and structures are, however, dependant upon domains.

A *subject* is defined, for purposes of CS 10110's protection mechanisms, as a combination of the current principle (user), the current process being executed, and the domain the process is currently being executed in. In addition to principle, process, and domain, which are identified by UIDs, subject may include a Tag, which is a user assigned identification code used where added security is required. For a given process, principle and process are constant but the domain is determined by the procedure currently being executed. A process's associated subject is therefore variable along the path of execution of the process.

Having discussed and defined the above terms, certain features of CS 10110 will next be briefly described.

d. Multi-Program Operation

CS 10110 is capable of concurrently executing two or more programs and selecting the sequence of execution of programs to make most effective use of CS 10110's resources. This is referred to as multiprogramming. In this regard, CS 10110 may temporarily suspend execution of one program, for example when a resource or certain information required for that program is not immediately available, and proceed to execute another program until the required resource or information becomes available. For example, particular information required by a first program may not be available in MEM 10112 when called for. JP 10114 may, as discussed further below, suspend execution of the first program, transfer a request for that information to IOS 10116, and proceed to call and execute a second program. IOS 10116 would fetch the requested information from ED 10124 and transfer it into MEM 10112. At some time after IOS 10116 notifies JP 10114 that the requested information is available in MEM 10112, JP 10114 could suspend execution of the second program and resume execution of the first program.

e. Multi-Language Operation

As previously described, CS 10110 is a multiple language machine. Each program written in a high level user language, such as COBOL or FORTRAN, is compiled into a corresponding Soft (S) Language program. That is, in terms of a conventional computer system, each user level language has a corresponding machine language, classically defined as an assembly language. In contrast to classical assembly languages, S-Languages are mid-level languages wherein each command in a user's high level language is replaced by, in general, two or three S-Language instructions, referred to as S-Interpreters. Certain S-Interpreters interpret S-Interpreters and provide corresponding sequences of microinstructions for detailed control of CS 10110. CS 10110's instruction set and operation may therefore be tailored to each user's program, regardless of the particular user language, so as to most efficiently execute the user's program. Computer System 10110 may, for example, execute programs in both FORTRAN and COBOL with comparable efficiency. In addition, a user may write a program in more than one high level user language without loss of efficiency. For example, a user may write a portion of his program in COBOL, but may wish to write certain portions in FORTRAN. In such cases, the COBOL portions would be compiled into COBOL S-Interpreters and executed with the COBOL dialect S-Interpreter. The FORTRAN portions would be compiled into FORTRAN S-Interpreters and executed with a FORTRAN dialect S-Interpreter. The present embodiment of CS 10110 utilizes a uniform format for all S-Interpreters. This feature allows simpler S-Interpreter structures and increases efficiency of S-Interpreter interpretation because it is not necessary to provide means for interpreting each dialect individually.

f. Addressing Structure

Each object created for use in, or by operation of, a CS 10110 is permanently assigned a Unique Identifier (UID). An object's UID allows that object to be uniquely identified and located at any time, regardless of which particular CS 10110 it was created by or for or where it is subsequently located. Thus
 5 each time a new object is defined, a new and unique UID is allocated, much as social security numbers are allocated to individuals. A particular piece of information contained in an object may be located by a logical address comprising the object's UID, an offset from the start of the object of the first bit of the segment, and the length (number of bits) of the information segment. Data within an object may therefore be addressed on a bit granular basis. As will be described further in following discussions, UID's are used within a CS
 10 10110 as logical addresses, and, for example, as pointers. Logically, all addresses and pointers in CS 10110 are UID addresses and pointers. As previously described and as described below, however, short, temporary unique identifiers, valid only within JP 10114 and referred to as Active Object Numbers are used within JP 10114 to reduce the width of address buses and amount of address information handled.

An object becomes active in CS 10110 when it is transferred from backing store CED 10124 to MEM
 15 10112 for use in executing a process. At this time, each such object is assigned an Active Object Number (AON). AONs are short unique identifiers and are related to the object's UIDs through certain CS 10110 information structures described below. AONs are used only within JP 10114 and are used in JP 10114, in place of UIDs, to reduce the required width of JP 10114's address buses and the amount of address data handled in JP 10114. As with UID logical addresses, a piece of data in an object may be addressed through
 20 a bit granular AON logical address comprising the object's AON, an offset from the start of the object of the first bit of the piece, and the length of the piece.

The transfer of logical addresses, for example pointers, between MEM 10112 (UIDA) and JP 10114 (AONs) during execution of a process requires translations between UIDs and AONs. As will be described in a later discussion, this translation is accomplished, in part, through the information structures
 25 mentioned above. Similarly, translation of logical addresses to physical addresses in MEM 10112, to physically access information stored in MEM 10112, is accomplished through CS 10110 information structures relating AON logical addresses to MEM 10112 physical addresses.

Each operand appearing in a program is assigned a Name when the program is compiled. Thereafter, all references to the operands are through their assigned Names. As will be described in detail in a later
 30 discussion, CS 10110's addressing structure includes a mechanism for recognizing Names as they appear in an instruction stream and Name Tables containing directions for resolving Names to AON logical addresses. AON logical addresses may then be evaluated, for example translated into a MEM 10112 physical address, to provide actual operands. The use of Names to identify operands in the instructions stream (process) (1) allows a complicated address to be replaced by a simple reference of uniform format;
 35 (2) does not require that an operation be directly defined by data type in the instruction stream; (3) allows repeated references to an operand to be made in an instruction stream by merely repeating the operand's Name; and, (4) allows partially completed Name to address translations to be stored in a cache to speed up operand references. The use of Names thereby substantially reduces the volume of information required in the instruction stream for operand references and increases CS 10110 speed and efficiency by performing
 40 operands references through a parallel operating, underlying mechanism.

Finally, CS 10110 address structure incorporates a set of Architectural Base Pointers (ABPs) for each process. ABPs provide an addressing framework to locate data and procedure information belonging to a process and are used, for example, in resolving Names to AON logical addresses.

g. Protection Mechanism

CS 10110's protection mechanism is constructed to prevent a user from (1) gaining access to or
 45 disrupting another user's process, including data, and (2) interfering with or otherwise subverting the operation of CS 10110. Access rights to each particular active object are dynamically granted as a function of the currently active subject. A subject is defined by a combination of the current principle (user), the
 50 current process being executed, and the domain in which the process is currently being executed. In addition to principle, process, and domain, subject may include a Tag, which is a user assigned identification code used where added security is required. For a given process, the principle and process are constant but the domain is determined by the procedure currently being executed. A process's associated subject is therefore variable along the path of execution of the process.

In a present embodiment of CS 10110, procedures having KOS domain of execution have access to
 55 objects in KOS, EOS, DBMS, and User domains; procedures having EOS domain of execution have access to objects in EOS, DBMS, and User domains; procedures having DBMS domain of execution have access to objects in DBMS and User domains; and procedures having User domain of execution have access only to objects in User domain. A user cannot, therefore, obtain access to objects in KOS domain of execution and cannot influence CS 10110's low level, secure operating system. The user's process may, however, call for
 60 execution a procedure having KOS domain of execution. At this point the process's subject is in the KOS domain and the procedure will have access to certain objects in KOS domain.

In a present embodiment of CS 10110, also described in a later discussion, each object has associated
 65 with it an Access Control List (ACL). An ACL contains an Access Control Entry (ACE) for each subject having access to that object. ACEs specify, for each subject, access rights a subject has with regard to that object.

There is normally no relationship, other than that defined by an object's ACL, between subjects and objects. CS 10110, however, supports Extended Type Objects having Extended ACLs wherein a user may specifically define which subjects have what access rights to the object.

5 In another embodiment of CS 10110, described in a following discussion, access rights are granted on a dynamic basis. In executing a process, a procedure may call a second procedure and pass an argument to the called procedure. The calling procedure will also pass selected access rights to that argument to the called procedure. The passed access rights exist only for the duration of the call.

10 In the dynamic access embodiment, access rights are granted only at the time they are required. In the ACL embodiment, access rights are granted upon object creation or upon specific request. In either embodiment, each procedure to which arguments may be passed in a cross-domain call has associated with it an Access Information Array (AIA). A procedure's AIA states what access rights a calling procedure (subject) must have before the called procedure can operate on the passed argument. CS 10110's protection mechanisms compare the calling procedure's access rights to the rights required by the called procedure. This ensures that a calling procedure may not ask a called procedure to do what the calling procedure is not allowed to do. Effectively, a calling procedure can pass to a called procedure only the access rights held by the calling procedure.

15 Having described the general structure and operation and certain features of CS 10110, those and other features of CS 10110 operation will next be described in greater detail.

20 B. Computer System 10110 Information Structures and Mechanisms (Figs. 102, 103, 104, 105)

CS 10110 contains certain information structures and mechanisms to assist in efficient execution of processes. These structures and mechanisms may be considered as falling into three general types. The first type concerns the processes themselves, i.e., procedure and data objects comprising a user's process or directly related to execution of a user's process. The second type are for management, control, and execution of processes. These structures are generally shared by all processes active in CS 10110. The third type are CS 10110 micromachine information structures and mechanisms. These structures are concerned with the eternal operation of the CS 10110 micromachine and are private to the CS 10110 micro-machine.

25 a. Introduction (Fig. 102)

30 Referring to Fig. 102, a pictorial representation of CS 10110 (MEM 10112, FU 10120, and EU 10122) is shown with certain information structures and mechanisms depicted therein. It should be understood that these information structures and mechanisms transcend or "cut across" the boundaries between MEM 10112, FU 10120, EU 10122, and IOS 10116. Referring to the upper portion of Fig. 103 Process Structures 10210 contains those information structures and mechanisms most closely concerned with individual processes, the first and third types of information structures described above. Process Structures 10210 reside in MEM 10112 and Virtual Processes 10212 include Virtual Processes (VP) 1 through N. Virtual Processes 10212 may contain, in a present embodiment of CS 10110, up to 256 VP's. As previously described, each VP includes certain objects particular to a single user's process, for example stack objects previously described and further described in a following description. Each VP also includes a Process Object containing certain information required to execute the process, for example pointers to other process information.

35 Virtual Processor State Blocks (VPSBs) 10218 include VPSBs containing certain tables and mechanisms for managing execution of VPs selected for execution by CS 10110.

40 A particular VP is bound into CS 10110 when a Virtual Process Dispatcher, described in a following discussion selects that VP as eligible for execution. The selected VPs Process Object, as previously described, is swapped into a VPSB. VPSBs 10218 may contain, for example 16 or 32 State Blocks so that CS 10110 may concurrently execute to 16 or 32 VPs. When a VP assigned to a VPSB is to be executed, the VP is swapped onto the Information structures and mechanisms shown in FU 10120 and EU 10122. FU Register and Stack Mechanism (FURSM) 10214 and EU Register and Stack Mechanism (EURSM) 10216, shown respectively in FU 10120 and EU 10122, comprise register and stack mechanisms used in execution of VPs bound to CS 10110. These register and stack mechanisms, as will be discussed below, are also used for certain CS 10110 process management functions. Procedure Objects (POs) 10213 contains Procedure Objects (POs) 1 to N of the processes executing in CS 10110.

45 Addressing Mechanisms (AM) 10220 are a part of CS 10110's process management system and are generally associated with Computer System 10110 addressing functions as described in following discussions. UID/AON Tables 10222 is a structure for relating UID's and AON's, previously discussed. Memory Management Tables 10224 includes structures for (1) relating AON logical addresses and MEM 10112 physical addresses; (2) managing MEM 10112's physical address space; (3) managing transfer of information between MEM 10112 and CS 10110's backing store (ED 10124) and, (4) activating objects into CS 10110; Name Cache (NC) 10226 and Address Translation Cache (ATC) 10228 are acceleration mechanisms for storing addressing information relating to the VP currently bound to CS 10110. NC 10226, described further below, contains information relating operand Names to AON addresses. ATC 10228, also discussed further below, contains information relating AON addresses to MEM 10112 physical addresses.

50 Protection Mechanisms 10230, depicted below AM 10220, include protection Tables 10232 and Protection Cache (PC) 10234. Protection Tables 10232 contain information regarding access rights to each

object active in CS 10110. PC 10234 contains protection information relating to certain objects of the VP currently bound to CS 10110.

Microinstruction Mechanisms 10236, depicted below PM 10230, includes Micro-code (M Code) Store 10238, FU (Micro-code) M Code Structure 10240, and EU Micro-code (M Code) Structure 10242. These structures contain microinstruction mechanisms and tables for interpreting SInS and controlling the detailed operation of CS 10110. Micro-instruction Mechanisms 10232 also provide microcode tables and mechanisms used, in part, in operation of the low level, secure operating system that manages and controls CS 10110's physical resources.

Having thus briefly described certain CS 10110 information structures and mechanisms with the aid of Fig. 102, those information structures and mechanisms will next be described in further detail in the order mentioned above. In these descriptions it should be noted that, in representation of MEM 10112 shown in Fig. 102 and in other figures of following discussions, the addressable memory space of MEM 10112 is depicted. Certain portions of MEM 10112 address space have been designated as containing certain information structures and mechanisms. These structures and mechanisms have real physical existence in MEM 10112, but may vary in both location and volume of MEM 10112 address space they occupy. Assigning position of a single, large memory to contain these structures and mechanisms allows these structures and mechanisms to be reconfigured as required for most efficient operation of CS 10110. In an alternate embodiment, physically separate memories may be used to contain the structures and mechanisms depicted in MEM 10112, rather than assigned portions of a single memory.

b. Process Structure 10210 (Figs. 103, 104, 105)

Referring to Fig. 103, a partial schematic representation of Process Structures 10210 is shown. Specifically, Fig. 103 shows a Process (P) 10310 selected for execution, and its associated Procedure Objects (POs) in Process Objects (POs) 10213. P 10310 is represented in Fig. 103 as including four procedure objects in POs 10213. It is to be understood that this representation is for clarity of presentation; a particular P 10310 may include any number of procedure objects. Also for clarity of presentation, EURSM 10216 is not shown as EURSM 10216 is similar to FURSM 10214. EURSM 10216 will be described in detail in the following detailed discussions of CS 10110's structure and operation.

As previously discussed, each process includes certain data and procedure object. As represented in Fig. 103 for P 10310 the procedure objects reside in POs 10213. The data objects include Static Data Areas and stack mechanisms in P 10310. POs, for example KOS Procedure Object (KOSPO) 10318, contain the various procedures of the process, each procedure being a sequence of SInS defining an operation to be performed in executing the process. As will be described below, Procedure Objects also contain certain information used in executing the procedures contained therein. Static Data Areas (SDAs) are data objects generally reserved for storing data having an existence for the duration of the process. P 10310's stack mechanisms allow stacking of procedures for procedure calls and returns and for swapping processes in and out of JP 10114. Macro-Stacks (MAS) 10328 to 10334 are generally used to store automatic data (data generated during execution of a procedure and having an existence for the duration of that procedure). Although shown as separate from the stacks in P 10310, the SDAs may be contained with MASs 10328 to 10334. Secure Stack (SS) 10336 stores, in general, CS 10110 micro-machine state for each procedure called. Information stored in SS 10336 allows machine state to be recovered upon return from a called procedure, or when binding (swapping) a VP into CS 10110.

As shown in P 10310, each process is structured on a domain basis. A P 10310 may therefore include, for each domain, one or more procedure objects containing procedures having that domain as their domain of execution, an SDA and an MAS. For example, KOS domain of P 10310 includes KOSPO 10318, KOSSDA 10326, and KOSMAS 10334. P 10310's SS 10336 does not reside in any single domain of P 10310, but instead is a stack mechanism belonging to CS 10110 micromachine.

Having described the overall structure of a P 10310, the individual information structures and mechanisms of a P 10310 will next be described in greater detail.

1. Procedure Objects (Fig. 103)

KOSPO 10318 is typical of CS10110 procedure objects and will be referred to for illustration in the following discussion. Major components of KOSPO 10318 are Header 10338, External Entry Descriptor (EED) Area 10340, Internal Entry Descriptor (IED) Area 10342, S-Op Code Area 10344, Procedure Environment Descriptor (PED) 10348, Name Table (NT) 10350, and Access Information Array (AIA) Area 10352.

Header 10338 contains certain information identifying PO 10318 and indicating the number of entries in EED area 10340, discussed momentarily.

EED area 10340 and IED area 10342 together contain an Entry Descriptor (ED) for each procedure in KOSPO 10318. KOSPO 10318 is represented as containing Procedures 1, 2, and 11, of which Procedure 11 will be used as an example in the present discussion. EDs effectively comprise an index through certain all information in KOSPO 10318 can be located. IEDs form an index to all KOSPO 10318 procedures which may be called only from other procedures contained in KOSPO 10318. EEDs form an Index to all KOSPO 10318 procedures which may be called by procedures external to KOSPO 10318. Externally callable procedures are distinguished aid, as described in a following discussion of CS 10110's protection mechanisms, in

confirming external calling procedure's access rights.

Referring to ED 11, ED for procedure 11, three fields are shown therein. Procedure Environment Descriptor Offset (PEDO) field indicates the start, relative to start of KOSPO 10318, of Procedure 11's PED in PED Area 10348. As will be discussed further below, a procedure's PED contains a set of pointers for locating information used in the execution of that procedure. PED Area 10348 contains a PED for each procedure contained in 10318. In the present embodiment of CS 10110, a single PED may be shared by two or more procedures. Code Entry Point (CEP) field indicates the start, relative to Procedure Base Pointer (PBP) which will be discussed below, of Procedure 11's SIN Code and SIN Code Area 10344. Finally, ED 11's Initial Frame Size (IFS) field indicates the required Initial Frame Size of the KOSMAS 10334 frame storing Procedure 11's automatic data.

PED 11, Procedure 11's PED in PED Area 10348, contains a set of pointers for locating information used in execution of Procedure 11. The first entry in PED 11 is a header containing information identifying PED 11. PED 11's Procedure Base Pointer (PBP) entry is a pointer providing a fixed reference from which other information in PO 10318 may be located. In a specific example, Procedure 11's CEP indicates the location, relative to PBP, of the start of Procedure 11's S-Op code in S-Op Code Area 10344. As will be described further below, PBP is a CS 10110 Architectural Base Pointer (ABP). CS 10110's ABP's are a set of architectural pointers used in CS 10110 to facilitate addressing of CS 10110's address space. PED 11's Static Data Pointer (SDP) entry points to data, in PO 10318, specifying certain parameters of P 10310's KOSSDA 10326. Name Table Pointer (NTP) entry is a pointer indicating the location, in NT 10350, of Name Table Entry's (NTE's) for Procedure 11's operands. NT 10350 and NTE's will be described in greater detail in the following discussion of Computer System 10110's Addressing Structure. PED 11's S-Interpreter Pointer (SIP) entry is a pointer, discussed in greater detail in a following discussion of CS 10110's microcode structure, pointing to the particular S-Interpreter (SINT) to be used in interpreting Procedure 11's SIN Code.

Referring finally to AIA 10352, AIA 10352 contains, as previously discussed, information pertaining to access rights required of any external procedure calling a 10318 procedure. There is an AIA 10352 entry for each PO 10318 procedure which may be called by an external procedure. A particular AIA entry may be shared by one or more procedures having an ED in EED Area 10340. Each EED contains certain information, not shown for clarity of presentation, indicating that that procedure's corresponding AIA entry must be referred to, and the calling procedure's access rights confirmed, whenever that procedure is called.

2. Stack Mechanism (Figs. 104, 105)

As previously described, P10310's stack mechanisms include SS 10336, used in part for storing machine state, and MAS's 10328 to 10334, used to store local data generated during execution of P 10310's procedures. P 10310 is represented as containing an MAS for each CS 10110 domain. In an alternate embodiment of CS 10110, a particular P 10310 will include MAS's only for those domains in which that P 10310 is executing a procedure.

Referring to MAS's 10328 to 10334 and SS 10336, P 10310 is represented as having had eleven procedure calls. Procedure 0 has called Procedure 1, Procedure 1 has called Procedure 2, and so on. Each time a procedure is called, a corresponding stack frame is constructed on the MAS of the domain in which the called procedure is executed. For example, Procedures 1, 2, and 11 execute in KOS domain; MAS frames for Procedures 1, 2, and 11 therefore are placed on KOSMAS 10334. Similarly, Procedures 3 and 9 execute in EOS domain, so that their stack frames are placed on EOSMAS 10332. Procedures 5 and 6 execute in DBMS domain, so that their stack frames are placed on DBMSMAS 10330. Procedures 4, 7, 8, and 10 execute in User domain with their stack frames being placed on USERMAS 10328. Procedure 11 is the most recently called procedure and procedure 11's stack frame on KOSMAS 10334 is referred to as the current frame. Procedure 11 is the procedure which is currently being executed when VP 10310 is bound to CS 10110.

SS 10336, which is a stack mechanism of CS 10110 micromachine, contains a frame for each of Procedures 1 to 11. Each SS 10336 frame contains, in part, CS 10110 operating state for its corresponding procedure.

Referring to Fig. 104, a schematic representation of a typical MAS, for example KOSMAS 10334, is shown. KOSMAS 10334 includes Stack Header 10410 and a Frame 10412 for each procedure on KOSMAS 10334. Each Frame 10412 includes a Frame Header 10414, and may contain a Linkage Pointer Block 10416, a Local Pointer Block 10418, and a Local (Automatic) Data Block 10420.

KOSMAS 10334 Stack Header 10410 contains at least the following information:

- (1) an offset, relative to Stack Header 10410, indicating the location of Frame Header 10414 of the first frame on KOSMAS 10334;
- (2) a Stack Top Offset (STO) indicating location, relative to start of KOSMAS 10334, of the top of KOSMAS 10334; top of KOSMAS 10334 is indicated by pointer STO pointing to the top of the last entry of Procedure 11 Frame 10412's Local Data Block 10420;
- (3) an offset, relative to start of KOSMAS 10334, indicating location of Frame Header 10414 of the current top frame of KOSMAS 10334; in Fig. 104 this offset is represented by Frame Pointer (FP), an ABP discussed further below;
- (4) the VP 10310's UID;
- (5) a UID Pointer indicating location of certain domain environment information, described further in a

following discussion:

(6) a signaller pointer indicating the location of certain routines for handling certain CS 10110 operating system faults;

(7) a UID pointer indicating location of KOSSDA 10326; and

5 (8) a frame label sequencer containing pointers to headers of frames in other domains; these pointers are used in executing non-local go-to operations.

KOSMAS 10334 Stack Header 10410 thereby contains information for locating certain important points in KOSMAS 10334's structure, and for locating certain information pertinent to executing procedures in KOS domain.

10 Each Frame Header 10414 contains at least the following information:

(1) offsets, relative to the Frame Header 10414, indicating the locations of Frame Headers 10414 of the previous and next frames of KOSMAS 10334;

(2) an offset, relative to the Frame Header 10414, indicating the location of the top of that Frame 10412;

(3) information indicating the number of passed arguments contained in that Frame 10412;

15 (4) a dynamic back pointer, in UID/Offset format, to the previous Frame 10412 if that previous Frame 10412 resides in another domain;

(5) a UID/Offset pointer to the environmental descriptor of the procedure calling that procedure;

20 (6) a frame label sequence containing information indicating the locations of other Frame Headers 10414 in KOSMAS 10334; this information is used to locate other frames in KOSMAS 10334 for the purpose of executing local go-to operations. Frame Headers 10414 thereby contain information for locating certain important points in KOSMAS 10334 structure, and certain data pertinent to executing the associated procedures. In addition, Frame Headers 10414, in combination with Stack Header 10410, contain information for linking the activation records of each VP 10310 MAS, and for linking together the activation records of the individual MAS's.

25 Linkage Pointer Blocks 10416 contain pointers to arguments passed from a calling procedure to the called procedure. For example, Linkage Pointer Block 10416 of Procedure 11's Frame 10412 will contain pointers to arguments passed to Procedure 11 from Procedure 10. The use of linkage pointers in CS 10110's addressing structure will be discussed further in a following discussion of CS 10110's Addressing Structure. Local Data Pointer Blocks 10418 contain pointers to certain of the associated procedure's local data. Indicated in Fig. 104 is a pointer, Frame Pointer (FP), pointing between top most Frame 10412's Linkage Pointer Block 10416 and Local Data Pointer Block 10418. FP, described further in following discussions, is an ABP to MAS Frame 10412 of the process's current procedure.

Each Frame 10412's Local (Automatic) Data Block 10420 contains certain of the associated procedure's automatic data.

35 As described above, at each procedure call a MAS frame is constructed on top of the MAS of the domain in which the called procedure is executed. For example, when Procedure 10 calls Procedure 11 a Frame Header 10414 for Procedure 11 is constructed and placed on KOSMAS 10334. Procedure 11's linkage pointers are then generated, and placed in Procedure 11's Linkage Pointer Block 10416. Next Procedure 11's local pointers are generated and placed in Procedure 11's Local Pointer Block 10418. Finally, Procedure 11's local data is placed in Procedure 11's Local Data Block 10420. During this operation, USERMAS 10328's frame label sequence is updated to include an entry pointing to Procedure 11's Frame Header 10414. KOSMAS 10334's Stack Header 10410 is updated with respect to STO to the new top of KOSMAS 10334. Procedure 2's Frame Header 10414 is updated with respect to offset to Frame Header 10414 of Procedure 11 Frame 10412, and with respect to frame label sequence indicating location of Procedure 11's Frame Header 40 10414. As Procedure 11 is then the current procedure, FP is updated to a point between Linkage Pointer Block 10416 and Local Pointer Block 10418 of Procedure 11's Frame 10412. Also, as will be discussed below, a new frame is constructed on SS 10336 or Procedure 11. CS 10110 will then proceed to execute Procedure 11. During execution of Procedure 11, any further local data generated may be placed on the top of Procedure 11's Local Data Block 10420. The top of stack offset information in Procedure 11's Frame Header 45 10414 and in KOSMAS 10334 Stack Header 10410 will be updated accordingly.

MAS's 10328 to 10334 thereby provide a per domain stack mechanism for storing data pertaining to individual procedures, thus allowing stacking of procedures without loss of this data. Although structured on a domain basis, MAS's 10328 to 10334 comprise a unified logical stack structure threaded together through information stored in MAS stack and frame headers.

55 As described above and previously, SS 10336 is a CS 10110 micromachine stack structure for storing, in part, CS 10110 micromachine state for each stacked VP 10310 procedure. Referring to Fig. 105, a partial schematic representation of a SS 10336 Stack Frame 10510 is shown. SS 10336 Stack Header 10512 and Frame Headers 10514 contain information similar to that in MAS Stack Headers 10410 and Frame Headers 10414. Again, the information contained therein locates certain points within SS 10336 structure, and threads together SS 10336 with MAS's 10328 to 10334.

60 SS 10336 Stack Frame 10510 contains certain information used by the CS 10110 micromachine in executing the VP 10212 procedure with which this frame is associated. Procedure Pointer Block 10516 contains certain pointers including ABPs, used by CS 10110 micromachine in locating information within VP 10310's information structures. Micro-Routine Frames (MRFs) 10518 together comprise Micro-Routine Stack (MRS) 10520 within each SS 10336 Stack Frame 10510. MRS Stack 10520 is associated with the 65

EP 0 067 556 B1

internal operation of CS 10110 microroutines executed during execution of the VP 10212 procedure associated with the Stack Frame 10510. SS 10336 is thus a dual function CS 10110 micromachine stack. Pointer Block 10516 entries effectively define an interface between CS 10110 micromachine and the current procedure of the current process. MRS 10520 comprise a stack mechanism for the internal operations of CS 10110 micromachine.

Having briefly described Virtual Processes 10212, FURSM 10214 will be described next. As stated above, EURSM 10216 is similar in operation to FURSM 10214 and will be described in following detailed descriptions of CS 10110 structure and operation.

3. FURSM 10214 (Fig. 103)

Referring again to Fig. 103, FURSM 10214 includes CS 10110 micromachine information structures used internally to CS 10110 micromachine in executing the procedures of a P 10310. When a VP, for example P 10310, is to be executed, certain information regarding that VP is transferred from the Virtual Processes 10212 to FURSM 10214 for use in executing that procedure. In this respect, FURSM 10214 may be regarded as an acceleration mechanism for the current Virtual Process 10212.

FURSM 10214 includes General Register File (GRF) 10354, Micro Stack Pointer Register Mechanism (MISPR) 10356, and Return Control Word Stack (RCWS) 10358. GRF 10354 includes Global Registers (GRs) 10360 and Stack Registers (SRs) 10362. GRs 10360 include Architectural Base Registers (ABRs) 10364 and Micro-Control Registers (MCRs) 10366. Stack Registers 10362 include Micro-Stack (MIS) 10368 and Monitor Stack (MOS) 10370.

Referring first to GRF 10354, and assuming for example that Procedure 11 of P 10310 is currently being executed, GRF 10354 primarily contains certain pointers to P 10310 data used in execution of Procedure 11. As previously discussed, CS 10110's addressing structure includes certain Architectural Base Pointers (ABP's) for each procedure. ABPs provide a framework for accessing CS 10110's address space. The ABPs of each procedure include a Frame Pointer (FP), a Procedure Base Pointer (PBP), and a Static Data Pointer (STP). As discussed above with reference to KOSPO 10318, these ABPs reside in the procedure's PEDs. When a procedure is called, these ABP's are transferred from that procedure's PED to ABR's 10364 and reside therein for the duration of that procedure. As indicated in Fig. 103, FP points between Linkage Pointer Block 10416 and Local pointer Blocks 10418 of Procedure 11's Frame 10412 on KOSMAS 10334. PBP points to the reference point from which the elements of KOSPO 10318 are located. SDP points to KOSSDA 10326. If Procedure 11 calls, for example, a Procedure 12, Procedure 11's ABPs will be transferred onto Procedure Pointer Block 10516 of SS 10336 Stack Frame 10510 for Procedure 11. Upon return to Procedure 11, Procedure 11's ABPs will be transferred from Procedure Pointer Block 10516 to ABR's 10364 and execution of Procedure 11 resumed.

MCRs 10366 contain certain pointers used by CS 10110 micromachine in executing Procedure 11. CS 10110 micromachine pointers indicated in Fig. 103 include Program Counter (PC), Name Table Pointer (NTP), S-Interpreter Pointer (SIP), Secure Stack Pointer (SSP), and Secure Stack Top Offset (SSTO). NTP and SIP have been previously described with reference to KOSPO 10318 and reside in KOSPO 10318. NTP and SIP are transferred into MCR's 10366 at start of execution of Procedure 11. PC, as indicated in Fig. 103, is a pointer to the Procedure 11 SIN currently being executed by CS 10110. PC is initially generated from Procedure 11's PBP and CEP and is thereafter incremented by CS 10110 micromachine as Procedure 11's SIN sequences are executed. SSP and SSTO are, as described in a following discussion, generated from information contained in SS 10336's Stack Header 10512 and Frame Headers 10514. As indicated in Fig. 103 SSP points to start of SS 10336 while SSTO indicates the current top frame on SS 10336, whether Procedure Pointer Block 10516 or a MRF 10518 of MRS 10520, by indicating an offset relative to SSP. If Procedure 11 calls a subsequent procedure, the contents of MCR's 10366 are transferred into Procedure 11's Procedure Pointer Block 10516 on SS 10336, and are returned to MCR's 10366 upon return to Procedure 11.

Registers 10360 contain further pointers, described in following detailed discussions of CS 10110 operation, and certain registers which may be used to contain the current procedure's local data.

Referring now to Stack Registers 10362, MIS 10368 is an upward extension, or acceleration, of MRS 10520 of the current procedure. As previously stated, MRS 10520 is used by CS 10110 micromachine in executing certain microroutines during execution of a particular procedure. MIS 10368 enhances the efficiency of CS 10110 micromachine in executing these microroutines by accelerating certain most recent MRFs 10518 of that procedure's MRS 10520 into FU 10120. MIS 10368 may contain, for example, up to the eight most recent MRFs 10518 of the current procedures MRS 10520. As various microroutines are called or returned from, MRS 10520 MRF's 10518 are transferred accordingly between SS 10336 and MIS 10368 so that MIS 10368 always contains at least the top MRF 10518 of MRS 10520, and at most eight MRFs 10518 of MRS 10520. MISPR 10356 is a CS 10110 micromachine mechanism for maintaining MIS 10368. MISPR 10356 contains a Current Pointer, a Previous Pointer, and a Bottom Pointer. Current Pointer points to the top-most MRF 10518 on MIS 10368. Previous Pointer points to the previous MRF 10518 on MIS 10368, and Bottom Pointer points to the bottom-most MRF 10518 on MIS 10368. MISPR 10356's Current, Previous and Bottom Pointers are updated as MRFs 10518 are transferred between SS 10336 and MIS 10368. If Procedure 11 calls a subsequent procedure, all Procedure 11 MRFs 10518 are transferred from MIS 10368 to Procedure 11's MRS 10520 on SS 10336. Upon return to Procedure 11, up to seven of Procedure 11's MRFs 10518

frames are returned from SS 10336 to MIS 10368.

Referring to MOS 10370, MOS 10370 is a stack mechanism used by CS 10110 micromachine for certain microroutines for handling fault or error conditions. These microroutines always run to completion, so that MOS 10370 resides entirely in FM 10120 and is not an extension of a stack residing in a P 10310 in MEM 10112. MOS 10370 may contain, for example, eight frames. If more than eight successive fault or error conditions occur, this is regarded as a major failure of CS 10110. Control of CS 10110 may then be transferred to DP 10118. As will be described in a following discussion, diagnostic programs in DP 10118 may then be used to diagnose and locate the CS 10110 faults or errors. In other embodiments of CS 10110 MOS 10370 may contain more or fewer stack frames, depending upon the degree of self diagnosis and correction capability desired for CS 10110.

RCWS 10358 is a two-part stack mechanism. A first part operates in parallel with MIS 10368 and a second part operates in parallel with MOS 10370. As previously described, CS 10110 is a microcode controlled system. RCWS is a stack for storing the current microinstruction being executed by CS 10110 micromachine when the current procedure is interrupted by a fault or error condition, or when a subsequent procedure is called. That portion of RCWS 10358 associated with MIS 10368 contains an entry for each MRF 10518 residing in MIS 10368. These RCWS 10358 entries are transferred between SS 10336 and MIS 10368 in parallel with their associated MRFs 10518. When resident in SS 10336, these RCWS 10358 entries are stored within their associated MRFs 10518. That portion of RCWS 10358 associated with MOS 10370 similarly operates in parallel with MOS 10370 and, like MOS 10370, is not an extension of an MEM 10112 resident stack.

In summary, each process active in CS 10110 exists as a separate, complete, and self-contained entity, or Virtual Process, and is structurally organized on a domain basis. Each Virtual Process includes, besides procedure and data objects, a set of MAS's for storing local data of that processes procedures. Each Virtual Process also includes a CS 10110 micromachine stack, SS 10336, for storing CS 10110 micromachine state pertaining to each stacked procedure of the Virtual Process. CS 10110 micromachine includes a set of information structures, register 10360, MIS 10368, MOS 10370, and RCWS 10358, used by CS 10110 micromachine in executing the Virtual Process's procedures. Certain of these CS 10110 micromachine information structures are shared with the currently executing Virtual Process, and thus are effectively acceleration mechanisms for the current Virtual Process, while others are completely internal to CS 10110 micromachine.

A primary feature of CS 10110 is that each process' macrostacks and secure stack resides in MEM 10112. CS 10110's macrostack and secure stacks are therefore effectively unlimited in depth.

Yet another feature of CS 10110 micromachine is the use of GRF 10354. GRF 10354 is, in an embodiment of CS 10110, a unitary register array containing for example, 256 registers. Certain portions, or address locations, of GRF 10354 are dedicated to, respectively, GRs 10360, MIS 10368, and MOS 10370. The capacities of GR 10360, MIS 10368, and MOS 10370, may therefore be adjusted, as required for optimum CS 10110 efficiency, by reassignment of GRF 10354's address space. In other embodiments of CS 10110, GRs 10360, MIS 10368, and MOS 10370 may be implemented a functionally separate registers arrays.

Having briefly described the structure and operation of Process Structures 10210, VP State Block 10218 will be described next below.

C. Virtual Processor State Blocks and Virtual Process Creation (Fig. 102)

Referring again to Fig. 102, VP State Blocks 10218 is used in management and control of processes. VP State Blocks 10218 contains a VP State Block for each Virtual Process (VP) selected for execution by CS 10110. Each such VP State Block contains at least the following information: (1) the state, or identification number of a VP;

- (2) entries identifying the particular principle and particular process of the VP;
- (3) an AON pointer to that VP's secure stack (e.g., SS 10336);
- (4) the AON's of that VP's MAS stack objects (e.g., MAS's 10328 to 10334); and,
- (5) certain information used by CS 10110's VP Management System.

The information contained in each VP State Block thereby defines the current state of the associated VP.

A Process is loaded into CS 10110 by building a primitive access record and loading this access record into CS 10110 to appear as an already existing VP. A VP is created by creating a Process Object, including pointers to macro- and secure-stack objects created for that VP, micromachine state entries, and a pointer to the user's program. CS 10110's KOS then generates Macro- and Secure-Stack Objects with headers for that process and, as described further below, loads protection information regarding that process' objects into protection Structures 10230. CS 10110's KOS then copies this primitive machine state record into a vacant VPSB selected by CS 10110's VP Manager, thus binding the newly created VP into CS 10110. At that time a KOS initializer procedure completes creation of the VP for example by calling in the user's program through a compiler. The newly created VP may then be executed by CS 10110.

Having briefly described VP State Blocks 10218 and creation of a VP, CS 10110's Addressing Structures 10220 will be described next below.

D. Addressing Structure 10220 (Figs. 103, 106, 107, 108)

1. Objects, UID's, AON's, Names, and Physical Addresses (Fig. 106)

As previously described, the data space accessible to CS 10110 is divided into segments, or containers, referred to as objects. In an embodiment of CS 10110, the addressable data space of each object has a capacity of 2^{32} bits of information and is structured into 2^{18} pages with each page containing 2^{14} bits of information.

Referring to Fig. 106A, a schematic representation of CS 10110's addressing structure is shown. Each object created for use in, or by operation of, a CS 10110 is permanently assigned a unique identifier (UID). An object's UID allows an object to be uniquely identified and located at any future point in time. Each UID is an 80 bit number, so that the total addressable space of all CS 10110's includes 2^{80} objects wherein each object may contain up to 2^{32} bits of information. As indicated in Fig. 106, each 80 bit UID is comprised of 32 bits of Logical Allocation Unit Identifier (LAUID) and 48 bits of Object Serial Number (OSN). LAUIDs are associated with individual CS 10110 systems. LAUIDs identify the particular CS 10110 system generating a particular object. Each LAUID is comprised of a Logical Allocation Unit Group Number (LAUGN) and a Logical Allocation Unit Serial Number (LAUSN). LAUGNs are assigned to individual CS 10110 systems and may be guaranteed to be unique to a particular system. A particular system may, however, be assigned more than one LAUGN so that there may be a time varying mapping between LAUGNs and CS 10110 systems. LAUSNs are assigned within a particular system and, while LAUSNs may be unique within a particular system, LAUSNs need not be unique between systems and need not map onto the physical structure of a particular system.

OSNs are associated with individual objects created by an LAU and are generated by an Architectural Clock in each CS 10110. Architectural clock is defined as a 64 bit binary number representing increasing time. Least significant bit of architectural clock represents increments of 600 picoseconds, and most significant bit represents increments of 127 years. In the present embodiment of CS 10110, certain most significant and least significant bits of architectural clock time are disregarded as generally not required practice. Time indicated by architectural clock is measured relative to an arbitrary, fixed point in time. This point in time is the same for all CS 10110s which will ever be constructed. All CS 10110s in existence will therefore indicate the same architectural clock time and all UIDs generated will have a common basis. The use of an architectural clock for generation of OSNs is advantageous in that it avoids the possibility of accidental duplication of OSNs if a CS 10110 fails and is subsequently reinitiated.

As stated above, each object generated by or for use in a CS 10110 is uniquely identified by its associated UID. By appending Offset (O) and Length (L) information to an object's UID, a UID logical address is generated which may be used to locate particular segments of data residing in a particular object. As indicated in Fig. 106, O and L fields of a UID logical address are each 32 bits. O and L fields can therefore indicate any particular bit, out of 2^{32-1} bits, in an object and thus allow bit granular addressing of information in objects.

As indicated in Fig. 106 and as previously described, each object active in CS 10110 is assigned a short temporary unique identifier valid only within JP 10114 and referred to as an Active Object Number (AON). Because fewer objects may be active in a CS 10110 than may exist in a CS 10110's address space, AON's are, in the present embodiment of CS 10110, 14 bits in length. A particular CS 10110 may therefore contain up to 2^{14} active objects. An object's AON is used within JP 10114 in place of that object's UID. For example, as discussed above with reference to process structures 10210, a procedure's FP points to start of that procedure's frame on its process' MAS. When that FP is residing in SS 10336, it is expressed as a UID. When that procedure is to be executed, FP is transferred from SS 10336 to ABR's 10364 and is translated into the corresponding AON. Similarly, when that procedure is stacked, FP is returned to SS 10336 and in doing so is translated into the corresponding UID. Again, a particular data segment in an object may be addressed by means of an AON logical address comprising the object's AON plus associated 32 bit Offset (O) and Length (L) fields.

Each operand appearing in a process is assigned a Name and all references to a process's operands are through those assigned Names. As indicated in Fig. 106B, in the present embodiment of CS 10110 each Name is an 8, 12, or 16 bit number. All Names within a particular process will be of the same length. As will be described in a following discussion, Names appearing during execution of a process may be resolved, through a procedure's Name Table 10350 or through Name Cache 10226, to an AON logical address. As described below, an AON logical address corresponding to an operand Name may then be evaluated to a MEM 10112 physical address to locate the operand referred to.

The evaluation of AON logical addresses to MEM 10112 physical addresses is represented in Fig. 106C. An AON logical address's L field is not involved in evaluation of an AON logical address to a physical address and, for purposes of clarity of presentation, is therefore not represented in Fig. 106C. AON logical address L field is to be understood to be appended to the addresses represented in the various steps of the evaluation procedure shown in Fig. 106C.

As described above, objects are 2^{32} bits structured into 2^{18} pages with each page containing a 2^{14} bits of data. MEM 10112 is similarly physically structured into frames with, in the present embodiment of CS 10110, each frame containing 2^{14} bits of data. In other embodiments of CS 10110, both pages and frames may be of different sizes but the translation of AON logical addresses to MEM 10112 physical addresses will be similar to that described momentarily.

An AON logical address O field was previously described as a 32 bit number representing the start, relative to start of the object, of the addressed data segment within the object. The 18 most significant bits of O field represent the number (P) of the page within the object upon which the first bit of the addressed data occurs. The 14 least significant bits of O field represent the offset (O_p), relative to the start of the page, within that page of the first bit of the addressed data. AON logical address O field may therefore, as indicated in Fig. 106C, be divided into an 18 bit page (P) field and a 14 bit offset within page (O_p) field. Since, as described above, MEM 10112 physical frame size is equal to object page size, AON logical address O_p field may be used directly as an offset within frame (O_f) field of the physical address. As will be described below, an AON logical address AON and P fields may then be related to the frame number (FN) of the MEM 10112 frame in which that page resides, through Addressing Mechanisms 10220.

Having briefly described the relationships between UIDs, UID Logical Addresses, Names, AONs, AON Logical Addresses, and MEM 10112 Physical Addresses, Addressing Mechanisms 10220 will be described next below.

2. Addressing Mechanisms 10220 (Fig. 107)

Referring to Fig. 107, a schematic representation of Computer System 10110's Addressing Mechanisms 10220 is shown. As previously described, Addressing Mechanisms 10220 comprise UID/AON Tables 10222, Memory Management Tables 10224, Name Cache 10226, and Address Translation Unit 10228.

UID/AON Tables 10222 relate each object's UID to its assigned AON and include AOT Hash Table (AOTHT) 10710, Active Object Table (AOT) 10712, and Active Object Table Annex (AOTA) 10714.

An AON corresponding to a particular UID is determined through AOTHT 10710. The UID is hashed to provide a UID index into AOTHT 10710, which then provides the corresponding AON. AOTHT 10710 is effectively an acceleration mechanism of AOT 10712 to, as just described, provide rapid translation of UIDs to AONs. AONs are used as indexes into AOT 10712, which provides a corresponding AOT Entry (AOTE). An AOTE as described in following detailed discussions of CS 10110, includes, among other information, the UID corresponding to the AON indexing the AOTE. In addition to providing translation between AONs and UIDs, the UID of an AOTE may be compared to an original UID to determine the correctness of an AON from AOTHT 10710.

Associated with AOT 10712 is AOTA 10714. AOTA 10714 is an extension of AOT 10712 and contains certain information pertaining to active objects, for example the domain of execution of each active procedure object.

Having briefly described CS 10110's mechanism for relating UIDs and AONs, CS 10110's mechanism for resolving operand Names to AON logical addresses will be described next below.

3. Name Resolution (Figs. 103, 108)

Referring first to Fig. 103, each procedure object in a VP, for example KOSPO 10318 in VP 10310, was described as containing a Name Table (NT) 10350. Each NT 10350 contains a Name Table Entry (NTE) for each operand whose Name appears its procedure. Each NTE contains a description of how to resolve the corresponding Name to an AON Logical Address, including fetch mode information, type of data referred to by that Name, and length of the data segment referred to.

Referring to Fig. 108, a representation of an NTE is shown. As indicated, this NTE contains seven information fields: Flag, Base (B), Predisplacement (PR), Length (L), Displacement (D), Index (I), and Inter-element Spacing (IES). Flag Field, in part, contains information describing how the remaining fields of the NTE are to be interpreted, type of information referred to by the NTE, and how that information is to be handled when fetched from MEM 10112. L Field, as previously described, indicates length, or number of bits in, the data segment. Functions of the other NTE fields will be described during the following discussions.

In a present embodiment of CS 10110, there are five types of NTE: (1) base (B) is not a Name, address resolution is not indirect; (2) B is not a Name, address resolution is indirect; (3) B is a Name, address resolution is indirect; (4) B is a Name, address resolution is indirect. A fifth type is an NTE selecting a particular element from an array of elements. These five types of NTE and their resolution will be described below, in the order mentioned.

In the first type, B is not a Name and address resolution is not indirect, B Field specifies an ABR 10364 containing an AON plus offset (AON/O) Pointer. The contents of D Field are added to the O Field of this pointer, and the result is the AON Logical Address of the operand. In the second type, B is not a Name and address resolution is indirect, B Field again specifies an ABR 10364 containing an AON/O pointer. The contents of PR Field are added to the O Field of the AON/O pointer to provide an AON Logical Address of a Base Pointer. The Base Pointer AON Logical Address is evaluated, as described below, and the Base Pointer fetched from MEM 10112. The contents of D Field are added to the O Field of the Base Pointer and the result is the AON Logical Address of the operand.

NTE types 3 and 4 correspond, respectively to NTE types 1 and 2 and are resolved in the same manner except that B Field contains a Name. The B Field Name is resolved through another NTE to obtain an AON/O pointer which is used in place of the ABR 10364 pointers referred to in discussion of types 1 and 2.

The fifth type of NTE is used in references to elements of an array. These array NTEs are resolved in the

same manner as NTE types 1 through 4 above to provide an AON Logical Address of the start of the array. I and IES Fields provide additional information to locate a particular element in the array. I Field is always Name which is resolved to obtain an operand value representing the particular element in the array. IES Field provides information regarding spacing between elements of the array, that is the number of bits between adjacent element of the array. IES Field may contain the actual IES value, or it may contain a Name which is resolved to an AON Logical Address leading to the inter-element spacing value. The I and IES values, obtained by resolving the I and IES Fields as just described, are multiplied together to determine the offset, relative to the start of the array, of the particular element referred to by the NTE. This within array offset is added to the O Field of the AON Logical Address of the start of the array to provide the AON Logical Address of the element.

In the current embodiment of CS 10110, certain NTE fields, for example B, D, and Flag fields, always contain literals. Certain other fields, for example, IES, D, PRE, and L fields, may contain either literals or names to be resolved. Yet other fields, for example I field, always contain names which must be resolved.

Passing of arguments from a calling procedure to a called procedure has been previously discussed with reference to Virtual Processes 10212 above, and more specifically with regard to MAS's 10328 to 10334 of VP 10310. Passing of arguments is accomplished through the calling and called procedure's Name Tables 10350. In illustration, a procedure W(a, b, c) may wish to pass arguments a, b, and c to procedure X(u, v, w), where arguments, v and w correspond to arguments a, b, and c. At compilation, NTEs are generated for arguments a, b, and c in procedure W's procedure object, and NTEs are generated for arguments u, v and w in Procedure X's procedure object. Procedure X's NTEs for u, v, and w are constructed to resolve to point to pointers in Linkage Pointer Block 10416 of Procedure X's Frame 10412 in MAS. To pass arguments a, b, and c from Procedure W to Procedure X, the NTEs of arguments a, b, and c are resolved to AON Logical Addresses (i.e., AON/O form). Arguments a, b, and c's AON Logical Addresses are then translated to corresponding UID addresses which are placed in Procedure X's Linkage Pointer Block 10416 at those places pointed to by Procedure X's NTEs for u, v, and w. When Procedure X is executed, the resolution of Procedure X's NTEs for u, v, and w will be resolved to locate the pointers, in Procedure X's Linkage Pointer Block 10416 to arguments a, b, and c. When arguments are passed in this manner, the data type and length information are obtained from the called procedure's NTEs, rather than the calling procedure's NTEs. This allows the calling procedure to pass only a portion of, for example, arguments a, b, or c, to the called procedure and thus may be regarded as a feature of CS 10110's protection mechanisms.

Having briefly described resolution of Names to AON/Offset addresses, and having previously described translation of UID addresses to AON addresses, the evaluation of AON addresses to MEM 10112 physical addresses will be described next below.

4. Evaluation of AON Addresses to Physical Addresses (Fig. 107)

Referring again to Fig. 107, a partial schematic representation of CS 10110's Memory Management Table 10224 is shown. Memory Hash Table (MHT) 10716 and Memory Frame Table (MFT) 10718 are concerned with translation of AON addresses into MEM 10112 physical addresses and will be discussed first. Working Set Matrix (WSM) 10720 and Virtual Memory Manager Request Queue (VMMRQ) 10722 are concerned with management of MEM 10112's available physical address base and will be discussed second. Active Object Request Queue (AORQ) 10728 and Logical Allocation Unit Directory (LAUD) 10730 are concerned with locating inactive objects and management of which objects are active in CS 10110 and will be discussed last.

Translation of AON/O Logical Addresses to MEM 10112 physical addresses was previously discussed with reference to Fig. 106C. As stated in that discussion, objects are divided into pages. Correspondingly, the AON/O Logical Address' O Field is divided into an 18 bit page number (P) Field and a 14 bit offset within a page (O_p) Field. MEM 10112 is structured into frames, each of which in the present embodiment of CS 10110 is equal to a page of an object. An AON/O address' O_p Field may therefore be used directly as an offset within frame (O_f) of the corresponding physical address. The AON and P fields of an AON address must, however, be translated into a MEM 10112 frame represented by a corresponding Frame Number (FN).

Referring now to Fig. 107, an AON address' AON and P Fields are "hashed" to generate an MHT index which is used as an index into MHT 10716. Briefly, "hashing" is a method of indexing, or locating, information in a table wherein indexes to the information are generated from the information itself through a "hashing function". A hashing function maps each piece of information to the corresponding index generated from it through the hashing function. MHT 10716 then provides the corresponding FN of the MEM 10112 frame in which that page is stored. FNs are used as indexes into MFT 10718, which contains, for each FN, an entry describing the page stored in that frame. This information includes the AON and P of the page stored in that MEM 10112 frame. An FN from MHT 10716 may therefore be used as an index into MFT 10718 and the resulting AON/P of MFT 10718 compared to the original AON/P to confirm the correctness of the FN obtained from MHT 10716. MHT 10716 is an effectively acceleration mechanism of MFT 10718 to provide rapid translation of AON address to MEM 10112 physical addresses.

MFT 10718 also stores "used" and "modified" information for each page in MEM 10112. This information indicates which page frames stored therein have been used and which have been modified.

This information is used by CS 10110 in determining which frames may be deleted from MEM 10112, or are free, when pages are to be written into MEM 10112 from backing store (ED 10124). For example, if a page's modified bit indicates that that page has not been written into, it is not necessary to write that page back into backing store when it is deleted from MEM 10112; instead, that page may be simply erased.

5 Referring finally to ATU 10228, ATU 10228 is an acceleration mechanism for MHT 10716. AON/O addresses are used directly, without hashing, as indexes into ATU 10228 and ATU 10228 correctly provides corresponding FN and O outputs. A CS 10110 mechanism, described in a following detailed discussion of CS 10110 operation, continually updates the contents of ATU 10228 so that ATU 10228 contain the FN's and O_p (O_p) of the pages most frequently referenced by the current process. If ATU 10228 does not contain a
10 corresponding entry for a given AON input, an ATU fault occurs and the FN and O information may be obtained directly from MHT 10716.

Referring now to WSM 10720 and VMMRQ 10722, as previously stated these mechanisms are concerned with the management of MEM 10112's available address space. For example, if MHT 10716 and MFT 10718 do not contain an entry for a page referenced by the current procedure, an MHT/MFT fault
15 occurs and the reference page must be fetched from backing store (ED 10124) and read into MEM 10112. WSM 10720 contains an entry for each page resident in MEM 10112. These entries are accessed by indexes comprising the Virtual Processor Number (VPN) of the virtual process making a page reference and the P of the page being referenced. Each WSM 10720 entry contains 2 bits stating whether the particular page is part of a VP's working set, that is, used by that VP, and whether that page has been referenced by that VP.
20 This information, together with the information contained in that MFT 10718 entries described above, is used by CS 10110's Virtual Memory Manager (VMM) in transferring pages into and out of MEM 10112.

CS 10110's VMM maintains VMMRQ 10722, which is used by VMM to control transfer of pages into and out of MEM 10112. VMMRQ 10722 includes Virtual Memory Request Counter (VMRC) 10724 and a Queue of Virtual Memory Request Entries (VMREs) 10726. As will be discussed momentarily, VMRC 10724 tracks the
25 number of currently outstanding request for pages. Each VMRE 10726 describes a particular page which has been requested. Upon occurrence of a MHT/MFT (or page) fault, VMRC 10724 is incremented, which initiates operation of CS 10110's VMM, and a VMRE 10726 is placed in the queue. Each VMRE 10726 comprises the VPN of the process requesting the page and the AON/O of the page requested. At this time, the VP making the request is swapped out of JP 10114 and another VP bound to JP 10114. VMM allocates
30 MEM 10112 frame to contain the requested page, using the previously described information in MFT 10718 and WSM 10720 to select this frame. In doing so, VMM may discard a page currently resident in MEM 10112 for example, on the basis of being the oldest page, an unused page, or an unmodified page which does not have to be written back into backing store. VMM then requests an I/O operation to transfer the requested page into the frame selected by the VMM. While the I/O operation is proceeding, VMM generates new entries in MHT 10716 and MFT 10718 for the requested page, cleans the frame in MEM 10112 which is to be
35 occupied by that page, and suspends operation. IOS 10116 will proceed to execute the I/O operation and writes the requested page directly into MEM 10112 in the frame specified by VMM. IOS 10116 then notifies CS 10110's VMM that the page now resides in memory and can be referenced. At some later time, that VP requesting that page will resume execution and repeat that reference. Going first to ATU 10228, that VP will
40 take an ATU 10228 fault since VP 10212 has not yet been updated to contain that page. The VP will then go to MHT 10716 and MFT 10718 for the required information and, concurrently, WSM 10720 and ATU 10228 will be updated.

In regard to the above operations, each VP active in CS 10110 is assigned a Page Fault Frequency Time Factor (PFFT) which is used by CS 10110's VMM to adjust that VP's working set so that the interval between
45 successive page faults for that VP lies in an optimum time range. This assists in ensuring CS 10110's VMM is operating most efficiently and allows CS 10110's VMM to be tuned as required.

The above discussions have assumed that the page being referenced, whether from a UID/O address, an AON/O address, or a Name, is resident in an object active in CS 10110. While an object need not have a page in MEM 10112 to be active, the object must be active to have a page in MEM 10112. A VP, however,
50 may reference a page in an object not active in CS 10110. If such a reference is made, the object must be made active in CS 10110 before the page can be brought into MEM 10112. The result is an operation similar to the page fault operation described above. CS 10110 maintains an Active Object Manager (AOM), including Active Object Request Queue (AORQ) 10728, which are similar in operation to CS 10110's VMM and VMMRQ 10722. CS 10110's AOM and AORQ 10728 operate in conjunction with AOTHT 10710 and AOT
55 10712 to locate inactive objects and make them active by assigning them AON's and generating entries for them in AOTHT 10710, AOT 10712, and AOTA 10714.

Before a particular object can be made active in CS 10110, it must first be located in backing store (ED 10124). All objects on backing store are located through a Logical Allocation Unit Directory (LAUD) 10730, which is resident in backing store. An LAUD 10730 contains entries for each object accessible to the
60 particular CS 10110. Each LAUD 10730 entry contains the information necessary to generate an AOT 10712 entry for that object. An LAUD 10730 is accessed through a UID/O address contained in CS 10110's VMM. A reference to an LAUD 10730 results in MEM 10112 frames being assigned to that LAUD 10730, and LAUD 10730 being transferred into MEM 10112. If an LAUD 10730 entry exists for the referenced inactive object, the LAUD 10730 entry is transferred into AOT 10712. At the next reference to a page in that object, AOT
65 10712 will provide the AON for that object but, because the page has not yet been transferred into MEM

10112, a page fault will occur. This page fault will be handled in the manner described above and the referenced page transferred into MEM 10112.

5 Having briefly described the structure and operation of CS 10110's Addressing Structure, including the relationship between UIDs, Names, AONs, and Physical Addresses and the mechanisms by which CS 10110 manages the available address space of MEM 10112, CS 10110's protection structures will be described next below.

E. CS 10110 Protection Mechanisms (Fig. 109)

10 Referring to Fig. 109, a schematic representation of Protection Mechanisms 10230 is shown. Protection Tables 10232 include Active Primitive Access Matrix (APAM) 10910, Active Subject Number Hash Table (ASNHT) 10912, and Active Subject Table (AST) 10914. Those portions of Protection Mechanism 10230 resident in FU 10120 include ASN Register 10916 and Protection Cache (PC) 10234.

15 As previously discussed, access rights to objects are arbitrated on the basis of subjects. A subject has been defined as a particular combination of a principle, Process, and Domain (PPD), each of which is identified by a corresponding UID. Each object has associated with it an Access Control List (ACL) 10918 containing an ACL Entry (ACLE) for each subject having access rights to that object.

20 When an object becomes active in CS 10110 (i.e., is assigned an AON) each ACLE in that object's ACL 10918 is written into APAM 10910. Concurrently, each subject having access rights to that object, and for which there is an ACLE in that object's ACL 10918, is assigned an Active Subject Number (ASN). These ASNs are written into ASNHT 10912 and their corresponding PPDs are written into AST 10914. Subsequently, the ASN of any subject requesting access to that object is obtained by hashing the PPD of that subject to obtain a PPD index into ASNHT 10912. ASNHT 10912 will in turn provide a corresponding ASN. An ASN may be used as an index into AST 10914. AST 10914 will provide the corresponding PPD, which may be compared to an original PPD to confirm the accuracy of the ASN.

25 As described above, APAM 10910 contains an ACL 10918 for each object active in CS 10110. The access rights of any particular active subject to a particular active object are determined by using that subject's ASN and that object's AON as indexes into APAM 10910. APAM 10910 in turn provides a 4 bit output defining whether that subject has Read (R) Write (W) or Execute (E) rights with respect to that object, and whether that particular entry is Valid (V).

30 ASN Register 10916 and PC 10234 are effectively acceleration mechanisms of Protection Tables 10232. ASN Register 10916 stores the ASN of a currently active subject while PC 10234 stores certain access right information for objects being used by the current process. PC 10234 entries are indexed by ASNs from ASN register 10916 and by a mode input from JP 10114. Mode input defines whether the current procedure intends to read, write, or execute with respect to a particular object having an entry in PC 10234. Upon receiving ASN and mode inputs, PC 10234 provides a go/nogo output indicating whether that subject has the access rights required to execute the intended operation with respect to that object.

35 In addition to the above mechanism, each procedure to which arguments may be passed in a cross-domain call has associated with it an Access Information Array (AIA) 10352, as discussed with reference to Virtual Processes 10212. A procedure's AIA 10352 states what access rights a calling procedure (subject) must have to a particular object (argument) before the called procedure can operate on the passed argument. CS 10110's protection mechanisms compare the calling procedure's access rights to the rights required by the called procedure. This insures the calling procedure may not ask a called procedure to do what the calling procedure is not allowed to do. Effectively, a calling procedure can pass to a called procedure only the access rights held by the calling procedure.

40 Finally, PC 10234, APAM 10910, or AST 10914 faults (i.e., misses) are handled in the same manner as described above with reference to page faults in discussion of CS 10110's Addressing Mechanisms 10220. As such, the handling of protection misses will not be discussed further at this point.

45 Having briefly described structure and operation of CS 10110's Protection Mechanisms 10230, CS 10110's Micro-Instruction Mechanisms 10236 will be described next below.

50 F. CS 10110 Micro-Instruction Mechanism (Fig. 110)

As previously described, CS 10110 is a multiple language machine. Each program written in a high level user language is compiled into a corresponding S-Language program containing instructions expressed as S-Instructions (SINs). CS 10110 provides a set, or dialect, of microcode instructions, referred to as S-Interpreters (SINTs) for each S-Language. SINTs interpret SINs and provide corresponding sequences of microinstructions for detailed control of CS 10110.

55 Referring to Fig. 110, a partial schematic representation of CS 10110's Micro-Instruction Mechanisms 10236 is shown. At system initialization all CS 10110 microcode, including SINTs and all machine assist microcode, is transferred from backing store to Micro-Code Control Store (mCCS) 10238 in MEM 10112. The Micro-Code is then transferred from mCCS 10238 to FU Micro-Code Structure (FUmC) 10240 and EU Micro-Code Structure (EUmC) 10242. EUmC 10242 is similar in structure and operation to FUmC 10240 and thus will be described in following detailed descriptions of CS 10110's structure and operation. Similarly, CS 10110 machine assist microcode will be described in following detailed discussions. The present discussion will concern CS 10110's S-Interpreter mechanisms.

60 CS 10110's S-Interpreters (SINTs) are loaded into S-Interpreter Table (SITT) 11012, which is represented

in Fig. 110 as containing S-Interpreters 1 to N. Each SIT contains one or more sequences of micro-code; each sequence of microcode corresponds to a particular SIN in that S-Language dialect. S-Interpreter Dispatch Table (SDT) 11010 contains S-Interpreter Dispatchers (SDs) 1 to N. There is one SD for each SINT in SITT 11012, and thus a SD for each S-Language dialect. Each SD comprises a set of pointers. Each pointer
 5 in a particular SD corresponds to a particular SIN of that SD's dialect and points to the corresponding sequence of microinstructions for interpreting that SIN in that dialect's SIT in SITT 11012. In illustration, as previously discussed when a particular procedure is being executed the SIP for that procedure is transferred into one of mCR's 10366. That SIP points to the start of the SD for the SIT which is to be used to interpret the SINs of that procedure. In Fig. 110, the SIP in mCRs 10366 is shown as pointing to the start of
 10 SD2. Each S-Op appearing during execution of that procedure is an offset, relative to the start of the selected SD, pointing to a corresponding SD pointer. That SD pointer in turn points to the corresponding sequence of microinstructions for interpreting that SIN in the corresponding SIT in SITT 11012. As will be described in following discussions, once the start of a microcode sequence for interpreting an SIN has been selected, CS 10110 micromachine then proceeds to sequentially call the microinstructions of that sequence
 15 from SITT 11012 and use those microinstructions to control operation of CS 10110.

G. Summary of Certain CS 10110 Features and Alternate Embodiments

The above Introductory Overview has described the overall structure and operation and certain features of CS 101, that is, CS 10110. The above Introduction has further described the structure and
 20 operation and further features of CS 10110 and, in particular, the physical implementation and operation of CS 10110's information, control, and addressing mechanisms. Certain of these CS 10110 features are summarized next below to briefly state the basic concepts of these features as implemented in CS 10110. In addition, possible alternate embodiments of certain of these concepts are described.

First, CS 10110 is comprised of a plurality of independently operating processors, each processor
 25 having a separate microinstruction control. In the present embodiment of CS 10110, these processors include FU 10120, EU 10122, MEM 10112 and IOS 10116. Other such independently operating processors, for example, special arithmetic processors such as an array processor, or multiple FU 10120's, may be added to the present CS 10110.

In this regard, MEM 10112 is a multiport processor having one or more separate and independent ports
 30 to each processor in CS 10110. All communications between CS 10110's processors are through MEM 10112, so that MEM 10112 operates as the central communications node of CS 10110, as well as performing memory operations. Further separate and independent ports may be added to MEM 10112 as further processors are added to CS 10110. CS 10110 may therefore be described as comprised of a plurality of separate, independent processors, each having a separate microinstruction control and having a separate
 35 and independent port to a central communications and memory node which in itself is an independent processor having a separate and independent microinstruction control. As will be further described in a following detailed description of MEM 10112, MEM 10112 itself is comprised of a plurality of independently operating processors, each performing memory related operations and each having a separate microinstruction control. Coordination of operations between CS 10110's processors is achieved by
 40 passing "messages" between the processors, for example, SOP's and descriptors.

CS 10110's addressing mechanisms are based, first, upon UID addressing of objects. That is, all information generated for use in or by operation of a CS 10110, for example, data and procedures, is structured into objects and each object is assigned a permanent UID. Each UID is unique within a particular
 45 CS 10110 and between all CS 10110's and is permanently associated with a particular object. The use of UID addressing provides a permanent, unique addressing means which is common to all CS 10110's, and to other computer systems using CS 10110's UID addressing.

Effectively, UID addressing means that the address (or memory) space of a particular CS 10110 includes the address space of all systems, for example disc drives or other CS 10110s, to which that particular CS 10110 has access. UID addressing allows any process in any CS 10110 to obtain access to any
 50 object in any CS 10110 to which it has physical access, for example, another CS 10110 on the other side of the world. This access is constrained only by CS 10110's protection mechanism. In alternate embodiments of CS 10110, certain UIDs may be set aside for use only within a particular CS 10110 and may be unique only within that particular CS 10110. These reserved UIDs would, however, be a limited group known to all CS 10110 systems as not having uniqueness between systems, so that the unique object addressing
 55 capability of CS 10110's UID addressing is preserved.

As previously stated, AONs and physical descriptors are presently used for addressing within a CS 10110, effectively as shortened UIDs. In alternate embodiments of CS 10110, other forms of AONs may be used, or AONs may be discarded entirely and UIDs used for addressing within as well as between CS
 60 10110s.

CS 10110's addressing mechanisms are also based upon the use of descriptors within and between CS
 10110s.

Each descriptor includes an AON or UID field to identify a particular object, an offset field to specify a bit granular offset within the object, and a length field to specify a particular number of bits beginning at the specified offset. Descriptors may also include a type, or format field identifying the particular format of the
 65 data referred to by the descriptor. Physical descriptors are used for addressing MEM 10112 and, in this

case, the AON or UID field is replaced by a frame number field referring to a physical location in MEM 10112.

As stated above, descriptors are used for addressing within and between the separate, independent processors (FU 10120, EU 10122, MEM 10112, and IOS 10116) comprising CS 10110, thereby providing common, system wide bit granular addressing which includes format information. In particular, MEM 10112 responds to the type information fields of descriptors by performing formatting operations to provide requestors with data in the format specified by the requestor in the descriptor. MEM 10112 also accepts data in a format specified in a descriptor and reformats that data into a format most efficiently used by MEM 10112 to store the data.

As previously described, all operands are referred to in CS 10110 by Names wherein all Names within a particular S-Language dialect are of a uniform, fixed size and format. A K value specifying Name size is provided to FU 10120, at each change in S-Language dialect, and is used by FU 10120 in parsing Names from the instruction stream. In an alternate embodiment of CS 10110, all Names are the same size in all S-Language dialects, so that K values, and the associated circuitry in FU 10120's parser, are not required.

Finally, in descriptions of CS 10110's use of SOPs, FU 10120's microinstruction circuitry was described as storing one or more S-Interpreters. S-Interpreters are sets of sequences of microinstructions for interpreting the SOPs of various S-Language dialects and providing corresponding sequences of microinstructions to control CS 10110. In an alternate embodiment of CS 10110, these S-Interpreters (SITT 11012) would be stored in MEM 10112. FU 10120 would receive SOPs from the instruction stream and, using one or more S-Interpreter Base Pointers (that is, architectural base pointers pointing to the SITT 11012 in MEM 10112), address the SITT 11012 stored in MEM 10112. MEM 10112 would respond by providing, from the SITT 11012 in MEM 10112, sequences of microinstructions to be used directly in controlling CS 10110. Alternately, the SITT 11012 in MEM 10112 could provide conventional instructions usable by a conventional CPU, for example, Fortran or machine language instructions. This, for example, would allow FU 10120 to be replaced by a conventional CPU, such as a Data General Corporation Eclipse®.

Having briefly summarized certain features of CS 10110, and alternate embodiments of certain of these features, the structure and operation of CS 10110 will be described in detail below.

2. DETAILED DESCRIPTION OF CS 10110 MAJOR SUBSYSTEMS

(Figs. 201—206, 207—274)

Having previously described the overall structure and operation of CS 10110, the structure and operation of CS 10110's major subsystems will next be individually described in further detail. As previously discussed, CS 10110's major subsystems are, in the order in which they will be described, MEM 10112, FU 10120, EU 10122, IOS 10116, and DP 10118. Individual block diagrams of MEM 10112, FU 10120, EU 10122, IOS 10116, and DP 10118 are shown in, respectively, Figures 201 through 205. Figures 201 through 205 may be assembled as shown in Fig. 206 to construct a more detailed block diagram of CS 10110 corresponding to that shown in Fig. 101. For the purposes of the following descriptions, it is assumed that Figs. 201 through 205 have been assembled as shown in Fig. 206 to construct such a block diagram. Further diagrams will be presented in following descriptions as required to convey structure and operation of CS 10110 to one of ordinary skill in the art.

As previously described, MEM 10112 is an intelligent, prioritizing memory having separate and independent ports MIO 10128 and MJP 10140 to, respectively, IOS 10116 and JP 10114. MEM 10112 is shared by and is accessible to both JP 10114 and IOS 10116 and is the primary memory of CS 10110. In addition, MEM 10112 is the primary path for information transferred between the external world (through IOS 10116) and JP 10114.

As will be described further below, MEM 10112 is a two-level memory providing fast access to data stored therein. MEM 10112 first level is comprised of a large set of random access arrays and MEM 10112 second level is comprised of a high speed cache whose operation is generally transparent to memory users, that is JP 10114 and IOS 10116. Information stored in MEM 10112, in either level, appears to be bit addressable to both JP 10114 and IOS 10116. In addition, MEM 10112 presents simple interfaces to both JP 10114 and IOS 10116. Due to a high degree of pipe lining (concurrent and overlapping memory operations) MEM 10112 interfaces to both JP 10114 and IOS 10116 appear as if each HP 10114 and IOS 10116 have full access to MEM 10112. This feature allows data transfer rates of up to, for example, 63.6 megabytes per second from MEM 10112 and 50 megabytes per second to MEM 10112.

In the following descriptions, certain terminology used on those descriptions will be introduced first, followed by description of MEM 10112 physical organization. Then MEM 10112 port structures will be described, followed by descriptions of MEM 10112's control organization and control flow. Next, MEM 10112's interfaces to JP 10114 and IOS 10116 will be described. Following these overall descriptions the major logical structures of MEM 10112 will be individually described, starting at MEM 10112's interfaces to JP 10114 and IOS 10116 and proceeding inwardly to MEM 10112's first (or bulk) level of data stored. Finally, certain features of MEM 10112 microcode control structure will be described.

A. MEM 10112 (Figs. 201, 206, 207—237)

a. Terminology

Certain terms are used throughout the following descriptions and are defined here below for reference

by the reader.

A word is 32 bits of data

A byte is 8 bits of data

A block is 128 bits of data (that is, 4 words).

A block is always aligned on a block boundary, that is the low order 7 bits of logical or physical address are zero (see Chapter 1, Sections A.f and D. Descriptions of CS 10110 Addressing).

The term aligned refers to the starting bit address of a data item relative to certain address boundaries. A starting bit address is block aligned when the low order 7 bits of starting bit address are equal to zero, that is the starting bit address falls on a boundary between adjacent blocks. A word aligned starting bit address means that the low order 5 bits of starting bit address are zero, the starting bit address points to a boundary between adjacent words. A byte aligned starting bit address means that the low order 3 bits of starting bit address are zero, the starting bit address points to a boundary between adjacent bytes.

Bit granular data has a starting bit address falling within a byte, but not on a byte boundary, or the address is aligned on a byte boundary but the length of the data is bit granular, that is not a multiple of 8 bits.

b. MEM 10112 Physical Structure (Fig. 201)

Referring to Fig. 201, a partial block diagram of MEM 10112 is shown. Major functional units of MEM 10112 are Main Store Bank (MSB) 20110, including Memory Arrays (MA's) 20112, Bank Controller (BC) 20114, Memory Cache (MC) 20116, including Bypass Write File (BYF) 20118, Field Isolation Unit (FIU) 20120, and Memory Interface Controller (MIC) 20122.

MSB 20110 comprises MEM 10112's first or bulk level of storage. MSB 20110 may include from one to, for example, 16 MA 20112's. Each MA 20112 may have a storage capacity, for example, 256 K-byte, 512 K-byte, 1 M-byte, or 2 M-bytes of storage capacity. As will be described further below, MA 20112's of different capacities may be used together in MSB 20110. Each MA 20112 has a data input connected in parallel to Write Data (WD) Bus 20124 and a data output connected in parallel to Read Data (RD) Bus 20126. MA's 20112 also have control and address ports connected in parallel to address and control (ADCTL) Bus 20128. In particular, Data Inputs 20124 of Memory Arrays 20112 are connected in parallel to Write Data (WD) Bus 20126, and Data Outputs 20128 of Memory Arrays 20112 are connected in parallel to Read Data (RD) Bus 20130. Control Address Ports 20132 of Memory Arrays 20112 are connected in parallel to Address and Control (ADCTL) Bus 20134.

Data Output 20136 of Bank Controller 20114 is connected to WD Bus 20126 and Data Input 20138 of BC 20114 is connected to RD Bus 20130. Control and Address Port 20140 of BC 20114 is connected to ADCTL Bus 20134. BC 20114's Data Input 20142 is connected to MC 20116's Data Output 20144 through Store Back Data (SBD) Bus 20146. BC 20114's Store Back Address Input 20148 is connected to MC 20116 Store Back Address Output 20150 through Store Back Address (SBA) Bus 20152. BC 20114's Read Data Output 20154 is connected to MC 20116's Read Data Input 20156 through Read Data Out (RDO) Bus 20158. BC 20114's Control Port 20160 is connected to Memory Control (MCNTL) Bus 20164.

MC 20116 has Output 20166 connected to MIO Bus 10131 through MIO Port 10128, and Port 20168 connected to MOD Bus 10144 through MJP Port 10140. Control Port 20170 of MC 20116 is connected to MCNTL Bus 20164. Input 20172 of BYF 20118 is connected to IOM Bus 10130 through MIO Port 10128, and Output 20176 is connected to SBD Bus 20146 through Bypass Write In (BWI) Bus 20178.

Finally, FIU 20120 has an Output 20180 and an Input 20182 connected to, respectively, MIO Bus 10129 and IOM Bus 10130 through MIO Port 10128. Input 20184 and Port 20186 are connected to, respectively, JPD Bus 10142 and MOD Bus 10144 through MJP Port 10140. Control Port 20188 is connected to MCNTL Bus 20164. Referring finally to MIC 20122, MIC 20122 has Control Port 20190 and Input 20192 connected to, respectively, IOMC Bus 10131 and IOM Bus 10130 through MIO Port 10128. Control Port 20194 and Input 20196 are connected, respectively, to JPMC Bus 10147 and Physical Descriptor (PD) Bus 10146 through MJP Port 10140. Control Port 20198 is connected to MCNTL Bus 20164.

c. MEM 10112 General Operation

Referring first to MEM 10112's interface to IOS 10116, this interface includes MIO Bus 10129, IOM Bus 10130, and IOMC Bus 10131. Read and Write Addresses and data to be written into MEM 10112 are transferred from IOS 10116 to MEM 10112 through IOM Bus 10130. Data read from MEM 10112 is transferred to IOS 10116 through MIO Bus 10129. IOMC 10131 is a Bi-directional Control bus between MEM 10112 and IOS 10116 and, as described further below, transfers control signals between MEM 10112 and IOS 10116 to control transfer of data between MEM 10112 and IOS 10116.

MEM 10112's interface to JP 10114 is MJP Port 10140 and includes JPD Bus 10142, MOD Bus 10144, PD Bus 10146, and JPMC Bus 10147. Physical descriptors, that is MEM 10112 physical read and write addresses, are transferred from JP 10114 to MEM 10112 through PD Bus 10146. S Ops, that is sequences of S Instructions and operand names, are transferred from MEM 10112 to JP 10114 through MOD Bus 10144 while data to be written into MEM 10112 from JP 10114 is transferred from JP 10114 to MEM 10112 through JPD Bus 10142. JPMC Bus 10147 is a By-directional Control bus for transferring command and control signals between MEM 10112 and JP 10114 for controlling transfer of data between MEM 10112 and JP 10114. As will be described further below, MJP Port 10140, and in particular MOD Bus 10144 and PD Bus

EP 0 067 556 B1

10146, is generally physically organized as a single port that operates as a dual port. In a first case, MJP Port 10140 operates as a Job Processor Instruction (JI) Port for transferring S Ops from MEM 10112 to JP 10114. In a second case, MOD 10144 and PD 10146 operate as a Job Processor Operand (JO) Port for transfer of operands, from MEM 10112 to JP 10114, while JPD Bus 10142 and PD Bus transfer operands from JP 10114 to MEM 10112.

Referring to MSB 20110, MSB 20110 contains MEM 10112's first, or bulk, level of storage capacity. MSB 20110 may contain from one to, for example, 16 MA's 20112. Each MA 20112 contains a dynamic, random access memory array and may have a storage capacity of, for example 256 Kilo-bytes, 512 Kilo-bytes, 1 Mega-bytes, or 2 Mega-bytes. MEM 10112 may therefore have a physical capacity of up to, for example, 16 Mega-bytes of bulk storage. As will be described further below. MA 20112's of different capacity may be used together in MSB 20110, for example, four 2 Mega-byte MA 20112's and four 1 Megabyte MA 20112's.

BC 20114 controls operation of MA's 20112 and is the path for transfer of data to and from MA's 20112. In addition, BC 20114 performs error detection and correction on data transferred into and out of MA's 20112, refreshes data stored in MA's 20112, and, during a refresh operations, performs error detection and correction of data stored in MA's 20112.

MC 20116 comprises MEM 10112's second, or cache, level of storage capacity and contains, for example 8 Kilo-bytes of high speed memory. MC 20116, including BYF 20118, is also the path for data transfer between MSB 20110 (through BC 20114) and JP 10114 and IOS 10116. In general, all read and write operations between JP 10114 and IOS 10116 are through MC 20116. IOS 10116 may, however, perform read and write operations of complete blocks by-passing MC 20116. Block write operations from IOS 10116 are accomplished through BYF 20118 while block read operations are performed through a data transfer path internal to MC 20116 and shown and described below. All read and write operations between MEM 10112 and JP 10114, however, must be performed through the cache internal to MC 20116, as will be shown and described further below.

As also shown and described below, FIU 20120 includes write data registers for receiving data to be written into MEM 10112 from JP 10114 and IOS 10116, and circuitry for manipulating data read from MSB 20110 so that MEM 10112 appears as a bit addressable memory. FIU 20120, in addition to providing bit addressability of MEM 10112, performs right and left alignment of data, zero fill of data, sign extension operations, and other data manipulation operations described further below. In performing these data manipulation operations on data read from MEM 10112 to JP 10114, MOD Bus 10144 is used as a data path internal to MEM 10112 for transferring of data from MC 20116 to FIU 20120, and from FIU 20120 to MC 20116. That is, data to be transferred to JP 10114 is read from MC 20116, transferred through MOD Bus 10144 to FIU 20120, manipulated by FIU 20120, and transferred from FIU 20120 to JP 10114 through MOD Bus 10144.

MIC 20122 contains circuitry controlling operation of MEM 10112 and, in particular, controls MEM 10112's interface with JP 10114 and IOS 10116. MIC 20122 receives MEM 10112 read and write request, that is read and write addresses through PD Bus 10146 and IOM Bus 10130 and control signals through JPMC Bus 10147 and IOMC Bus 10131, and provides control signals to BC 20114, MC 20116, and FIU 20120 through MCNTL Bus 20164.

Having described the overall structure and operation of MEM 10112, the structure and operation of MEM 10112's Port, MIO Port 10128, and MJP Port 10140, will be described next, followed by descriptions of MEM 10112's control structure and the control and flow of MEM 10112 read and write requests.

d. MEM 10112 Port Structure

MEM 10112 port structure is designed to provide a simple interface to JP 10114 and IOS 10116. While providing fast and flexible operation in servicing MEM 10112 read and write requests from JP 10114 and IOS 10116. In this regard, MEM 10112, as will be described further below, may handle up to 4 read and write requests concurrently and up to, for example, a 63.6 M-byte per second data rate. In addition MEM 10112 is capable of performing bit granular addressing, block read and write operations, and data manipulations, such as alignment and filling, to enable JP 10114 and IOS 10116 to operate most efficiently.

MEM 10112 effectively services requests from three ports. These ports are MIO Port 10128 to IOS 10116, hereafter referred to as IO Port, and JI and JO Ports, described above, to JP 10114. These three ports share the entire address base of MEM 10112, but IOS 10116, for example, may be limited from making full use of MEM 10112's address space. Each port has a different set of allowed operations. For example, JO Port can use a bit granular addresses but can reference only 32 bits of data on each request. JI Port can make read requests only to word align 32 bit data items. IO Port may reference bit granular data, and, as described further below, may read or write up to 16 bytes on each read or write request. The characteristics of each of these ports will be discussed next below.

1. IO Port Characteristics

IOS 10116 may access MEM 10112 in either of two modes. The first mode is block transfers by-passing or through the cache in MC 20116, and the second is non-block transfer through the cache and MC 20116.

Block by-passes may occur for both read and write operations. A read or write operation is eligible for a block by-pass if the data is on block boundaries, is 16 bytes long, and the read or write request is not accompanied by a control signal indicating that an encache (load into MC 20116's cache) operation is to be

EP 0 067 556 B1

performed. A by-pass operation takes place only if the block address, that is the physical address of the block in MEM 10112 does not address a currently encached block, that is the block is not present in MC 20116's cache. If the block is encached in MC 20116's cache, the read or write transfer is to MC 20116's cache.

5 Partial block references, that is non-full block transfers will go through MC 20116's cache. If a cache miss occurs, that is the reference data is not present in MC 20116's cache, MEM 10112's control structures transfer the data to or from MSB 20110 and update MC 20116's cache. It should be noted that partial blocks may be as short as one byte, or up to 15 bytes long. A starting byte address may be anywhere within a block, but the partial block's length may not cross a block boundary.

10 Bit length transfers, that is transfers of data items having a length of 1 to 16 bits and not a multiple of a byte, or where address is not on a byte boundary, go through MC 20116's cache. These operations may cross byte, word, or block boundaries but may not cross page boundaries. These specific operations requested by IO port determines whether a read or write request is a partial block or bit length transfer.

15 2. JO Port Characteristics

All read or write requests from JO Port must go through MC 20116's cache; by-pass operations may not be performed. The data transferred between MEM 10112 and JP 10114 is always 32 bits in length but, of the 32 bits passed, from zero to 32 bits may be valid data. JP 10114 determines the location of valid data within the 32 bits by referring to certain FIU specification bits provided as part of the read or write request. As will be described further below, FIU specification bits, and other control bits, are provided to MIC 20122 by JP 10114 through JPMC Bus 10147 when each read or write request is made.

20 While MEM 10112 does not perform block by-pass operations to JP 10114, MEM 10112 may perform a cache read-through operation. Such operations occur on a JP 10114 read request wherein the requested data is not present in MC 20116's cache. If the JP 10114 read request is for a full word, which is word aligned, MEM 10112's Load Manager, discussed below, transfers the requested data directly to JP 10114 while concurrently loading the requested data into MC 20116's cache. This operation is referred to as a "hand-off" operation. These operations may also be performed by IO Port for 16 bit half words aligned on the right hand half word of a 32 bit word, or if a full block is handed left and loaded into MC 20116's cache.

30 3. JI Port Characteristics

All JI Port requests are satisfied through MC 20116's cache; MEM 10112 does not perform by-pass operations to JI Port. JI Port requests are always read requests for full-word aligned words and are handed off, as described above, if a cache miss occurs. In most other respects, JI Port requests are similar to JO Port requests.

35 Having described the overall structure and operation of MEM 10112, including MEM 10112's input and output ports to JP 10114 and IOS 10116, MEM 10112's control structure will be described next below.

e. MEM 10112 Control Structure and Operation (Fig. 207)

40 Referring to Fig. 207, a more detailed block diagram of MIC 20116 is shown. Fig. 207 will be referred to in conjunction with Fig. 201 in the following discussion of MEM 10112's control structure.

1. MEM 10112 Control Structure

45 Referring first to Fig. 207, MCNTL Bus 20164 is represented as including MCNTL-BC Bus 20164A, MCNTL-MC Bus 20164B, and MCNTL-FIU Bus 20164C. Buses 20164A, 20164B, and 20164C are branches of MCNTL Bus 20164 connected to, respectively, BC 20114, MC 20116, and FIU 20120. Also represented in Fig. 207 are PD Bus 10146 and JPMC Bus 10147 to JP 10114, and IOM Bus 10130 and IOMC Bus 10131 to IOS 10116.

JO Port Address Register (JOPAR) 20710 and JI Port Address Register (JIPAR) 20712 have inputs connected from PD Bus 10146. IO Port Address Register (IOPAR) 20714 has an input connected from IOM Bus 10130. Port Control Logic (PC) 20716 has a bi-directional input/outputs connected from JPC 10147 and IOMC Bus 10131. By-pass Read/Write Control Logic (BR/WC) 20718 has a bidirectional input/output connected from IOMC Bus 10131.

50 Outputs of JOPAR 20710, JIPAR 20712, and IOPAR 20714 are connected to inputs of Port Request Multiplexer (PRMUX) 20720 through, respectively, Buses 20732, 20734, 20736. PRMUX 20720's output in turn is connected to Bus 20738. Branches of Bus 20738 are connected to inputs of Load Pointers (LP) 20724, Miss Control (MISSC) 20726, and Request Manager (RM) 20722, and to Buses MCNTL-MC 20164B and MCNTL-FIU 20164C.

55 Outputs of PC 20716 are connected to inputs of JOPAR 20710, JIPAR 20712, IOPAR 20714, PRMUX 20720, and LP 20724 through Bus 20738. Bus 20740 is connected between an input/output of PC 20716 and in input/output of RM 20722.

60 An output of BR/WC 20718 is connected to MCNTL-MC Bus 20164B through Bus 20742. Inputs of BR/WC 20718 are connected from outputs of RM 20722 and Read Queue (RQ) 20728 through, respectively, Buses 20744 and 20746.

65 RM 20722 has outputs connected to MCNTL-BC Bus 20164A, MCNTL-FIU Bus 20164C, and input of MISSC 20726, and an input of LP 20724 through, respectively, Buses 20748, 20750, 20752, and 20754.

MISSC 20726's output is connected to MCNTL-BC Bus 20164A. Outputs of LP 20724 are connected to MCNTL-MC Bus 20164B and to an input of LM 20730 through, respectively, Buses 20756 and 20758. RQ 20728's input is connected from MCNTL-MC Bus 20164B through Bus 20760 and RQ 20728 has outputs connected to an input of LP 20724, through Bus 20762, and as previously described to an input of BRWC 20718 through Bus 20746. Finally, LM 20730's output is connected to MCNTL-MC Bus 20164B through Bus 20764.

Having described the structure of MIC 20716 with reference to Fig. 207, and having previously described the structure of MEM 10112 with reference to Fig. 201, MEM 10112's control structure operation will next be described with reference to both figures 201 and 207.

10

2. MEM 10112 Control Operation

Referring first to Fig. 207, JOPAR 20710, JIPAR 20712, and IOPAR 20714 are, as previously described, connected from PD Bus 10146 from JP 10114 and IOM Bus 10130 from IOS 10116. JPAR 20710, JIPAR 20712, and IOPAR 20714 receive read and write request addresses from JP 10114 and IOS 10116 and store these addresses for subsequent service by MEM 10112. As will be described further below, these address inputs from JP 10114 and IOS 10116 include FIU information specifying what data manipulation operations must be performed by FIU 20120 before requested data is transferred to the requestor or written into MEM 10112, information regarding the destination data read from MEM 10112 is to be provided to, information regarding the type of operation to be performed by MEM 10112, and information regarding operand length. Request address information received and stored in JOPAR 20710, JIPAR 20712, and IOPAR 20714 is retained therein until MEM 10112 has initiated service of the corresponding requests. MEM 10112 will accept further request address information into a given port register only after a previous request into that port has been serviced or aborted. Address information outputs from JOPAR 20710, JIPAR 20712, and IOPAR 20714 is transferred through PRMUX 20720 to Bus 20738 and from there to RM 20722, MC 20116, and FIU 20120 as service of individual requests is initiated. As will be described below, this address information will be transferred through PRMUX 20720 and Bus 20738 to LP 20724 for use in servicing a cache miss upon occurrence of a MC 20116 miss.

PC 20716 receives command and control signals pertinent to each requested memory operation from JP 10114 and IOS 10116 through JPMC Bus 10147 and IOSB Bus 10131. PC 20716 includes request arbitration logic and port state logic. Request arbitration logic determines the sequence in which IO, JI, JO ports are serviced, and when each port is to be serviced. In determining the sequence of port service, request arbitration logic uses present port state information for each port from the port state logic, information from JPMC Bus 10147 and IOMC Bus 10131 regarding each incoming request, and information from RM 20722 concerning the present state of operation of MEM 10112. Port state logic selects each particular port to be serviced and, by control signals through Bus 20738, enables transfer of each port's request address information from JOPAR 20710, JIPAR 20712, and IOPAR 20714 through PRMUX 20720 to Bus 20738 for use by the remainder of MEM 10112's control logic in servicing the selected port. In addition to request information received from JP 10114 and IOS 10116 through JPMC Bus 10147 and IOMC Bus 10131, port state logic utilizes information from RM 20722 and, upon occurrence of a cache miss, from LM 20730 (for clarity of presentation, this connection is not represented in Fig. 207). Port state logic also controls various port state flag signals, for example port availability signals, signals indicating valid requests, and signals indicating that various ports are waiting service.

RM 20722 controls execution of service for each request. RM 20722 is a microcode controlled "micromachine" executing programs called for by requested MEM 10112 operations. Inputs of RM 20722 include request address information from IOPAR 20714, JIPAR 20712, and JOPAR 20710, including information regarding the type of MEM 10112 operation to be performed in servicing a particular request, interrupt signals from other MEM 10112 control elements, and, for example, start signals from PC 20716's request arbitration logic. RM 20722 provides control signals to FIU 20120, MC 20116, and most other parts of MEM 10112's control structure.

Referring to Fig. 201, MC 20116's cache is, for example, an 8 Kilo-byte, four set associative cache used to provide rapid access to a subset of data stored in MSB 20110. The subset of MSB 20110 data stored in MC 20116's cache at any time is the data most recently used by JP 10114 or IOS 10116. MC 20116's cache, described further below, includes tag store comparison logic for determining encached addresses, a data store containing corresponding encached data, and registers and logic necessary to up-date cache contents upon occurrence of a cache miss. Registers and logic for servicing cache misses includes logic for determining the least recently used cache entry and registers for capture and storage of information regarding missed cache references, for example modify bits and replacement page numbers. Inputs to MC 20116 are provided from RM 20722, LM 20730 (discussed further below), FIU 20120, MSB 20110 (through BC 20114), LP 20724 (described further below) and address information from PRMUX 20720. Outputs of MC 20116 include data and go to FIU 20120 (through MOD Bus 10144), the data requestors (JP 10114 and IOS 10116), and a MC 20116 Write Back File (described further below).

As previously described, FIU 20120 includes logic necessary to make MEM 10112 appear bit addressable. In addition, FIU 20120 includes logic for performing certain data manipulation operations as required by the requestors (JP 10114 or IOS 10116). Data is transferred into FIU 20120 from MC 20116 through that portion of MOD Bus 10144 internal to MEM 10112, is manipulated as required, and is then

transferred to the requestor through MOD Bus 10144 or MIO Bus 10129. In the case of writes requiring read-modify-write of encached data, the data is transferred back to MC 20116 through MOD Bus 10144 after manipulation. In general, data manipulation operations include locating requested data onto selected MOD Bus 10144 or MIO Bus 10139 lines and filling unused bus lines as specified by the requestor. Data inputs to
 5 FIU 20120 may be provided from MC 20116 or JP 10114 through MOD Bus 10144 or from IOS 10116 through IOM Bus 10130. Data outputs from FIU 20120 may be provided to MC 20116, JP 10114, or IOS 10116 through these same buses. Control information is provided to FIU 20120 from RM 20722 through Bus 20748 and MCNTL-FIU Bus 20164C. Address information may be provided to FIU 20120 from JOPAR 20710, JIPAR 20712, or IOPAR 20714 through PRMUX 20720, Bus 20738, and MCNTL-FIU Bus 20164C.

10 Returning to Fig. 207, MISSC 20726 is used in handling MC 20116 misses. In the event of a request referring to data not in MC 20116's cache, MISSC 20726 stores block address of the reference and type of operation to be performed, this information being provided from an address register in MC 20116 and from RM 20722. MISSC 20726 utilizes this information in generating a command to BC 20114, through MCNTL-BC Bus 20164A, for a data read from MSB 20110 to obtain the referenced data. BC 20114 places this
 15 command in a queue, or register, and subsequently executes the commanded read operation. MISSC 20726 also generates an entry into RQ 20728 (described further below) indicating the type of operation to be performed when referenced data is subsequently read from MSB 20110.

RQ 20728 is, for example, a three-level deep queue storing information indicating operations associated with data being read from MSB 20110. Two kinds of operation may be indicated: block by-pass reads and cache loads. If a cache load is specified, that is a read and store to MC 20116's cache, is indicated, RM 20722 is interrupted and forced to place other MEM 10112 operations in idle until cache load is completed. A block by-pass read operation results in by-pass read control (described below) assuming control of the data from MSB 20110. Inputs to RQ 20728 are control signals from RM 20752, MISSC 20726, and BC 20114. RQ 20728 provides control outputs to LP 20724 (described below) LM 20730 (described
 25 below) RM 20722, and by-pass read control (described below).

LP 20724 is a set of registers for storing information necessary for servicing MC 20116 misses that result in order to load MC 20116's tag store. LM 20730 uses this information when data stored in MSB 20110 and read from MSB 20110 to service a MC 20116 cache miss, becomes available through BC 20114. Inputs to LP 20724 include the address of the missing reference, provided from JOPAR 20710, JIPAR 20712, or
 30 IOPAR 20714 through PRMUX 20720 and Bus 20738, commands from RM 20722, and a control signal from RQ 20728. LP 20724 outputs include addresses of missed references to MC 20116, through Bus 20756 and MNCTL-MC 20164B, and command signals to LM 20730 and BRAWC 20718.

LM 20730, referred to above, controls loading of MC 20116's cache with data from MSB 20110 after occurrence of a cache miss. RQ 20728, referred to above, indicates, for each data read from MSB 20110,
 35 whether the data read is the result of a MC 20116 cache miss. If the data is read from MSB 20110 as a result of a cache miss, LM 20730 proceeds to issue a sequence of control signals for loading the data from MSB 20110 and its associated address into MC 20116's cache. This data is transferred into MC 20116's cache data store while the block address, from LP 20724 is transferred into the tag store (described in the following discussion) of MC 20116's cache. If the transfer of data into MC 20116's cache replaces data previously
 40 resident in that cache, and that previous data is "dirty", that is has been written into so as to be different from an original copy of the data stored on MSB 20110, the modified data resident in MC 20116's cache must be written back into MSB 20110. This operation is performed through a Write Back File contained in MC 20116 and described below. In the event of such an operation, LM 20730 initiates a write back operation by MC 20116 and BC 20114, also as described below.

45 As will be described further in a following description, all MC 20116 cache load operations are full 4 word blocks. A request resulting in a MC 20116 cache miss may result in a "hand-off", that is a read operation of a full 4 word block. Handoff operations also may be of single 32 bit words wherein a 32 bit word aligned word is transferred from JP 10114 or a 16 bit operand aligned on the right half-word is transferred from IOS 10116. In such a handoff operation, LM 20730 will send a valid request signal to the
 50 requesting port and a handoff operation will be performed. Otherwise, a waiting signal will be sent to the requesting port and the request will re-enter the priority queue of PC 20716 for subsequent execution. To accomplish these operations, LM 20730 receives input from RQ 20728, (not shown in Fig. 207 for clarity of presentation) and LP 20724. LM 20730 provides outputs to port state logic of PC 20716, to MC 20116, MC 20116's Write Back File and MC 20116's Write Back Address Register and to BC 20114.

55 Referring to Fig. 201, as previously discussed IOS 20116 may request a full block write operation directly to MSB 20110. Such a by-pass write request may be honored if the block being transferred is not encached in MC 20116's cache. In such a case, RM 20722 will initiate the transfer setting up By-pass Write Control logic in BRAWC 20718, and may then pass control of the operation over to BRAWC 20718's By-Pass Write Control logic for completion. By-pass Write Control may then accept the remaining portion of the
 60 data block from IOS 10116, generating appropriate hand shaking signals through IOMC Bus 10131, and load the data block into BYF 20118 and MC 20116. MISSC 20726 will provide a by-pass write command to BC 20114, through MNCTL-PC Bus 20164A. BC 20114 will then transfer the data block from BYF 20118 and into MA's 20112 and MSB 20110.

65 As previously described, BYF 20118 receives data from IOM Bus 10130 and provides data output to BC 20114 through BWY Bus 20178 and SBD Bus 20146. BYF 20118 is capable of simultaneously accepting data

EP 0 067 556 B1

from IOM Bus 10130 while reading data out to BC 20114. Control of writing data into BYF 20118 is provided from BRWC 20718's By-Pass Write Control logic.

IOS 10116 may, as previously described, request a full block read operation by-passing MC 20116's cache. In such a case, BRWC 20718's by-pass read control handles data transfer to IOS 10116 and generates required hand shaking signals to IOS 10116 through IOMC Bus 10131. The data path for by-pass read operations is through a data path internal to MC 20116, rather than through BYF 20118. This internal data path is RDO Bus 20158 to MIO Bus 10129.

As previously described, BC 20114 manages all data transfers to and from MA's 20112 in MSB 20110. BC 20114 receives requests for data transfers from RM 20722 in an internal queue register. All data transfers to and from MSB 20110 are full block transfers with block aligned addresses. On data write operations, BC 20114 receives data from BWF 20118 or from MC 20116's Write Back File and transfers the data into MA's 20112. During read operations, BC 20114 fetches the data block from MA's 20112 and places the data block on RDO Bus 20158 while signalling to MIC 20122 that the data is available. As described above, MIC 20122 tracks and controls transfer of data and BYF 20118, MC 20116, and MC 20116's Write Back File, and directs data read from MSB 20110 to the appropriate destination, MC 20116's Data Store, JP 10114, or IOS 10116.

In addition to the above operations, BC 20114 controls refresh of MA's 20112 and performs error detection and correction operations. In this regard, BC 20114 performs two error detection and correction operations. In the first, BC 20114 detects single and double bit errors in data read from MSB 20110 and corrects single bit errors. In the second, BC 20114 reads data stored in MA's 20112 during refresh operations and performs single bit error detection. Whenever an error is detected, during either read operations or refresh operations, BC 20114 makes a record of that error in an error log contained in BC 20114 (described further in a following description). Both JP 10114 and IOS 10116 may read BC 20114's error log, and information from BC 20114's error log may be recorded in a CS 10110 maintenance log and to assist in repair and trouble shooting of CS 10110. BC 20114's error log may be addressed directly by RM 20722 and data from BC 20114's error log is transferred to JP 10114 or IOS 10116 in the same manner as data stored in MSB 20110.

Referring finally to MA's 20112, each MA 20112 contains an array of dynamic semiconductor random access memories. Each MA 20112 may contain 256 Kilo-bytes, 512 Kilo-bytes, 1 Mega-bytes, or 2 Mega-bytes of data storage. The storage capacity of each MA 20112 is organized as segments of 256 Kilo-bytes each. In addressing a particular MA 20112, BC 20114 selects that particular MA 20112 as will be described further below. BC 20114 concurrently selects a segment within that MA 20112, and a block of four words within that segment. Each word may comprise 39 bits of information, 32 bits of data and 7 bits of error correcting code. The full 39 bits of each MA 20112 word are transferred between BC 20114 and MA's 20112 during each read and write operation. Having briefly described the general structure and operation of MEM 10112, certain types of operations which may be performed by MEM 10112 will be described next below.

f. MEM 10112 Operations

MEM 10112 may perform two general types of operation. The first type are data transfer operations and the second type are memory maintenance operations. Data transfer operations may include read, write, and read and set. Memory maintenance operations may include read error log, repair block, and flush cache. Except during a flush cache operation, the existence of MC 20116 and its operation is invisible to the requestors, that is JP 10114 and IOS 10116.

A MEM 10112 read operation transfers data from MS 10112 to a requestor, either JP 10114 or IOS 10116. A read data transfer is asynchronous in that the requestor cannot predict elapsed time between submission of a memory operation request and return of requested data. Operation of a requestor in MEM 10112 is coordinated by a requested data available signal transmitted from MEM 10112 to the requestor.

A MEM 10112 write operation transfers data from either JP 10114 or IOS 10116 to MEM 10112. During such operations, JP 10114 is not required to wait for a signal from MEM 10112 that data provided to MEM 10112 from JP 10114 has been accepted. JP 10114 may transfer data to MEM 10112's JO Port whenever a JO Port available signal from MEM 10112 is present; read data is accepted immediately without further action or waiting required of JP 10114. Word write operations from IOS 10116 are performed in a similar manner. On block write operations, however, IOS 10116 is required to wait for a data taken signal from MEM 10112 before sending the 2nd, 3rd and 4th words of a block.

MEM 10112 has a capability to perform "lock bit" operations. In such operations, a bit granular read of the data is performed and the entire operand is transmitted to the requestor. At the same time, the most significant bit of the operand, that is the Lock Bit, is set to one in the copy of data stored in MEM 10112. In the operand sent to the requestor, the lock bit remains at its previous value, the value before the current read and set operation. Test and set operations are performed by performing read and set operations wherein the data item length is specified to be one bit.

As previously described, MEM 10112 performs certain maintenance operations, including error detection. MEM 10112's Error Log in BC 20114 is a 32 bit register containing an address field and an error code field. On a first error to occur, the error type and in some cases, such as ERCC errors on read data stored in MSB 20110, the address of the data containing the error are stored in BC 20114's Error Log Register. An interrupt signal indicating detection of an error is raised at the same that information

regarding the error is stored in the Error Log. If multiple errors occur before Error Log is read and reset, the information regarding the first error will be retained and will remain valid. The Error Log code field will, however, indicate that more than one error has occurred.

5 JP 10114 may request a read Error Log operation referred to as a "Read Log and Reset" operation. In this operation, MEM 10112 reads the entire contents of Error Log to JP 10114, resets Error Log Register, and resets the interrupt signal indicating presence of an error. IOS 10116, as discussed further below, is limited to reading 16 bits at a time from MEM 10112. It therefore requires two read operations to read Error Log. First read operation to IOS 10116 reads an upper 16 bits of Error Log data and does not reset Error Log. The second read operation is performed in the same manner as a JP 10114 Read Log and Reset operation, 10 except that only the low order 16 bits of Error Log are read to IOS 10116.

MEM 10112 performs repair block operations to correct parity or ERCC errors in data stored in MC 20116's Cache or in data stored in MA's 20112. In a repair block procedure, parity bits for data stored in MC 20116's Cache, or ERCC check bits of data stored in MA's 20112, are modified to agree with the data bits of data stored therein. In this regard, repaired uncorrectible errors, such as two bit errors of data in MA's 15 20112, will have good ERCC and parity values. Until a repair block operation is performed, any read request directed to bad data, that is data having parity or ERCC check bits indicating invalid data, will be flagged as invalid. Repair block operations therefore allow such data to be read as valid, for example to be used in a data correction operation. Errors are ignored and not logged in BC 20114's Error Log in repair block operations. A write operation into an area containing bad data may be accomplished if MEM 10112's 20 internal operation does not require a read-modified-write procedure. Only byte aligned writes of integral byte length data residing in MC 20116 and word aligned writes of integral word lengths of data in MSP 20110 do not require read-modified-write operation. By utilizing such write operations, it is therefore possible to overwrite bad data by use of normal write operations before or instead of repair block operations.

25 MEM 10112 performs a cache flush operation in event of a power failure, that is when MEM 10112 goes into battery back-up operation. In such an event, only MA's 20112 and BC 20114 remain powered. Before JP 10114 and IOS 10116 lose power, JP 10114 and IOS 10116 must transfer to MEM 10112 any data, including operating state, to be saved. This is accomplished by using a series of normal write operations. After conclusion of these write operations, both JP 10114 and IOS 10116 transmit a flush cache request to MEM 30 10112. Upon receiving two flush cache requests, MEM 10112 flushes MC 20116's Cache so that all dirty data encached in MC 20116's Cache is transferred into MA's 20112 before power is lost. If only JP 10114 or IOS 10116 is operating, DP 10118 will detect this fact and will have transmitted an enabling signal (FLUSHOK) to MEM 10112 during system initialization. FLUSHOK enables MEM 10112 to perform cache flush upon receiving a single flush cache request. After a cache flush operation, no further MEM 10112 operations are possible until Dp 10118 resets a power failure lock-out signal to enable MEM 10112 to resume normal 35 operation.

Having described MEM 10112's overall structure and operation and certain operations which may be performed by MEM 10112, MEM 10112's interfaces to JP 10114 and IOS 10116 will be described next below.

40 g. MEM 10112 Interfaces to JP 10114 and IOS 10116 (Figs. 209, 210, 211, 204)

As previously described, MJP Port 10140 and MIO Port 10128 logically function as three independent ports. These ports are an IO Port to IOS 10116, a JP Operand Port to JP 10114 and a JP Instruction Port to JP 10114. Referring to Figs. 209, 210, and 211, diagrammatic representations of IO Port 20910, JP Operand (JPO) Port 21010, and JP Instruction (JPI) port 21110 are shown respectively.

45 IO Port 20910 handles all IOS 10116 requests to MEM 10112, including transfer of both instructions and operands. JPO Port 21010 is used for read and write operations of operands, for example numeric values, to and from JP 10114. JPI Port 21110 is used to read SInS, that is SOPs and operand NAMEs, from MEM 10112 to JP 10114. Memory service requests to a particular port are serviced in the order that the requests are provided to the Port. Serial order is not maintained between requests to different ports, but ports may 50 be serviced in the order of their priority. In one embodiment of the present invention, IO Port 20910 is accorded highest priority, followed by JPO port 21010, and lastly by JPI Port 21110, with requests currently contained in a port having priority over incoming requests. As described above and will be described in more detail in following descriptions, MEM 10112 operations are pipelined. This pipelining allows interleaving of requests from IO Port 20910, JPO Port 21010, and JPI port 21110, as well as overlapping 55 service of requests at a particular port. By overlapping operations it is meant that one operation servicing a particular port begins before a previous operation servicing that port has been completed.

1. IO Port 20910 Operating Characteristics (Figs. 209, 204)

60 Referring first to Fig. 209, a diagrammatic representation of IO port 20910 is shown. Signals are transmitted between IO Port 20910 and IOS 10116 through MIO Bus 10129, IOM Bus 10130, and IOMC Bus 10131. MIO Bus 10129 is a unidirectional bus having inputs from MC 20116 and FIU 20120 and dedicated to transfers of data and instructions from MEM 10112 to IOS 10118. IOM Bus 10130 is likewise a unidirectional bus and is dedicated to the transfer, from IOS 10118 to MEM 10112, of read addresses, write addresses, and data to be written into MEM 10112. IOM Bus 10130 provides inputs to BYF 20118, FIU 20120, and MIC 20122. IOMC 65 Bus 10131 is a set of dedicated signal lines for the exchange of control signals between IOS 10118 and MEM

10112.

Referring first to MIO Bus 10129, MIO Bus 10129 is a 36 bit bus receiving read data inputs from MC 20116's Cache and from FIU 20120. A single read operation from MEM 10112 to IOS 10116 transfers one 32 bit word (or 4 bytes) of data (MIO(0—31)) and four bits of odd parity (MIOP(0—3)), or one parity bit per byte.

Referring next to IOM Bus 10130, a single transfer from IOS 10116 to MEM 10112 includes 36 bits of information which may comprise either a memory request comprising a physical address, a true length, and command bits. These memory requests and data are multiplexed onto IOM 10130 by IOS 10116.

Data transfers from IOS 10116 to MEM 10112 each comprise a single 32 bit data word (IOM(0—31)) and four bits of odd parity (IOMP(0—3)) or one parity bit per byte. Such data transfers are received by either BYF 20118 or FIU 20120.

Each IOS 10116 memory request to MEM 10112, as described above, an address field, a length field, and an operation code field. Address and length fields occupy the 32 IOM Bus 10130 lines used for transfer of data to MEM 10112 in IOS 10116 write operations. Length field includes four bits of information occupying bits (IOM(03)) of IOM Bus 10130 and address field contains 27 bits of information occupying bits (IOM(4—31)) of IOM Bus 10130. Together, address and length field specify a physical starting address and true length of the particular data item to be written into or read from MEM 10112. Operation code field specifies the type of operation to be performed by MEM 10112. Certain basic operation codes comprise 3 bits of information occupying bits (IOMP (32—36)) of IOM Bus 10130; as described above. These same lines are used for transfer of parity bits during data transfers. Certain operations which may be requested of MEM 10112 by IOS 10116 are, together with their corresponding command code fields, are;

000 = read,
 001 = read and set,
 010 = write,
 011 = error,
 100 = read error log (first half),
 101 = read error log (second half) and reset,
 110 = repair block, and
 111 = flush cache.

Two further command bits may specify further operations to be performed by MEM 10112. A first command bit, indicates to MEM 10112 during write operations whether it is desirable to encache the data being written into MEM 10112 in MC 20116's Cache. IOS 10116 may set this bit to zero if reuse of the data is unlikely, thereby indicating to MEM 10112 that MEM 10112 should avoid encaching the data. IOS 10116 may set this bit to one if the data is likely to be reused, thereby indicating to MEM 10112 that it is preferable to encache the data. A second command bit is referred to a CYCLE. CYCLE command bit indicates to MEM 10112 whether a particular data transfer is a single cycle operation, that is a bit granular word, or a four cycle operation, that is a block aligned block or a byte aligned partial block.

IOMC 10131 includes a set of dedicated lines for exchange of control signals between IOS 10116 and MEM 10112 to coordinate operation of IOS 10116 and MEM 10112. A first such signal is Load IO Request (LIOR) from IOS 10116 to MEM 10112. When IOS 10116 wishes to load a memory request into MEM 10112, IOS 10116 asserts LIOR to MEM 10112. IOS 10116 must assert LIOR during the same system cycle during which the memory request, that is address, length, and command code fields, are valid.

If LIOR and IO Port Available (IOPA) signals, described below, are asserted during the same clock cycle, MEM 10112's port is loaded from IOS 10116 and IOPA is dropped, indicating the request has been accepted. If a load of a request is attempted and IOPA is not asserted, MEM 10112 remains unaware of the request, LIOR remains active, and the request must then be repeated when IOPA is asserted.

IOPA is a signal from MEM 10112 to IOS 10116 which is asserted by MEM 10112 when MEM 10112 is available to accept a new request from IOS 10116. IOPA may be asserted while a previous request from IOS 10116 is completing operation if the address, length, and operation code fields of the previous request are no longer required by MEM 10112, for example in servicing bypass operations.

IO Data Taken (TIOMD) is a signal from MEM 10112 to IOS 10116 indicating that MEM 10112 has accepted data from IOS 10116. IOS 10116 places a first data word on IOM Bus 10130 on the next system clock cycle after a write request is loaded; that is, LIOR has been asserted, a memory request presented, and IOPA dropped. MEM 10112 then takes that data word on the clock edge beginning the next system clock cycle. At this point, MEM 10112 asserts TIOMD to indicate the data has been accepted. On a single word operations TIOMD is not used by IOS 10116 as a first data word is always accepted by MEM 10112 if IO Port 20910 was available. On block operations, a first data word is always taken but a delay may occur between acceptance of first and second words. IOS 10116 is required to hold the second word valid on IOM Bus 10130 until MEM 10112 responds with TIOMD to indicate that the block operation may proceed.

Data Available for IO (DAVIO) is a signal asserted by MEM 10112 to IOS 10116 indicating that data requested by IOS 10116 is available. DAVIO is asserted by MEM 10112 during the system clock cycle in which MEM 10112 places the requested data on MIO Bus 10129. In any single word type transfer, DAVIO is active for a single system clock transfer. In block type transfers, DAVIO is normally active for four consecutive system clock cycles. Upon event of a single cycle "bubble" resulting from detection and

correction of an ERCC error by BC 20114, DAVIO will remain high for four non-consecutive system clock cycles and with a single cycle bubble, a non-assertion, in DAVIO corresponding to the detection and correction of the error.

IO Memory Interrupt (IMINT) is a signal asserted by MEM 10112 to IOS 10116 when BC 20114 places a record of a detected error in BC 20114's Error Log, as described above.

Previous MIO Transfer Invalid (PMIOI) signal is similarly a signal asserted by MEM 10112 to IOS 10116 regarding errors in data read from MEM 10112 to IOS 10116. If an uncorrectible error appears in such data, that is an error in two or more data bits, the incorrect data is read to IOS 10116 and PMIOI signal asserted by MEM 10112. Correctible, or single bit, errors in data do not result in assertion of PMIOI. MEM 10112 will assert PMIOI to IOS 10116 of the next system clock cycle following MEM 10112's assertion of DAVIO.

Having described MEM 10112's interface to IOS 10116, and certain operations which IOS 10116 may request of MEM 10112, certain MEM 10112 operations within the capability of the Interface will be described next. First, operand transfers, for example of numeric data, between MEM 10112 and IOS 10116 may be bit granular with any length from one to sixteen bits. Operand transfers may cross boundaries within a page but may not cross physical page boundaries. As previously described, MIO Bus 10129 and IOM Bus 10130 are capable of transferring 32 bits of data at a time. The least significant 16 bits of these buses, that is bits 16 to 31, will contain right justified data during operand transfers. The contents of the most significant 16 bits of these buses is generally not defined as MEM 10112 generally does not perform fill operations on read operations to IO Port 20910, nor does IOS 10116 fill unused bits during write operations. During a read or write operation, only those data bits indicated by length field in the corresponding memory request are of significance. In all cases, however, parity must be valid on all 32 bits of MIO Bus 10129 and IOM Bus 10130.

Referring to Fig. 204, IOS 10116 includes Data Channels 20410 and 20412 each of which will be described further in a following detailed description of IOS 10116. Data Channels 20410 and 20412 each possess particular characteristics defining certain IO Port 20910 operations. Data Channel 20410 operates to read and write block aligned full and partial blocks. Full blocks have block aligned addresses and lengths of 16 bytes. Partial blocks have byte aligned addresses and lengths of 1 to 15 bytes; a partial block transfer must be within a block, that is not cross block boundaries. A full 4 word block will be transferred between IOS 10116 and MEM 10112 in either case, but only those blocks indicated by length of field in a corresponding MEM 10112 request are of actual significance in a write operation. Non-addressed bytes in such operations may contain any information so long as parity is valid for the entire data transfer. Data Channel 20412 preferably reads or writes 16 bits at a time on double byte boundaries. Such reads and writes are right justified on MIO Bus 10129 and IOM Bus 10130. The most significant 16 bits of these buses may contain any information during such operations so long as parity is valid for the entire 32 bits. Data Channel 20412 operations are similar to IOS 10116 operand read and write operations with double byte aligned addresses and lengths of 16 bits. Finally, instructions, for example controlling IOS 10116 operation, are read from MEM 10112 to IOS 10116 a block at a time. Such operations are identical to a full block data read.

Having described the operating characteristics of IO Port 20910, the operating characteristics of JPO Port 21010 will be described next.

2. JPO Port 21010 Operating Characteristics (Fig. 210)

Referring to Fig. 210, a diagramic representation of JPO Port 21010 is shown. As previously described, JPO Port 21010 is utilized for transfer of operands, for example numeric data, between MEM 10112 and JP 10114. JPO Port 21010 includes a request input (address, length, and operation information) to MIC 20122 from 36 bit PD Bus 10146, a write data input to FIU 20120 from 32 bit JPD Bus 10142, a 32 bit read data output from MC 20116 and FIU 20120 to 32 bit MOD Bus 10144, and bi-directional control inputs and outputs between MIC 20122 and JPMC Bus 10147.

Referring first to JPO Port 21010's read data output to MOD Bus 10144, MOD Bus 10144 is used by JPO Port 21010 to transfer data, for example operands, to JP 10114. MOD Bus 10144 is also utilized internal to MEM 10112 as a bidirectional bus to transfer data between MC 20116 and FIU 20120. In this manner, data may be transferred from MC 20116 to FIU 20120 where certain data format operations are performed on the data before the data is transferred to JP 10114 through MOD Bus 10144. Data may also be used to transfer data from FIU 20120 to MC 20116 after a data format operation is performed in a write operation. Data may also be transferred directly from MC 20116 to JP 10114 through MOD Bus 10144. Internal to MEM 10112, MOD Bus 10144 is a 36 bit bus for concurrent transfer of 32 bits of data, MOD Bus 10144 bits (MOD(0-31)), and 4 bits of odd parity, 1 bit per byte, MOD Bus 10144 bits (MODP(0-3)). External to MEM 10112, MOD Bus 10144 is a 32 bit bus, comprising bits (MOD(0-31)); parity bits are not read to JP 10114.

Data is written into MEM 10112 through JPD Bus 10142 to FIU 20120. As just described, data format operations may then be performed on this data before it is transferred from FIU 20120 to MC 20116 through MOD Bus 10144. In such operations, JPD Bus 10142 operates as a 32 bit bus carrying 32 bits of data, bits (JPD(0-31)), with no parity bits. JO Port 21010 generates parity for JPD Bus 10142 data to be written into MEM 10112 as this data is transferred into MEM 10112.

Memory requests are also transmitted to MEM 10112 from JP 10114 through JPD Bus 10142, which operates in this regard as a 40 bit bus. Each such request includes an address field, a length field, an FIU

EP 0 067 556 B1

field specifying data formatting operations to be performed, operation code field, and a destination code field specifying destination of data read from MEM 10112. Address field includes a 13 bit physical page number field, (JPPN(0—12)), and a 14 bit physical page offset field, (JPPO(0—13)). Length field includes 6 bits of length information, (JLNG(0—5)), and expresses true length of the data item to be written to or read from MEM 10112.

As JPD Bus 10142 and MOD Bus 10144 are each capable of transferring 32 bits of data in a single MEM 10112 read or write cycle, 6 bits of length information are required to express true length. As will be described in a following description, JP 10114 may provide physical page offset and length information directly to MEM 10112, performs logical page number to physical page number translations, and may perform a Protection Mechanism 10230 check on the resulting physical page number. As such, MEM 10112 expects to receive (JPPN(0—12)) later than (JPPO(0—13)) and (JLNG(0—5)). (JPPO(0—13)) and (JLNG(0—5)) should, however, be valid during the system clock cycle in which a JP 10114 memory request is loaded into MEM 10112.

Operation code field provided to MEM 10112 from JP 10114 is a 3 bit code, (JMCM(0—2)) specifying an operation to be formed by MEM 10112. Certain operations which JP 10114 may request of MEM 10112, and their corresponding operation codes, are:

000 = read;
001 = read and set;
010 = write;
011 = error;
100 = error;
101 = read error log and reset;
110 = repair block; and,
111 = flush cache.

Two bit FIU field, (JFIU(0—1)) specifies data manipulation operations to be performed in executing read and write operations. Among the data manipulation operations which may be requested by JP 10114, and their FIU fields, are:

00 = right justified, zero fill;
01 = right justified, sign extend;
10 = left justify, zero fill; and,
11 = left justify, blank fill.

For write operations, JPO Port 21010 may respond only to the most significant bit of FIU field, that is the FIU field bit specifying alignment.

Finally, destination field is a two bit field specifying a JP 10114 destination for data read from MEM 10112. This field is ignored for write operations to MEM 10112. A first bit of destination field, JPMDST, identifies the destination to be FU 10120, and the second field, EBMDST, specifies EU 10120 as the destination.

JPMC Bus 10147 includes dedicated lines for exchange of control signals between JPO Port 21010 and JP 10114. Among these control signals is Load JO Request (LJOR), which is asserted by JP 10114 when JP 10114 wishes to load a request into MEM 10112. LJOR is asserted concurrently with presentation of the memory request to MEM 10112 through PD Bus 10146. JO Port Available (JOPA) is asserted by MEM 10112 when JPO Port 21010 is available to accept a new memory request from JP 10114. If LJOR and JOPA are asserted concurrently, MEM 10112 accepts the memory request from JP 10114 and MEM 10112 drops JOPA to indicate that memory request has been accepted. As previously discussed, MEM 10112 may assert JOPA while a previous request is being executed and the PD Bus 10146 information, that is the memory request previously provided concerning the previous request, is no longer required.

If JP 10114 submits a memory request and JOPA is not asserted by MEM 10112, MEM 10112 does not accept the request and JP 10114 must resubmit that request when JOPA is asserted. Because, as described above, JPPN field of a memory request from JP 10114 may arrive late compared to the other fields of the request, MEM 10112 will delay loading of JPPN field for a particular request until the next system clock cycle after the request was initially submitted. MEM 10112 may also obtain this JPPN field at the same time it is being loaded into the port register by by-passing the port register.

JP 10114 may abort a memory request upon asserting Abort JP Request (ABJR). ABJR will be accepted by MEM 10112 during system clock cycle after accepting memory request from JP 10114 and ABJR will result in cancellation of the requested operation. A single ABJR line is provided for both JPO Port 21010 and JPI Port 21110 because, as described in a following description, MEM 10112 may accept only a single request from JP 10114, to either JPO Port 21010 or to JPI port 21110, during a single system clock cycle.

Upon completion of an operand read operation requested through JPO Port 21010 MEM 10112 may assert either of two data available signals to JP 10114. These signals are data available for FA(DAVFA) and data available for EB(DAVEB). As previously described, a part of each read request from JP 10114 includes a destination field specifying the intended destination of the requested data. As will be described further

below, MEM 10112 tracks such destination information for read requests and returns destination information with a corresponding information in the form of DAVFA and DAVEB. DAVFA indicates a destination in FIU 10120 while DAVEB indicates a destination in EU 10122. MEM 10112 may also assert signal zero filled (ZFILL) specifying whether read data for JPO Port 21010 is zero filled. ZFILL is valid only when DAVEB is asserted.

For JPO Port 21010 write request, the associated write data word should be valid on same system clock cycle as the request, or one system clock cycle later. JP 10114 asserts Load JP Write Data (LJWD) during the system clock cycle when JP 10114 places valid write data on JPD Bus 10142.

As previously discussed, when MEM 10112 detects an error in servicing a JP 10114 request MEM 10112 places a record of this error in MC 20116's Error Log. When an entry is placed in Error Log for either JPO Port 21010 or IO Port 20910, MEM 10112 asserts an interrupt flag signal indicating a valid Error Log entry is present. DP 10118 detects this flag signal and may direct the flag signal to either JP 10114 or IOS 10116, or both. IOS 10116 or JP 10114, as selected by DP 10118, may then read and reset Error Log and reset the flag. The interrupt flag signal is not necessarily directed to the requestor, JP 10114 or IOS 10116, whose request resulted in the error.

If an uncorrectable MEM 10112 error, that is an error in two or more bits of a single data word, is detected in a read operation the incorrect data is read to JP 10114 and an invalid data signal asserted. A signal, Previous MOD Transfer Invalid (PMODI), is asserted by MEM 10112 on the next system clock cycle following either DAVFA or DAVEB. PMODI is not asserted for single bit errors, instead the data is corrected and the corrected data read to JP 10114.

Having described JPO Port 21010's structure, and characteristics, JPI Port 21110 will be described next below.

3. JPI Port 21110 Operating Characteristics (Fig. 211)

Referring to Fig. 211, a diagramic representation of JPI Port 21110 is shown. JPI port 21110 includes an address input from PD Bus 10146 to FIU 20120, a data output to MOD Bus 10144 from MC 20116, and bi-directional control inputs and outputs from MIC 20122 to JPMC Bus 10147. As previously described, a primary function of JPI Port 21110 is the transfer of SOPs and operand NAMEs from MEM 10112 to JP 10114 upon request from JP 10114. JPI Port thereby performs only read operations wherein each read operation is a transfer of a single 32 bit word having a word aligned address.

Referring to JPI Port 21110 input from PD Bus 10146, read requests to MEM 10112 by JP 10114 for SOPs and operand NAMEs each comprise a 21 bit word address. As described above, each JPI Port 21110 read operation is of a single 32 bit word. As such, the five least significant bits of address are ignored by MEM 10112. For the same reason, a JPI Port 21110 request to MEM 10112 does not include a length field, an operation code field, an FIU field, or a destination code field. Length, operation code, and FIU code fields are not required since JPI Port 21110 performs only a single type of operation and destination code field is not required because destination is inherent in a JPI Port 21110 request.

The 32 bit words read from MEM 10112 in response to JPI Port 21110 requests are transferred to JP 10114 through MC 20116's 32 bit output to MOD Bus 10144. As in the case of JPO 21010 read outputs to JP 10114, JPI Port 21110 does not provide parity information to JP 10114.

Control signals exchange between JP 10114 and JPI Port 21110 through JPMC Bus 10147 include Load Ji Request (LJIR) and Ji Port Available (JIPA), which operate in the same manner as discussed with reference to JPO Port 21010. As previously described, JPO Port 21010 and JPI Port 21110 share a single Abort JP Request (ABJR) command. Similarly, JPO Port 21010 and JPI Port 21110 share previous MOD Transfer Invalid (PMODI) from MEM 10112. As described above, a JPI port 21110 request does not include a destination field as destination is implied. MEM 10112 does, however, provide a Data Available Signal (DAVFI) to JP 10114 when a word read from MEM 10112 in response to a JPI Port 21110 request is present on MOD Bus 10144 and valid.

Having described the overall structure and operation of MEM 10112, and the structure and operation of MEM 10112's interface to JP 10114 and IOS 10116, the structure and operation of FIU 20120 MEM 10112 will next be described in further detail.

h. FIU 20120 (Figs. 201, 230, 231)

As previously described, FIU 20120 performs certain data manipulation operations, including those operations necessary to make MEM 10112 bit addressable. Data manipulation operations may be performed on data being written into MEM 10112, for example, JP 10114 through JPD Bus 10142 or from IOS 10116 through IOM Bus 10130. Data manipulations operations may also be performed on data being read from MEM 10112 to JPD 10114 or IOS 10116. In case of data read to JP 10114, MOD Bus 10144 is used both as a MEM 10112 internal bus, in transferring data from MC 20116 to FIU 20120 for manipulation, and to transfer manipulated data from MEM 10112 to JP 10114. In case of data read to IOS 10116, MOD Bus 10144 is again used as MEM 10112 internal bus to read data from MC 20116 to FIU 20120 for subsequent manipulation. The manipulated data is then read from FIU 20120 to IOS 10116 through MIO Bus 10129.

Certain data manipulation operations which may be performed by FIU 20120 have been previously described. In general, a data manipulation operation consists of four distinct operations, and FIU 20120 may manipulate data in any possible manner which may be achieved through performing any combination

of these operations. These four possible operations are selection of data to be manipulated, rotation or shifting of that data, masking of that data, and transfer of that manipulated data to a selected destination. Each FIU 20120 data input will comprise a thirty-two bit data word and, as described above, may be selected from input provided from JPD Bus 10142, MOD Bus 10144, and IOM Bus 10130. In certain cases, an
 5 FIU 20120 data input may comprise two thirty-two bit words, for example, when a cross word operation is performed generating an output comprised of bits from each of two different thirty-two bit words. Rotation or shifting of a selected thirty-two bit data word enables bits within a selected word to be repositioned with respect to word boundaries. When used in conjunction with the masking operation, described momentarily, rotation and shifting may be reiterably performed to transfer any selected bits in a word to
 10 any selected locations in that word. As will be described further below, a masking operation allows any selected bits of a word to be affectively erased, thus leaving only certain other selected bits, or certain selected bits to be forced to predetermined values. A masking operation may be performed, for example, to zero fill or sign extend portions of a thirty-two bit word. In conjunction with a rotation or shifting operation, a masking operation may, for example, select a single bit of a thirty-two bit input word, position that bit in
 15 any selected bit location, and force all other bits of that word to zero. Each output of FIU 20120 is a thirty-two bit data word and, as described above, may be transferred on to MOD Bus 10144 or onto MIO Bus 10129. As will be described below, selection of a particular sequence of the above four operations to be performed on a particular data word is determined by control inputs provided from MIC 20122. These control inputs from MIC 20122 are decoded and executed by microinstruction control logic included within
 20 FIU 20120.

Referring to Fig. 230, a partial block diagram of FIU 20120 is shown. As indicated therein, FIU 20120 includes Data Manipulation Circuitry (DMC) 23010 and FIU Control Circuitry (FIUC) 23012. Data Manipulation Circuitry 23010 in turn includes FIUIO circuitry (FIUIO) 23014, Data Shifter (DS) 23016, Mask Logic (MSK) 23018, and Assembly Register (AR) logic 23020. Data manipulation circuitry 23010 will be
 25 described first followed by FIUC 23012. In describing data manipulation circuitry 23010, FIUIO 23014 will be described first, followed by DS 23016, MSK 23018, and AR 23020, in that order.

Referring to FIUIO 23014, FIUIO 23014 comprises FIU 20120's data input and output circuitry. Job Processor Write Data Register (JWDR) 23022, IO System Write Data Register (IWDR) 23024, and Write Input Data Register (RIDR) 23026 are connected from, respectively, JPD Bus 10142, IOM Bus 10130, and MOD Bus
 30 10144 for receiving data word inputs from, respectively, JP 10114, IOS 10116, and MC 20116. JWDR 23022, IWDR 23024 and RIDR 23026 are each thirty-six bit registers comprised, for example, of SN74S374 registers. Data words transferred into IWDR 23024 and RIDR 23026 are each, as previously described, comprised of a thirty-two data word plus four bits of parity. Data inputs from JP 10114 are, however, as previously described, thirty-two bit data words without parity. Job Processor Parity Generator (JPPG)
 35 23028 associated with JWDR 23022 is connected from JPD Bus 10142 and generates four bits of parity for each data input to JWDR 23022. JWDR 23022's thirty-six bit input thereby comprises thirty-two bits of data, directly from JPD Bus 10142, plus a corresponding four bits of parity from JPPG 23028.

Data words, thirty-two bits of data plus four bits of parity, are transferred into JWDR 23022, IWDR 23024, or RIDR 23026 when, respectively, input enable signals Load JWD (LJWD), Load IWD (LIWD) or Load RID (LRID) are asserted. LJWD is provided from FU 10120 while LIWD and LRID are provided from MIC
 40 20122.

Data words resident in JWDR 23022, IWDR 23024, or RIDR 23026 may be selected and transferred onto FIU 20120's Internal Data (IB) Bus 23030 by output enable signals JWD Enable Output (JWDEO), IWD Enable Output (IWDEO), an RID Enable Output (RIDEO). JWDEO, IWDEO, and RIDEO are provided from
 45 FIUC 23012 described below.

As will be described further below, manipulated data words from DS 23016 or AR 23020 will be transferred onto, respectively, Data Shifter Output (DSO) Bus 23032 or Assembly Register Output (ASYRO) Bus 23034 for subsequent transfer onto MOD Bus 10144 or MIO Bus 10129. Each manipulated data word appearing on DSO Bus 23032 or ASYRO Bus 23034 will be comprised of 32 bits of data plus 4 bits of parity.
 50 Manipulated data words present on DSO Bus 23032 may be transferred onto MOD Bus 10144 or MIO Bus 10129 through, respectively, DSO Bus To MOD Bus Driver Gate (DSMOD) 23036 or BSO Bus To MIO Bus Driver Gate (DSMIO) 23038. Manipulated data words present on ASYRO Bus 23034 may be transferred onto MOD Bus 10144 or MIO Bus 10129 through, respectively, ASYRO Bus To MOD Bus Driver Gate (ASYMOD) 23040 or ASYRO Bus To MIO Bus Driver Gate (ASYMIO) 23042. DSMOD 23036, DSMIO 23038, ASYMOD 23040, and ASYMIO 23042 are each comprised of, for example, SN74S244 drivers. A manipulated data
 55 word on DSO Bus 23032 be transferred through DSMOD 23036 to MOD Bus 10144 when driver gate enable signal Driver Shift To MOD (DRVSHFMOD) to DSMOD 23036 is asserted. Similarly, a manipulated data word on DSO Bus 23032 will be transferred through DSMIO 23038 to MIO Bus 10129 when driver gate enable signal Drive Shift Through MIO Bus (DRVSHFMIO) to DSMIO 23038 is asserted. Manipulated data
 60 words present on ASYRO Bus 23034 may be transferred onto MOD Bus 10144 or MIO Bus 10129 when, respectively, driver gate enable signal Drive Assembly To Mod Bus (DRVASYMOD) to ASYMOD 23040 or Drive Assembly To MIO Bus (DRVASYMIO) to ASYMIO 23042 are asserted. DRVSHFMOD, DRVSHFMIO, DRVASYMOD, and DRVASYMIO are provided, as described below, from FIUC 23012.

Registers IARM 23044 and BARMR 23046, which will be described further in a following description of
 65 DP 10118, are used by DP 10118 to, respectively, write data words onto IB 23030 and to Read data words

EP 0 067 556 B1

from MOD Bus 10144, for example manipulated data words from FIU 20120. Data word written into IARMR 23044 from DP 10118, that is 32 bits of data and 4 bits of parity, will be transferred onto IB Bus 23030 when register enable output signal IARM enable output (IARMEO) from FIUC 23012 is asserted. Similarly, a data word present on MOD Bus 10144, comprising 32 bits of data plus 4 bits of parity, will be written into BARMR 23046 when load enable signal Load BARMR (LDBARMR) to BARMR 23046 is asserted by MIC 20122. A data word written into BARMR 23046 from MOD Bus 10144 may then subsequently be read to DP 10118. IARMR 23044 and BARMR 23046 are similar to JWDR 23022, IWDR 23024, and IRDR 23026 and may be comprised, for example, of SN74S299 registers.

Referring finally to IO Parity Check Circuit (IOPC) 23048, IOPC 23048 is connected from IB Bus 23030 to receive each data word, that is 32 bits of data plus 4 bits of parity, appearing on IB Bus 23030. IOPC 23048 confirms parity and data validity of each data word appearing on IB Bus 23030 and, in particular, determines validity of parity and data of data words written into FIU 20120 from IOS 10116. IOPC 23048 generates output Parity Error (PER), previously discussed, indicating a parity error in data words from IOS 10116.

Referring to DS 23016, DS 23016 includes Byte Nibble Logic (BYNL) 23050, Parity Rotation Logic (PRL) 23052, and Bit Scale Logic (BSL) 23054. BYNL 23050, PRL 23052, and BSL 23054 may respectively be comprised of, for example, 25S10 shifters. BYNL 23050 is connected from IB Bus 23030 for receiving and shifting the 32 data bits of a data word selected and transferred onto IB Bus 23030. PRL 23052 is a 4 bit register similarly connected from IB Bus 23030 to receive and shift the 4 parity bits of a data word selected and transferred onto IB Bus 23030. Outputs of BYNL 23050 and PRL 23052 are both connected onto DSO Bus 23032, thus providing a 36 bit FIU 20120 data word output directly from BYNL 23050 and PRL 23052. BYNL 23050's 32 bit data output is also connected to BSL 23054's input. BSL 23054's 32 bit output is in turn provided to MSK 23018.

As previously described, DS 23016 performs data manipulation operations involving shifting of bits within a data word. In general, data shift operations performed by DS 23016 are rotations wherein data bits are right shifted, with least significant bits of data word being shifted into most significant bit position and most significant bits being translated towards least significant bit positions. DS 23016 rotation operations are performed in two stages. First stage is performed by BYNL 23050 and PRL 23052 and comprises right rotations on a nibble basis (a nibble is defined as 4 bits of data). That is, BYNL 23050 right shifts a data word by an integral number of 4 bit increments. A right rotation on a nibble by nibble basis may, for example, be performed when RM 20722 asserts FLIPHALF previously described. FLIPHALF is asserted for IOS 10116 half word read operations wherein the request data resides in the most significant 16 bits of a data word from MC 20116. BYNL 23050 will perform a right rotation of 4 nibbles to transfer the desired 16 bits of data into the least significant 16 bits of BYNL 23050's output. Resulting BYNL 23050 output, together with PRL 23052's parity bit output would then be transferred through DSO 23032 to MIO Bus 10129. In addition to performing data shifting operations, DS 23016 may transfer a data word, that is the 32 bits of data, directly to MSK 23018 when data manipulation to be performed does not require data shifting, that is shifts of 0 bits may be performed.

Because data bits are shifted by BYNL 23050 on a nibble basis, the relationship between the 32 data bits of a word and the corresponding 4 parity bits may be maintained if parity bits are similarly right rotated by an amount corresponding to right rotation of data bits. This relationship is true if the data word is shifted in multiples of 2 nibbles, that is 8 bits or 1 byte. PRL 23052 right rotates the 4 parity bits of a data word by an amount corresponding to right rotation of the corresponding 32 data bits in BYNL 23050. Right rotated outputs of BYNL 23050 and PRL 23052 therefore comprise a valid data word having 32 bits of data and 4 bits of parity wherein the parity bits are correctly related to the data bits. A right rotated data word output from BYNL 23050 and PRL 23052 may be transferred onto DSO Bus 23032 for subsequent transfer to MOD Bus 10144 or MIO Bus 10129 as described above. DSO 23032 is used as FIU 20120's output data path for byte write operations and "rotate read" operations wherein the required manipulation of a particular data word requires only an integral number of right rotations by bytes. Amount of right rotation of 32 bits of data in BYNL 23050 and 4 bits of parity in PRL 23052 is controlled by input signal shift (SHFT) (0—2) to BYNL 23050 and PRL 23052. As will be described below, SHFT (0—2) is generated, together with SHFT (3—4) controlling BSL 23054, by FIUC 23012. BYNL 23050 and PRL 23052, like BSL 23054 described below, are parallel shift logic chips and entire rotation operation of BYNL 23050 and PRL 23052 or BSL 23054 may be performed in a single clock cycle.

Second stage of rotation is performed by BSL 23054 which, as described above, receives the 32 data bits of a data word from BYNL 23050. BSL 23054 performs right rotation on a bit by bit basis with the shift amount being selectable between 0—3 bits. Therefore, BSL 23054 may rotate bits through nibble boundaries. BYNL 23050 and BSL 23054 therefore comprise a data shifting circuit capable of performing bit-by-bit right rotation by an amount from 1 bit to a full 32 bit right rotation.

Referring now to MSK 23018, MSK 23018 is comprised of 5 32 bit Mask Word Generators (MWG's) 23056 to 23064. MSK 23018 generates a 32 bit output to AR 23020 by selectively combining 32 bit mask word outputs of MWG's 23056 to 23064. Each mask word generated by one of MWG's 23056 to 23064 is effectively comprised a bit by bit combination of a set of enabling bits and a predetermined 32 bit mask word, generated by FIUC 23012 and MIC 20122. MWG's 23058 to 23064 are each comprised of for example, open collector NAND gates for performing these functions, while MWG 23056 is comprised of a PROM.

As just described, outputs of MWG's 23056 to 23064 are all open collector circuits so that any selected combination of mask word outputs from MWG's 23056 to 23064 may be ORed together to comprise the 32 bit output of MSK 23018.

MWG 23056 to MWG 23064 generate, respectively, mask word outputs Locked Bit Word (LBW) (0—31), Sign Extended Word (SEW) (0—31), Data Mask Word (DMW) (0—31), Blank Fill Word (BWF) (0—31), and Assembly Register Output (ARO) (0—31). Referring first to MWG 23064 and ARO (0—31), the contents of Assembly Register (ASYMR) 23066 in AR 23020 are passed through MWG 23064 upon assertion of enabling signal Assembly Output Register (ASYMOR). ARO (0—31) is thereby a copy of the contents of ASYMR 23066 and MWG 23064 allows the contents of ASYMR 23066 to be ORed with the selected combination of LBW (0—31), SEW (0—31), DMW (0—31), or BFW (0—31).

DMW (0—31) from MWG 23060 is generated by ANDing enable Input Data Mask (DMSK) (0—31) with the 32 bit output of DS 23016. DMSK (0—31) is a 32 bit enabling word generated, as described below, by FIUC 23012. FIUC 23012 may generate 4 different DMSK (0—31) patterns. Referring to Fig. 231, the 4 DMSKs (0—31) which may be generated by FIUC 23012 are shown. DMSKA (0—31) is shown in Line A of Fig. 231. In DMSKA (0—31) all bits to the left of but not including a bit designated by Left Bit Address (LBA) and all bits to the right of and not including a bit designated by Right Bit Address (RBA) are 0. All bits between, and including, those bits designated by LBA and RBA are 1's. DMSKB (0—31) is shown in Line B of Fig. 231 and is DMSKA (0—31) inverted. DMSKC (0—31) and DMSKD (0—31) are shown, respectively, in Lines C and D of Fig. 231 and are comprised of, respectively, all 0's or all 1's. As stated above DMSK (0—31) is ANDed with the 32 bit output of DS 23016. As such, DMSKC (0—31) may be used, for example, to inhibit DS 23016's output while DMSKD (0—31) may be used, for example, to pass DS 23016's output to AR 23020. DMSKA (0—31) and DMSKB (0—31) may be used, for example, to gate selected portions of DS 23016's output to AR 23020 where, for example, the selected portions of DS 23016's output may be ORed with other mask word outputs MSK 23018.

Referring next to MWG 23062, MWG 23062 generates BFW (0—31). BFW (0—31) is used in a particular operation wherein 32 bit data words containing 1 to 4 ASCII blanks are required to be generated wherein 1 bit/byte contains a logic one and remaining bits contain logic zeros. In this case, the ASCII blank bytes may contain logic 1's in bit positions 2, 10, 18, and 26.

Referring again to Fig. 231, Line E therein shows 32 bit right mask (RMSK) (0—31) which may be generated by FIUC 23012. In the most general case, RMSK contains zeros in all bit positions to the left of and including a bit position designated by RBA. When used in a blank fill operation, bit positions 2, 10, 18, and 26 may be selected to contain logic 1's depending upon those byte positions containing logic 1's, that is in those bytes containing ASCII blanks; these bytes to the right of RBA are determined by RMSK (0—31). RMSK (0—31) is enabled through MWG 23062 as BWF (0—31) when MWG 23062 is enabled by blank fill (BLNKFILL) provided from FIU 23012.

As described above, MWG's 23058 to 23064 and in particular MWG's 23060 and MWG 23062 are NAND gate operations. Therefore, the outputs of MWGs 23056 through 23064 are active low signals. The inverted output of ASYMR 23066 is used as an output to ASYRO 23034 to invert these outputs to active high.

MWG 23058, generating SEW (0—31), is used in generating sign extended or filled words. In sign extended words, all bit spaces to the left of the most significant bit of a 32 bit data word are filled with the sign bit of the data contained therein, the left most bits of the 32 bit word are filled with 1's or 0's depending on whether that word's sign bit indicates that the data contained therein is a positive or negative number.

Sign Select Multiplexor (SIGNSEL) 23066 is connected to receive the 32 data bits of a word present on IB Bus 23030. Sign Select (SGNSEL) (0—4) to SIGNSEL 23066 is derived from SBA (0—4), that is from SBA Bus 21226 from PRIMUX 20720. As previously described, SBA (0—4) is Starting Bit Address identifying the first or most significant bit of a data word. When a data word contains a signed number, most significant bit contains sign bit of that number. SGNSEL (0—4) input to SIGNSEL 23066 is used as a selection input and, when SIGNSEL is enabled by Sign Extend (SIGNEXT) from FIU 23012, selects the sign bit on IB Bus 23030 and provides that sign bit as an input to MWG 23058.

Sign bit input to MWG 23058 is ANDed with each bit of left hand mask (LMSK) (0—31) from FIUC 23012. Referring again to Fig. 231, LMSK (0—31) is shown on Line F thereof. LMSK (0—31) contains all 0's to the right of and including the bit space identified by LBA and 1's in all bit spaces to the left of that bit space identified by LBA. SEW (0—31) will therefore contain sign bit in all bit spaces to the left of the most significant bit of the data word present on output of MWG 23058. The data word on IB Bus 23030 may then be passed through DS 23016 and subjected to a DMSK operation wherein all bits to the left of the most significant bit are forced to 0. SEW (0—31) and DMW (0—31) outputs of MWG's 23058 and 23060 may then be ORed to provide the desired find extended word output.

LBW (0—31), provided by MWG 23056, is used in locked bit operations wherein the most significant data bit of a data word is in MEM 10112 forced to logic 1. SIGNSEL (0—4) is an address input to MWG 23056 and, as previously described, indicates most significant data bit of a data word present on an IB Bus 23030. MWG 23056 is enabled by input Lock (LOCK) from FIUC 23012 and the resulting LBW (0—31) will contain a single logic 1 in the bit space of the most significant data bit of the data word present on IB Bus 23030. The data word present on IB Bus 23030 may then be passed through DS 23016 and MWG 23060 to be ORed with LBW (0—31) so that that data words most significant data bit is forced to logic 1.

Referring to AR 23020, AR 23020 includes ASYMR 23066, which may be comprised for example of a

SN74S175 registers, and Assembly Register Parity Generator (ASYPG) 23070. As previously described, ASYMR 23066 is connected from MSK 23018 32 bit output. A 32 bit word present on MSK 23018's output will be transferred into ASYMR 23066 when ASYMR 23066 is enabled by Assembly Register Load (ASYMLD) from MIC 20122. The 32 bit word generated through DS 23016 and MSK 23018 will then be present on ASYRO Bus 23034 and may, as described above, then be transferred onto MOD Bus 10144 or MIO Bus 10129. ASYPG 23070 is connected from ASYMR 23066 32 bit output and will generate 4 parity bits for the 32 bit word presently on the 32 data lines of ASYRO Bus 23034. ASYPG 23070's 4 bit parity output is based on the 4 parity bit lines of ASYRO Bus 23034 and accompany the 32 bit data word present thereon.

Having described structure and operation of Data Manipulation Circuitry 23010, FIUC 23012 will be described next below.

Referring again to Fig. 230, FIUC 23012 provides pipelined microinstruction control of FIU 20120. That is, control signals are received from MIC 20122 during a first clock cycle and certain of the control signals are decoded by microinstruction logic to generate further FIUC 23012 control signals. During the second clock cycle, control signals received and generated during first clock cycle are provided to DMC 23010, some of which are further decoded to provide yet other control signals to control operation of FIUC 23012. FIUC 23012 includes Initial Decode Logic (IDL) 23074, Pipeline Registers (PPLR) 23072, Final Decoding Logic (FDL) 23076, and Enable Signal Pipeline Register (ESPR) 23098 with Enable Signal Decode Logic (ESDL) 23099.

IDL 23074 and Control Pipeline Register (CPR) 23084 of PPLR 23072 are connected from control outputs of MIC 20122 to receive control signals therefrom during a first clock cycle as described above. IDL 23074 provides outputs to control pipeline registers Right Bit Address Register (RBAR) 23086, Left Bit Address Register (LBAR) 23088 and Shift Register (SHFR) 23090 of PPLR 23072. CPR 23084 and SHFR 23090 provide control outputs directly to DMC 23010. As described above these outputs control DMC 23010 during second clock cycle.

CPR 23084, RBAR 23086, and LBAR 23088 provide outputs to FDL 23076 during second clock cycle and FDL 23076 in turn provides certain outputs directly to DMC 23010.

ESPR 23098 and ESDL 23099 receive enable and control signals from MIC 20122 and in turn provide enable and control signals to DMC 23010 and certain other portions of MEM 10112 circuitry.

IDL 23074 and FDL 23076 may be comprised, for example, of PROMs. CPR 23084, RBAR 23086, LBAR 23088, SHFR 23090, and ESPR 23098 may be comprised, for example, of SN74S194 registers. ESDL 23099 may be comprised of, for example, compatible decoders, such as logic gates.

Referring first to IDL 23074, IDL 23074 performs an initial decoding of circuitry control signals from MIC 20122 and provides further control signals used by FIUC 23012 in controlling FIU 20120. IDL 23074 is comprised of read-only memory arrays Right Bit Address Decoding Logic (RBADL) 23078, Left Bit Address Decoding Logic (LBADL) 23080, and Shift Amount Decoding Logic (SHFAMTDL) 23082. RBADL 23078 receives, as address inputs, Final Bit Address (FBA) (0—4), Bit Length Number (BLN) (0—4), and Starting Bit Address (SBA) (0—4). FBA, BLN and SBA define, respectively, the final bit, length, and starting bit of a requested data item as previously discussed with reference to PRMUX 20720. RBADL 23078 also receives chip select enable signals Address Translation Chip Select (ATCS) 00, 01, 02, 03, 04, and 15 from MIC 20122 and, in particular, RM 20722. When FIU 20120 is required to execute certain MSK 23018 operations, Inputs FBA (0—4), BLN (0—4), and SBA (0—4), together with an ATCS input, are provided to RBADL 23078 from MIC 20122. RBADL 23078 in turn provides output RBA (Right Bit Address) (0—4), which has been described above with reference to DMSK (0—31) and RMSK (0—31). LBADL 23080 is similar to RBADL 23078 and is provided with inputs BLN (0—4), FBA (0—4), SBA (0—4), and ATCS 06, 07, 08, 09, and 05 from MIC 20122. Again, for certain MSK 23018 operations, LBADL 23080 will generate Left Bit Address (LBA) (0—4), which has been previously discussed above with reference to DMSK (0—31) and LMSK (0—31).

RBA (0—4) and LBA (0—4) are, respectively, transferred to RBAR 23086 and LBAR 23088 at start of second clock cycle by Pipeline Load Enable signal PIPELD provided from MIC 20122. RBAR 23086 and LBAR 23088 in turn respectively provide outputs Register Right Address (RRAD) (0—4) and Register Left Address (RLAD) (0—4) as address inputs to Right Mask Decode Logic (RMSKDL) 23092, Left Mask Decode Logic (LMSKDL) 23094, and FDL 23076 at start of second clock cycle. RRAD (0—4) and RLAD (0—4) correspond respectively to RBA (0—4) and LBA (0—4).

RMSKDL 23092 and LMSKDL 23094 are ROM arrays, having, as just described, RRAD (0—4) and RLAD (0—4) as, respectively, address inputs and Mask Enable (MSKENBL) from CPR 23084 as enable inputs. Together, RMSKDL 23092 and LMSKDL 23094 generate, respectively, RMSK (0—31) and LMSK (0—31) to MSK 23018. RMSK (0—31) and LMSK (0—31) are provided as inputs to Exclusive Or/Exclusive Nor gating (XOR/XNOR) 23096. XOR/XNOR 23096 also receives enable and selection signal Out Mask (OUTMSK) from CPR 23084. RMSK (0—31) and LMSK (0—31) inputs to XOR/XNOR 23096 are used, as selected by OUTMSK from CPR 23084, to generate a selected DMSK (0—31) as shown in Fig. 231. DMSK (0—31) output of XOR/XNOR 23096 is provided, as described above, to MSK 23018.

Referring again to IDL 23074, SHFAMTDL 23082 decodes certain control inputs from MIC 20122 to generate, through SHFR 23090, control inputs SHFT (0—4) and SGNSEL (0—4) to, respectively, DS 23016, SIGNSEL 23068 and MWG 23056. Address inputs to the PROMs comprising SHFAMTDL 23082 include FBA (0—4), SBA (0—4), and FLIPHALF (FLIPHALF) from MIC 20122. FBA (0—4) and SBA (0—4) have been described above. FLIPHALF is a control signal indicating that, as described above, that 16 bits of data

requested by IOS 10116 resides in the upper half of a 32 bit data word and causes those 16 bits to be transferred to the lower half of FIU 20120's output data word onto MIO Bus 10129. MIC 20122 also provides chip enable signals ATCS 10, 11, 12, 13, and 14. Upon receiving these control inputs from MIC 20122, SHFAMTDL 23082 generates an output shift amount (SHFAMT) (0—4) which, together with SBA (0—4) from MIC 20122, is transferred into SHFR 23090 by PIPELD at start of second clock cycle. SHFR 23090 then provides corresponding outputs SHFT (0—4) and SIGNSEL (0—4). As described above, SIGNSEL (0—4) are provided to SIGNSEL 23068 and MWG 23056 and MSK 23018. SHFT (0—4) is provided as SHFT (0—2) and SHFT (3—4) to, respectively, BYNL 23050 and BSL 23054 and DS 23016.

Referring to CPR 23084, as described above certain control signals are provided directly to FIU 20120 circuitry without being decoded by IDL 23074 or FDL 23076. Inputs to CPR 23084 include Sign Extension (SIGNEXT) and Lock (LOCK) indicating, respectively, that FIU 20120 is to perform a sign extension operation through MWG 23058 or a lock bit word operation through MWG 23056. CPR 23084 provides corresponding outputs SIGNEXT and LOCK to MSK 23018 to select these operations. Input Assembly Output Register (ASYMOR) and Blank Fill (BLANKFILL) are passed through CPR 23084 as ASYMOR and BLANKFILL to, respectively, MWG 23064 and MWG 23062 to select the output of ASYMR 23066 as a mask or to indicate that MSK 23018 is to generate a blank filled word through MWG 23062. Inputs OUTMSK and MSKENBL to CPR 23084 are provided, as discussed above, as enable signals OUTMSK and MSKENBL to, respectively, EXORVENOR 23096 and RMSKDL 23092 and LMSKBL 23094 and generating RMSK (0—31), LMSK (0—31), and DMSK (0—31) as described above.

Referring finally to ESPR 23098 and ESDL 23099, ESPR 23098 and PPLR 23072 together comprise a pipeline register and ESDL 23099 decoding logic for providing enable signals to FIU 20120 and other MEM 10112 circuitry. ESPR 23098 receives inputs Drive MOD Bus (DRVMOD) (0—1), Drive MIO Bus (DRVMIO) (0—1), and Enable Register (ENREG) (0—1) from MIC 20122 as previously described. DRVMOD (0—1), DRVMIO (0—1), and ENREG (0—1) are transferred into ESPR 23098 by PIPELD as previously described with reference to PPLR 23072. ESPR 23098 provides corresponding outputs to ESDL 23099, which in turn decodes DRVMOD (0—1), DRVMIO (0—1), and ENREG (0—1) to provide enable signals to FIU 20120 and other MEM 10112 circuitry. Outputs DRVSHFMOD, DRVASYMOD, DRVSHFMIO, and DRVASYMIO are provided to DSMOD 23036, DSMIO 23038, ASYMOD 23040, ASYMIO 23042, and FIUIO 23014 to control transfer of FIU 20120 manipulated data words onto MOD Bus 10144 and MIO Bus 10129. Outputs IARMEO, JWDEO, IWDEO, and RIDEO are provided as output enable signals to IARMR 23044, JWDR 23022, IWDR 23024, and RIDR 23026 to transfer the contents of these registers onto IB Bus 23030 as previously described. Outputs DRVCAMOD, DRVAMIO, DRVBYMOD, and DRVBYMIO are provided to MC 20116 for use in controlling transfer of information onto MOD Bus 10144 and MIO Bus 10129.

Having described the structure and operation of MEM 10112 above, the structure and operation of FU 10120 will be described next below.

B. Fetch Unit 10120 (Figs. 202, 206, 101, 103, 104, 238)

As has been previously described, FU 10120 is an independently operating, microcode controlled machine comprising, together with EU 10122, CS 10110's micromachine for executing user's programs. Principal functions of FU 10120 include: (1) Fetching and interpreting instructions, that is SInS comprising SOPs and Names, and data from MEM 10112 for use by FU 10120 and EU 10122; (2) Organizing and controlling flow and execution of user programs; (3) Initiating EU 10122 operations; (4) Performing arithmetic and logic operations on data; (5) Controlling transfer of data from FU 10120 and EU 10122 to MEM 10112; and, (6) Maintaining certain stack register mechanisms. Among these stack and register mechanisms are Name Cache (NC) 10226, Address Translation Cache (ATC) 10228, Protection Cache (PC) 10234, Architectural Base Registers (ABRs) 10364, Micro-Control Registers (mCRs) 10366, Micro-Stack (MIS) 10368, Monitor Stack (MOS) 10370 of General Register File (GRF) 10354, Micro-Stack Pointer Register Mechanism (MISPR) 10356, and Return Control Word Stack (RCWS) 10358. In addition to maintaining these FU 10120 resident stack and register mechanisms, FU 10120 generates and maintains, in whole or part, certain MEM 10112 resident data structures. Among these MEM 10112 resident data structures are Memory Hash Table (MHT) 10716 and Memory Frame Table (MFT) 10718, Working Set Matrix (WSM) 10720, Virtual Memory Management Request Queue (VMMRQ) 10721, Active Object Table (AOT) 10712, Active Subject Table (AST) 10914, and Virtual processor State Blocks (VPSBs) 10218. In addition, a primary function of FU 10120 is the generation and manipulation of logical descriptors which, as previously described, are the basis of CS 10110's internal addressing structure. As will be described further below, while FU 10120's internal structure and operation allows FU 10120 to execute arithmetic and logic operations, FU 10120's structure includes certain features to expedite generation and manipulation of logical descriptors.

Referring to Fig. 202, a partial block diagram of FU 10120 is shown. To enhance clarity of presentation, certain interconnections within FU 10120, and between FU 10120 and EU 10122 and MEM 10112 are not shown by line connections but, as described further below, are otherwise indicated, such as by common signal names. Major functional elements of FU 10120 include Descriptor Processor (DESP) 20210, MEM 10112 Interface Logic (MEMINT) 20212, and Fetch Unit Control Logic (FUCTL) 20214. DSP 20210 is, in general, an arithmetic and logic unit for generating and manipulating entries for MEM 10112 and FU 10120 resident stack mechanisms and caches, as described above, and, in particular, for generation and manipulation of logical descriptors. In addition, as stated above, DSP 20210 is a general purpose Central

Processor Unit (CPU) capable of performing certain arithmetic and logic functions.

DESP 20210 includes AON Processor (AONP) 20216, Offset Processor (OFFP) 20218, Length Processor (LENP) 20220. OFFP 20218 comprises a general, 32 bit CPU with additional structure to optimize generation and manipulation of offset fields of logical descriptors. AONP 20216 and LENP 20220 comprise, respectively, processors for generation and manipulation of AON and length fields of logical descriptors and may be used in conjunction with OFFP 20218 for execution of certain arithmetic and logical operations. DESP 20210 includes GRF 10354, which in turn include Global Registers (GRs) 10360 and Stack Registers (SRs) 10362. As previously described, GR's 10360 includes ABRs 10364 and mCRs 10366 while SRs 10362 includes MIS 10368 and MOS 10370.

MEMINT 20212 comprises FU 10120's interface to MEM 10112 for providing Physical Descriptors (physical addresses) to MEM 10112 to read SInS and data from and write data to MEM 10112. MEMINT 20212 includes, among other logic circuitry, MC 10226, ATC 10228, and PC 10234.

FUCTL 20214 controls fetching of SInS and data from MEM 10112 and provides sequences of microinstructions for control of FU 10120 and EU 10122 in response to SOPs. FUCTL 20214 provides Name inputs to MC 10226 for subsequent fetching of corresponding data from MEM 10112. FUCTL 20214 includes, in part, MISPR 10356, RCWS 10358, Fetch Unit S-Interpreter Dispatch Table (FUSDT) 11010, and Fetch Unit S-Interpreter Table (FUSITT) 11012.

Having described the overall structure of FU 10120, in particular with regard to previous descriptions in Chapter 1 of this description, DESP 20210, MEMINT 20212, and FUCTL 20214 will be described in further detail below, and in that order.

1. Description Processor 20210 (Figs. 202, 101, 103, 104, 238, 239)

As described above, DESP 20210 comprises a 32 bit CPU for performing all usual arithmetic and logic operations on data. In addition, a primary function of DESP 20210 is generation and manipulation of entries for, for example, Name Tables (NTs) 10350, ATC 10228, and PC 10234, and generation and manipulation of logical descriptors. As previously described, with reference to CS 10110 addressing structure, logical descriptors are logical addresses, or pointers, to data stored in MEM 10112. Logical descriptors are used, for example, as architectural base pointers or microcontrol pointers in ABRs 10364 and mCRs 10366 as shown in Fig. 103, or as linkage and local pointers of Procedure Frames 10412 as shown in Fig. 104. In a further example, logical descriptors generated by DESP 20210 and corresponding to certain operand Names are stored in MC 10226, where they are subsequently accessed by those Names appearing in SInS fetched from MEM 10112 to provide rapid translation between operand Names and corresponding logical descriptors.

As has been previously discussed with reference to CS 10110 addressing structure, logical descriptors provided to ATU 10228, from DESP 20210 or NC 10226, are translated by ATU 10228 to physical descriptors which are actual physical addresses of corresponding data stored in MEM 10112. That data subsequently is provided to JP 10114, and in particular to FU 10120 or EU 10122, through MOD Bus 10144.

As has been previously discussed with reference to MEM 10112, each data read to JP 10114 from MEM 10112 may contain up to 32 bits of information. If a particular data item referenced by a logical descriptor contains more than 32 bits of data, DESP 20210 will, as described further below, generate successive logical descriptors, each logical descriptor referring to 32 bits or less of information, until the entire data item has been read from MEM 10112. In this regard, it should be noted that NC 10226 may contain logical descriptors only for data items of 255 bits or less in length. All requests to MEM 10112 for data items greater than 32 bits in length are generated by DESP 20210. Most of data items operated on by CS 10110 will, however, be 32 bits or less in length so that NC 10226 is capable of handling most operand Names to logical descriptor translations.

As described above, DESP 20210 includes AONP 20216, OFFP 20218, and LENP 20220. OFFP 20218 comprises a general purpose 32 bit CPU with additional logic circuitry for generating and manipulating table and cache entries, as described above, and for generating and manipulating offset fields of AON pointers and logical descriptors. AONP 20216 and LENP 20220 comprise logic circuitry for generating and manipulating, respectively, AON and length fields of AON pointers and logical descriptors. As indicated in Fig. 202, GRF 10354 is vertically divided in three parts. A first part resides in ANOP 20216 and, in addition to random data, contains AON fields of logical descriptors. Second and third parts reside, respectively, in OFFP 20218 and LENP 20220 and, in addition to containing random data, respectively contain offset and length fields of logical descriptors. AON, Offset, and length portions of GRF 10354 residing respectively in AONP 20216, OFFP 20218, and LENP 20220 are designated, respectively, as AONGRF, OFFGRF, and LENGRF. AONGRF portion of GRF 10354 is 28 bits wide while OFFGRF and LENGRF portions of GRF 10354 are 32 bits in width. Although shown as divided vertically into three parts, GRF 10354 is addressed and operates as a unitary structure. That is, a particular address provided to GRF 10354 will address corresponding horizontal segments of each of GRF 10354's three sections residing in AONP 20216, OFFP 20218, and LENP 20220.

a. Offset Processor 20218 Structure

Referring first to OFFP 20218, in addition to being a 32 bit CPU and generating and manipulating table and cache entries and offset fields of AON pointers and logical descriptors, OFFP 20218 is DESP 20210's

primary path for receiving data from and transferring data to MEM 10112. OFFP 20218 includes Offset Input Select Multiplexer (OFFSEL) 20238, OFFGRF 20234, Offset Multiplexer Logic (OFFMUX) 20240, Offset ALU (OFFALU) 20242, and Offset ALU A.Inputs Multiplexer (OFFALUSA) 20244.

5 OFFSEL 20238 has first and second 32 bit data inputs connected from, respectively, MOD Bus 10144 and JPD Bus 10142. OFFSEL 20238 has a third 32 bit data input connected from a first output of OFFALU 20242, a fourth 28 bit data input connected from a first output of AONGRF 20232, and a fifth 32 bit data input connected from OFFSET Bus 20228. OFFSEL 20238 has a first 32 bit output connected to input of OFFGRF 20234 and a second 32 bit data output connected to a first input of OFFMUX 20240. OFFMUX 20240 has second and third 32 bit data inputs connected from, respectively, MOD Bus 10144 and JPD Bus 10142. OFFMUX 10 20240 also has a fourth 5 bit data input connected from Bias Logic (BIAS) 20246 and LENP 20220, described further below, and fifth 16 bit data input connected from NAME Bus 20224. Thirty-two bit data output of OFFGRF 20234 and first 32 bit data output of OFFMUX 20240 are connected to, respectively, first and second data inputs of OFFALUSA 20244. A first 32 bit data output of OFFALUSA 20244 and a second 32 bit data output of OFFMUX 20240 are connected, respectively, to first and second data inputs of OFFALU 15 20242. A second 32 bit data output of OFFALUSA 20244 is connected to OFFSET Bus 20228. A first 32 bit data output of OFFALU 20242 is connected to JPD Bus 10142, to a first input of AON Input Select Multiplexer (AONSEL) 20248 and AONP 20216, and, as described above, to a third input of OFFSEL 20238. A second 32 bit data output of OFFALU 20242 is connected to OFFSET Bus 20228 and third 16 bit output is connected to NAME Bus 20224.

b. AON Processor 20216 Structure

Referring to AONP 20216, a primary function of AONP 20216 is that of containing AON fields of AON pointers and logical descriptors. In addition, those portions of AONGRF 20232 not otherwise occupied by AON pointers and logical descriptors may be used as a 28 bit wide general register area by JP 10114. These 25 portions of AONGRF 20232 may be so used either alone or in conjunction with corresponding portions of OFFGRF 20234 and LENGRF 20236. AONP 20216 includes AONSEL 20248 and AONGRF 20232. As previously described, a first 32 bit data input AONSEL 20248 is connected from a first data output of OFFALU 20242. A second 28 bit data input of AONSEL 20248 is connected from 28 bit output of AONGRF 20232 and from AON Bus 20230. A third 28 bit data input of AONSEL 20248 is connected from logic zero, 30 that is a 28 bit input wherein each input bit is set to logic zero. Twenty-eight bit data output of AONSEL 20248 is connected to data input of AONGRF 20232. As just described, 28 bit data output of AONGRF 20232 is connected to second data input of AONSEL 20248, and is connected to AON Bus 20230.

c. Length Processor 20220 Structure

35 Referring finally to LENP 20220, a primary function of LENP 20220 is the generation manipulation of length fields of AON pointers and physical descriptors. In addition, LENGRF 20236 may be used, in part, either alone or in conjunction with corresponding address spaces of AONGRF 20232 and OFFGRF 20234, as general registers for storage of data. LENP 20220 includes Length Input Select Multiplexer (LENSEL) 20250, LENGRF 20236, BIAS 20246, and Length ALU (LENALU) 20252. LENSEL 20250 has first and second data 40 inputs connected from, respectively, LENGTH Bus 20226 and OFFSET Bus 20228. LENGTH Bus 20226 is eight data bits, zero filled while OFFSET Bus 20228 is 32 data bits. LENSEL 20250 has a third 32 bit data input connected from data output of LENALU 20252. Thirty-two bit data output of LENSEL 20250 is connected to data input of LENGRF 20236 and to a first data input of BIAS 20246. Second and third 32 bit data inputs of BIAS 20246 are connected from, respectively, Constant (C) and Literal (L) outputs of FUSITT 45 11012 as will be described further below. Thirty-two bits data output of LENGRF 20236 is connected to JPD Bus 10142, to Write Length Input (WLI) input of NC 10226, and to a first input of LENALU 20252. Five bit output of BIAS 20246 is connected to a second input of LENALU 20252, to LENGTH Bus 20226, and, as previously described, to a fourth input of OFFMUX 20240. Thirty-two bit output of LENALU 20252 is connected, as stated above, to third input of LENSEL 20250.

50 Having described the overall operation and the structure of DESP 20210, operation of DESP 20210 will be described next below in further detail.

d. Descriptor Processor 20210 Operation

a.a. Offset Selector 20238

55 Referring to OFFP 20218, GRF 10354 includes GR's 10360 and SR's 10362. GR's 10360 in turn contain ABR's 10364, mCR's 10366, and a set of general registers. SR's 10362 include MIS 10368 and MOS 10370. GRF 10354 is vertically divided into three parts. AONGRF 20232 is 28 bits wide and resides in AONP 20216, LENGRF 10354 is 32 bits wide and resides in LENP 20220, and OFFGRF 20234 is 32 bits wide and resides in OFFP 20218. AONGRF 20232, OFFGRF 20234, and LENGRF 20236 may be comprised of Fairchild 93422s.

60 In addition to storing offset fields of AON pointers and logical descriptors, those portions of OFFGRF 20234 not reserved for ABR's 10365, mCR's 10366, and SR's 10362 may be used as general registers, alone or in conjunction with corresponding portions AONGRF 20232 and LENGRF 20236, when OFFP 20218 is being utilized as a general purpose, 32 bit CPU. OFFGRF 20234 as will be described further below, is addressed in parallel with AONGRF 20232 and LENGRF 20236 by address inputs provided from FUCTL 65 20214.

OFFSEL 20238 is a multiplexer, comprised for example of SN74S244s and SN74S257s, for selecting data inputs to be written into selected address locations of OFFGRF 20234. OFFSEL 20238's first data input is from MOD Bus 10144 and is the primary path for data transfer between MEM 10112 and DESP 20210. As previously described, each data read from MEM 10112 to JP 10114 is a single 32 bit word where between one and 32 bits may contain actual data. If a data item to be read from MEM 10112 contains more than 32 bits of data, successive read operations are performed until the entire data item has been transferred.

OFFSEL 20238's second data input is from JPD Bus 10142. As will be described further below, JPD Bus 10142 is a data transfer path by which data outputs of FU 10120 and EU 10122 are written into MEM 10112. OFFSEL 20238's input of JPD Bus 10142 thereby provides a wrap around path by which data present at outputs of FU 10120 or EU 10122 may be transferred back into DESP 20210 for further use. For example, as previously stated a first output of OFFALU 20242 is connected to JPD Bus 10142, thereby allowing data output of OFFP 20218 to be returned to OFFP 20218 for further processing, or to be transferred to AONP 20216 or LENP 20220 as will be described further below. In addition, output of LENGRF 20236 is also connected to JPD Bus 10142 so that length fields of AON pointers or physical descriptors, or data, may be read from LENGRF 20236 to OFFP 20218. This path may be used, for example, when LENGRF 20236 is being used as a general purpose register for storing data or intermediate results of arithmetic or logical operations.

OFFSEL 20238's third input is provided from OFFALU 20242's output. This data path thereby provides a wrap around path whereby offset fields or data residing in OFFGRF 20234 may be operated on and returned to OFFGRF 20234, either in the same address location as originally read from or to a different address location. OFFP 20218 wrap around path from OFFALU 20242's output to OFFSEL 20238's third input, and thus to OFFGRF 20234, may be utilized, for example, in reading from MEM 10112 a data item containing more than 32 bits of data. As previously described, each read operation from MEM 10112 to JP 10114 is of a 32 bit word wherein between one and 32 bits may contain actual data. Transfer of a data word containing more than 32 bits is accomplished by performing a succession of read operations from MEM 10112 to JP 10114. For example, if a requested data item contains 70 bits of data, that data item will be transferred in three consecutive read operations. First and second read operations will each transfer 32 bits of data, and final read operation will transfer the remaining 6 bits of data. To read a data item of greater than 32 bits from MEM 10112 therefore, DESP 20210 must generate a sequence of logical descriptors, each defining a successive 32 bit segment of that data item. Final logical descriptor of the sequence may define a segment of less than 32 bits, for example, six bits as in the example just stated. In each successive physical descriptor, offset field must be incremented by value of length field of the preceding physical descriptor to define starting addresses of successive data items segments to be transferred. Length field of succeeding physical descriptors will, in general, remain constant at 32 bits except for final transfer which may be less than 32 bits. Offset field will thereby usually be incremented by 32 bits at each transfer until final transfer. OFFP 20218's wrap around data path from OFFALU 20242's output to their input of OFFSEL 20238 may, as stated above, be utilized in such sequential data transfer operations to write incremented or decremented offset field of a current physical descriptor back into OFFGRF 20234 to be offset field of a next succeeding physical descriptor.

In a further example, OFFP 20218's wrap around path from OFFALU 20242's output to third input of OFFSEL 20238 may be used in resolving Entries in Name Tables 10350, that is Name resolutions. In Name resolutions, as previously described, offset fields of AON pointers, for example Linkage Pointers 10416, are successively added and subtracted to provide a final AON pointer to a desired data item.

OFFSEL 20238's fourth input, from AONGRF 20232's output, may be used to transfer data or AON fields from AONGRF 20232 to OFFGRF 20234 or OFFMUX 20240. This data path may be used, for example, when OFFP 20218 is used to generate AON fields of AON pointers or physical descriptors or when performing Name evaluations.

Finally, OFFSEL 20238's fifth data input from OFFSET Bus 20228 allows offset fields on OFFSET Bus 20228 to be written into OFFGRF 20234 or transferred into OFFMUX 20240. This data path may be used, for example, to copy offset fields to OFFGRF 20234 when JP 10114 is performing a Name evaluation.

Referring now to OFFMUX 20240, OFFMUX 20240 includes logic circuitry for manipulating individual bits of 32 bit words. OFFMUX 20240 may be used, for example, to increment and decrement offset fields by length fields when performing string transfers, and to generate entries for, for example, MHT 10716 and MFT 10718. OFFMUX 20240 may also be used to aid in generating and manipulating AON, OFFSET, and LENGTH fields of physical descriptors and AON pointers.

b.b. Offset Multiplexer 20240 Detailed Structure (Fig. 238)

Referring to Fig. 238, a more detailed, partial block diagram of OFFMUX 20240 is shown. OFFMUX 20240 includes Offset Multiplexer Input Selector (OFFMUXIS) 23810, which for example may be comprised of SN74S373s and SN74S244s and Offset Multiplexer Register (OFFMUXR) 23812, which for example may be comprised of SN74S374s. OFFMUX 20240 also includes Field Extraction Circuit (FEXT) 23814, which may for example be comprised of SN74S257s, and Offset Multiplexer Field Selector (OFFMUXFS) 23816, which for example may be comprised of SN74S257s and SN74S374s. Finally, OFFMUX 20240 includes Offset Scaler (OFFSCALE) 23818, which may for example be comprised of AMD 25S10s, Offset Inter-element Spacing Encoder (OFFIESENC) 23820, which may for example be comprised of Fairchild 93427s

and Offset Multiplexer Output Selector (OFFMUXOS) 23822, which may for example be comprised of AMD 255s, Fairchild 93427s, and SN74S244s.

Referring first to OFFMUX 20240's connections to other portions of OFFP 20218, OFFMUX 20240's first data input, from OFFSEL 20238, is connected to a first input of OFFMUXIS 23810. OFFMUX 20240's second input, from MOD Bus 10144, is connected to a second input of OFFMUXIS 23810. OFFMUX 20240's third input, from JPD Bus 10142, is connected to a first input of OFFMUXFS 23816 while OFFMUX 20240's fourth input, from BIAS 20246, is connected to a first input of OFFMUXOS 23822. OFFMUX 20240's fifth input, from NAME Bus 20224, is connected to a second input of OFFMUXFS 23816. OFFMUX 20240's first output, to OFFALUSA 20244, is connected from output of OFFMUXR 23812 while OFFMUX 20240's second output, to OFFALU 20242, is connected from output of OFFMUXOS 23822.

Referring to OFFMUX 20240's internal connections, 32 bit output of OFFMUXIS 23810 is connected to input OFFMUXR 23812 and 32 bit output of OFFMUXR 23812 is connected, as described above, as first output of OFFMUX 20240, and as a third input of OFFMUXFS 23816. Thirty-two bit output of OFFMUXR 23812 is also connected to input of FEXT 23814. OFFMUXFS 23816's first, second and third inputs are connected as described above. A fourth input of OFFMUXFS 23816 is a 32 bit input wherein 31 bits are set to logic zero and 1 bit to logic 1. A fifth input is a 32 bit input wherein 31 bits are set to logic 1 and 1 to logic 0. A sixth input of OFFMUXFS 23816 is a 32 bit literal (L) input provided from FUSITT 11012 and is a 32 bit binary number comprising a part of a microinstruction FUCTL 20214, described below. OFFMUXFS 23816's seventh and eighth input are connected from FEXT 23814. Input 7 comprises FIU and TYPE fields of Name Table Entries which have been read into OFFMUXR 23812. Input 8 is a general purpose input conveying bits extracted from a 32 bit word captured in OFFMUXR 23812. As indicated in Fig. 238, OFFMUXFS 23816's first, third, fourth, fifth, and sixth inputs are each 32 bit inputs which are divided to provide two 16 bit inputs each. That is, each of these 32 bit inputs is divided into a first input comprising bit 0 to 15 of that 32 bit input, and a second input comprising bits 16 to 31.

Thirty-two bit output of OFFMUXFS 23816 is connected to inputs of OFFSCALE 23818 and OFFIESENC 23820. As indicated in Fig. 238, Field Select Output (FSO) of OFFMUXFS 23816 is a 32 bit word divided into a first word including 0 to 15 and a second word including bits 16 to 31. Output FSO of OFFMUXFS 23816, as will be described further below, thereby reflects the divided structure of OFFMUXFS 23816's first, third, fourth, fifth, and sixth inputs.

Logical functions performed by OFFMUXFS 23816 in generating output FSO, and which will be described in further detail in following descriptions, include:

- (1) Passing the contents of OFFMUXR 23812 directly through OFFMUXFS 23816;
- (2) Passing a 32 bit word on JPD Bus 10142 directly through OFFMUXFS 23816;
- (3) Passing a literal value comprising a part of a microinstruction from FUCTL 20214 directly through OFFMUXFS 23816;
- (4) Forcing FSO to be literal values 0000 0000;
- (5) Forcing FSO to be literal value 0000 001;
- (6) Extracting Name Table Entry fields;
- (7) Accepting a 32 bit word from OFFMUXR 23812 or JPD Bus 10142, or 32 bits of a microinstruction from FUCTL 20214, and passing the lower 16 bits while forcing the upper 16 bits to logic 0;
- (8) Accepting a 32 bit word from OFFMUXR 23812 or JPD Bus 10142, or 32 bits of microinstruction from FUCTL 20214, and passing the higher 16 bits while forcing the lower 16 bits to logic 0;
- (9) Accepting a 32 bit word from OFFMUXR 23812, or JPD Bus 10142, or Name Bus 20224, or 32 bits of a microinstruction from FUCTL 20214, and passing the lower 16 bits while sign extending bit 16 to the upper 16 bits; and,
- (10) Accepting a 32 bit word from Name Bus 20224 and passing the lowest 8 bits while sign extending bit 24 to the highest 24 bits.

Thirty-two bit output of OFFSCALE 23818 and 3 bit output of OFFIESENC 23820 are connected, respectively, to second and third inputs of OFFMUXOS 23822. OFFMUXOS 23822's first input is, as described above, OFFMUX 20240's fourth input and is connected from output BIAS 20246. Finally, OFFMUXOS 23822's 32 bit output, OFFMUX (0—31) is OFFMUX 20240's second output and as previously described as connected to a second input of OFFALU 20242.

c.c. Offset Multiplexer 20240 Detailed Operation

a.a.a. Internal Operation

Having described the structure of OFFMUX 20240 as shown in Fig. 238, operation of OFFMUX 20240 will be described below. Internal operation of OFFMUX 20240, as shown in Fig. 238, will be described first, followed by description of OFFMUX 20240's operation with regard to DESP 20210.

Referring first to OFFMUXR 23812, OFFMUXR 23812 is a 32 bit register receiving either a 32 bit word from MOD Bus 10144, MOD (0—31), or a 32 bit word received from OFFSEL 20238, OFFSEL (0—31), and is selected by OFFMUXIS 23810. OFFMUXR 23812 in turn provides those selected 32 bit words from MOD Bus 10144 or OFFSEL 20238 as OFFMUX 20240's first data output to OFFALUSA 20244, as FEXT 23814's input, and as OFFMUXFS 23816's third input. OFFMUXR 23812's 32 bit output to OFFMUXFS 23816 is provided as two parallel 16 bit words designated as OFFMUXR output (OFFMUXRO) (0—15) and (16—31). As described above, OFFMUXFS 23816's output to OFFALUSA 20244 from OFFMUXR 23812 may be right shifted 16

places and the highest 16 bits zero filled.

FEXT 23814 receives OFFMUXRO (0—15) and (16—31) from OFFMUXR 23812 and extracts certain fields from those 16 bit words. In particular, FEXT 23814 extracts FIU and TYPE fields from NT 10350 Entries which have been transferred into OFFMUXR 23812. FEXT 23814 may then provide those FIU and TYPE fields as OFFMUXFS 23816's seventh input. FEXT 23814 may, selectively, extract certain other fields from 32 bit words residing in OFFMUXR 23812 and provide those fields as OFFMUXFS 23816's eighth input.

OFFMUXFS 23816 operates as a multiplexer to select certain fields from OFFMUXFS 23816's eight inputs and provide corresponding 32 bit output words, Field Select Output (FSO), comprised of those selected fields from OFFMUXFS 23816's inputs. As previously described, FSO is comprised of 2, parallel 16 bit words, FSO (0—15) and FSO (16—31). Correspondingly, OFFMUX 20240's third input, from JPD Bus 10142, is a 32 bit input presented as two 16 bit words, JPD (0—15) and JPD (16—31). Similarly, OFFMUXFS 23816's fourth, fifth, and sixth inputs are each presented as 32 bit words comprised of 2, parallel 16 bit words, respectively, "0" (0—15) and (16—31), "1" (0—15) and (16—31), and L (0—15) and (16—31). OFFMUXFS 23816's second input, from NAME Bus 20224, is presented as a single 16 bit word, NAME (16—31), while OFFMUXFS 23816's inputs from FEXT 23814 are each less than 16 bits in width. OFFMUXFS 23816 may, for a single 32 bit output word, select FSO (0—15) to contain one of corresponding 16 bit inputs JPD (0—15), "0" (0—15), "1" (0—15), or L (0—15). Similarly, FSO (16—31) of that 32 bit output word may be selected to contain one of NAME (16—31), JPD (16—31), O (16—31), 1 (16—31), L (16—31), or inputs 7 and 8 from FEXT 23814. OFFMUXFS 23816 therefore allows 32 bit words, comprised of two 16 bit fields, to be generated from selected portions of OFFMUXFS 23816's inputs.

OFFMUXFS 23816 32 bit output is provided as inputs to OFFSCALE 23818 and OFFIESENC 23820. Referring first to OFFIESENC 23820, OFFIESENC 23820 is used, in particular, in resolving, or evaluating, NT 10350 Entries (NTEs) referring to arrays of data words. As indicated in Fig. 108, word D of an NTE contains certain information relating to inter-element spacing (IES) of data words of an array. Word D of an NTE may be read from MEM 10112 to MOD Bus 10144 and through OFFMUX 20240 to input of OFFIESENC 23820. OFFIESENC 23820 then examines word D's IES field to determine whether inter-element spacing of that array is a binary multiple, that is 1, 2, 4, 8, 16, 32, or 64 bits. In particular, OFFIESENC 23820 determines whether 32 bit word D contains logic zeros in the most significant 25 bits and a single logic one in the least significant 7 bits. If inter-element spacing is such a binary multiple, starting addresses of data words of that array may be determined by left shifting of index (IES) to obtain offset fields of physical addresses of words in the array and a slower and more complex multiplication operation is not required. In such cases, OFFIESENC generates a first output, IES Encodeable (IESENC) to FUCTL 20214 to indicate that inter-element spacing may be determined by simple left shifting. OFFIESENC 23820 then generates encoded output, Encoded IES (ENCIES), to OFFMUXOS 23822. ENCIES is then a coded value specifying the amount of left shift necessary to translate index (IES) value into offsets of words in that array. As indicated in Fig. 238, ENCIES is OFFMUXOS 23822's third input.

OFFSCALE 23818 is a left shift shift network with zero fill of least significant bits, as bits are left shifted. Amount of shift by OFFSCALE 23818 is selectable between zero and 7 bits. Thirty-two bit words transferred into OFFSCALE 23818 from OFFSCALE 23818 from OFFMUXFS 23816 may therefore be left shifted, bit by bit, to selectively reposition bits within that 32 bit input word. In conjunction with OFFMUXFS 23816, and a wrap around connection provided by OFFALU 20242's output to OFFSEL 20238, OFFSCALE 23818 may be used to generate and manipulate, for example, entries for MHT 10716, MFT 10718, AOT 10712, and AST 10914, and other CS 10110 data structures.

OFFMUXOS 23822 is a multiplexer having first, second, and third inputs from, respectively, BIAS 20246, OFFSCALE 23818, OFFIESENC 23820. OFFMUXOS 23822 may select any one of these inputs as OFFMUX 20240's second output, OFFMUX (0—31). As previously described, OFFMUX 20240's second output is connected to a second input of OFFALU 20242.

Having described internal of OFFMUX 20240, operation of OFFMUX 20240 with regard to overall operation of DESP 20210 will be described next below.

b.b.b. Operation Relative to Descriptor Processor 20210

OFFMUX 20240's first input, from OFFSEL 20238, allows inputs to OFFSEL to be transferred through OFFMUXIS 23810 and into OFFMUXR 23812. This input allows OFFMUXR 23812 to be loaded, under control of FUCTL 20214 microinstructions, with any input of OFFSEL 20238. In a particular example, OFFALU 20242's output may be fed back through OFFSEL 20238's third input and OFFMUX 20240's first input to allow OFFMUX 20240 and OFFALU 20242 to perform reiterative operations on a single 32 bit word.

OFFMUX 20240's second input, from MOD Bus 10144, allows OFFMUXR 23812 to be loaded directly from MOD Bus 10144. For example, NTEs from a currently active procedure may be loaded into OFFMUXR 23812 to be operated upon as described above. In addition, OFFMUX 20240's second input may be used in conjunction with OFFSEL 20238's first input, from MOD Bus 10144, as parallel input paths to OFFP 20218. These parallel input paths allow pipelining of OFFP 20218 operations by allowing OFFSEL 20238 and OFFGRF 20234 to operate independently from OFFMUX 20240. For example, FU 10120 may initiate a read operation from MEM 10112 to OFFMUXR 23812 during a first microinstruction. The data so requested will appear on MOD Bus 10144 during a second microinstruction and may be loaded into OFFMUXR 23812 through OFFMUX 20240's second input from MOD Bus 10144. Concurrently, FU 10120 may initiate, at start

of second microinstruction, an independent operation to be performed by OFFSEL 20238 and OFFGRF 20234, for example loading output of OFFALU 20242 into OFFGRF 20234. Therefore, by providing an independent path into OFFMUX 20240 from MOD Bus 10144, OFFSEL 20238 is free to perform other, concurrent data transfer operations while a data transfer from MOD Bus 10144 to OFFMUX 20240 is being performed.

OFFMUX 20240's third input, from JPD Bus 10142, is a general purpose data transfer path. For example, data from LENGRF 20236 or OFFALU 20242 may be transferred into OFFMUX 20240 through JPD Bus 10142 and OFFMUX 20240's third input.

OFFMUX 20240's fourth input is connected from BIAS 20246 and primarily used during string transfers as described above. That is, length fields of physical descriptors generated for a string transfer may be transferred into OFFMUX 20240 through OFFMUX 20240's fourth input to increment or decrement, offset fields of those physical descriptors in OFFALU 20242.

OFFMUX 20240's fifth input is connected from NAME Bus 20224. As will be described further below, Names are provided to NC 10226 by FUCTL 20214 to call, from MC 10226, logical descriptors corresponding to Names appearing on MOD Bus 10144 as part of sequences of SInS.

As each Name is presented to NC 10226, that Name is transferred into and captured in Name Trap (NT) 20254. Upon occurrence of an NC 10226 miss, that is NC 10226 does not contain an entry corresponding to a particular Name, that Name is subsequently transferred from NT 20254 to OFFMUX 20240 through NAME Bus 20224 and OFFMUX 20240's fifth input. That Name, which is previously described as an 8, 12, or 16 bit binary number, may then be scaled, that is multiplied by a NTE size. That scaled Name may then be added to Name Table Pointer (NTP) from mCRs 10366 to obtain the address of a corresponding NTE in an NT 10350. In addition, a Name resulting in a NC 10226 miss, or a page fault in the corresponding NT 10350, or requiring a sequence of Name resolves, may be transferred into OFFGRF 20234 from OFFMUX 20240, through OFFALU 20242 and OFFSEL 20238 third input. That Name may subsequently be read, or restored, from OFFGRF 20234 as required.

Referring now to outputs of OFFMUX 20240, OFFMUX 20240's first output, from OFFMUXR 23812, allows contents of OFFMUXR 23812 to be transferred to first input of OFFALU 20242 through OFFALUSA 20244. OFFMUX 20240's second output, from OFFMUXOS 23822, is provided directly to second input of OFFALU 20242. OFFALU 20242 may be concurrently provided with a first input from OFFMUXR 23812 and a second input, for example a manipulated offset field, from OFFMUXOS 23822.

Referring to OFFALUSA 20244, OFFALUSA 20244 is a multiplexer. OFFALUSA 20244 may select either output of OFFGRF 20234 or first output of OFFMUX 20240 to be either first input of OFFALU 20242 or to be OFFP 20218's output to OFFSET Bus 20228. For example, an offset field from OFFGRF 20234 may be read to OFFSET Bus 20228 to comprise offset field of a current logical descriptor, and concurrently read into OFFALU 20242 to be incremented or decremented to generate offset field of a subsequent logical descriptor in a string transfer.

OFFALU 20242 is a general purpose, 32 bit arithmetic and logic unit capable of performing all usual ALU operations. For example, OFFALU 20242 may add, subtract, increment, or decrement offset fields of logical descriptors. In addition, OFFALU 20242 may serve as a transfer path for data, that is OFFALU 20242 may transfer input data to OFFALU 20242's outputs without operating upon that data. OFFALU 20242's first output, as described above, is connected to JPD Bus 10142, to third input of OFFSEL 20238, and to first input of AONSEL 20248. Data transferred or manipulated by OFFALU 20242 may therefore be transferred on to JPD Bus 10142, or wrapped around into OFFP 20218 through OFFSEL 20238 for subsequent or reiterative operations. OFFALU 20242's output to AONSEL 20248 may be used, for example, to load AON fields of AON pointers or physical descriptors generated by OFFP 20218 into AONGRF 20232. In addition, this data path allows FU 10120 to utilize AONGRF 20232 as, for example, a buffer or temporary memory space for intermediate or final results of FU 10120 operations.

OFFALU 20242's output to OFFSET Bus 20228 allows logical descriptor offset fields to be transferred onto OFFSET Bus 20228 directly from OFFALU 20242. For example, a logical descriptor offset field may be generated by OFFALU 20242 during a first clock cycle, and transferred immediately onto OFFSET Bus 20228 during a second clock cycle.

OFFALU 20242's third output is to NAME Bus 20224. As will be described further below, NAME Bus 20224 is address input (ADR) to NC 10226. OFFALU 20242's output to NAME Bus 20224 thereby allows OFFP 20218 to generate or provide addresses, that is Names, to NC 10226.

Having described operation of OFFP 20218, operation of LENP 20220 will be described next below.

e. Length Processor 20220 (Fig. 239)

Referring to Fig. 202, a primary function of LENP 20220 is generation and manipulation of logical descriptor length fields, including length fields of logical descriptors generated in string transfers. LENP 20220 includes LENGRF 20236, LENSEL 20250, BIAS 20246, and LENALU 20252. LENGRF 20236 may be comprised, for example, of Fairchild 93422s. LENSEL 20250 may be comprised of, for example, SN74S257s, SN74S157s, and SN74S244s, and LENALU 20252 may be comprised of, for example, SN74S381s.

As previously described, LENGRF 20236 is a 32 bit wide vertical section of GRF 10354. LENGRF 20236 operates in parallel with OFFGRF 20234 and AONGRF 20232 and contains, in part, length fields of logical descriptors. In addition, also as previously described, LENGRF 20236 may contain data.

LENSEL 20250 is a multiplexer having three inputs and providing outputs to LENGRF 20236 and first input of BIAS 20246. LENSEL 20250's first input is from Length Bus 20226 and may be used to write physical descriptor or length fields from LENGTH Bus 20226 into LENGRF 20236 or into BIAS 20246. Such length fields may be written from LENGTH Bus 20226 to LENGRF 20236, for example, during Name evaluation or resolve operations. LENSEL 20250's second input is from OFFSET Bus 20228. LENSEL 20250's second input may be used, for example, to load length fields generated by OFFP 20218 into LENGRF 20236. In addition, data operated upon by OFFP 20218 may be read into LENGRF 20236 for storage through LENSEL 20250's second input.

LENSEL 20250's third input is from output of LENALU 20252 and is a wrap around path to return output of LENALU 20252 to LENGRF 20236. LENSEL 20250's third input may, for example, be used during string transfers when length fields of a particular logical descriptor is incremented or decremented by LENALU 20252 and returned to LENGRF 20236. This data path may also, for example, be used in moving a 32 bit word from one location in LENGRF 20236 to another location in LENGRF 20236. As stated above, LENSEL 20250's output is also provided to first input BIAS and allows data appearing at first, second, or third inputs of LENSEL 20250 to be provided to first input of BIAS 20246.

BIAS 20246, as will be described in further detail below, generates logical descriptor length fields during string transfers. As described above, no more than 32 bits of data may be read from MEM 10112 during a single read operation. A data item of greater than 32 bits in length must therefore be transferred in a series, or string, of read operations, each read operation transferring 32 bits or less of data. String transfer logical descriptor length fields generated BIAS 20246 are provided to LENGTH Bus 20226, to LENALU 20252 second input, and to OFFMUX 20240's fourth input, as previously described. These string transfer logical descriptor length fields, referred to as bias fields are provided to LENGTH Bus 20226 by BIAS 20246 to be length fields of the series of logical descriptors generated by DESP 20210 to execute a string transfer. These bias fields are provided to fourth input OFFMUX 20240 to increment or decrement offset fields of those logical descriptors, as previously described. These bias fields are provided to second input of LENALU 20252, during string transfers, to correspondingly decrement the length field of a data item being read to MEM 10112 in a string transfer. BIAS 20246 will be described in greater detail below, after LENALU 20252 is first briefly described.

a.a. Length ALU 20252

LENALU 20252 is a general purpose, 32 bit arithmetic and logic unit capable of executing all customary arithmetic and logic operations. In particular, during a string transfer of a particular data item LENALU 20252 receives that data item's length field from LENGRF 20236 and successive bias fields from BIAS 20246. LENALU 20252 then decrements that logical descriptor's current length field to generate length field to be used during next read operation of the string transfer, and transfers new length field back into LENGRF 20236 through LENSEL 20250's third input.

b.b. BIAS 20246 (Fig. 239)

Referring to Fig. 239, a partial block diagram of BIAS 20246 is shown. BIAS 20246 includes Bias Memory (BIASM) 23910, Length Detector (LDET) 23912, Next Zero Detector (NXTZRO) 23914, and Select Bias (SBIAS) 23916. Input of LDET 23912 is first input of BIAS 20246 and connected from output of LENSEL 20250. Output of LDET 23912 is connected to data input of BIASM 23910, and data output of BIASM 23910 is connected to input of NXTZRO 23914. Output of NXTZRO 23914 is connected to a first input of SBIAS 23916. A second input of SBIAS 23916 is BIAS 20246's second input, LB, and is connected from an output of FUCTL 20214. A third input of SBIAS 23916 is BIAS 20246's third input, L, and is connected from yet another output of FUCTL 20214. Output of SBIAS 23916 is output of BIAS 20246 and, as described above, is connected to LENGTH Bus 20226, to a second input of LENALU 20252, and to fourth input of OFFMUX 20240.

BIASM 23910 is a 7 bit wide random access memory having a length equal to, and operating and addressed in parallel with, SR's 10362 of GRF 10354. BIASM 23910 has an address location corresponding to each address location of SR's 10362 and is addressed concurrently with those address locations in SR's 10362. BIASM 23910 may be comprised, for example, of AMD 27SO3As.

BIASM 23910 contains a bias value of each logical descriptor residing in SR's 10362. As described above, a bias value is a number representing number of bits to be read from MEM 10112 in a particular read operation when a data item having a corresponding logical descriptor, with a length field stored LENGRF 20236, is to be read from MEM 10112. Initially, bias values are written into BIASM 23910, in a manner described below, when their corresponding length fields are written into LENGRF 20236. If a particular data item has a length of less than 32 bits, that data item's initial bias value will represent that data item's actual length. For example, if a data item has a length of 24 bits the associated bias value will be a 6 bit binary number representing 24. That data item's length field in LENGRF 20236 will similarly contain a length value of 24. If a particular item has a length of greater than 32 bits for example, 70 bits as described in a previous example, that data item must be read from MEM 10112 in a string transfer operation. As previously described, a string transfer is a series of read operations transferring 32 bits at a time from MEM 10112, with a final transfer of 32 bits or less completing transfer of that data item. Such a data item's initial length field entry in LENGRF 20236 will contain, using the same example as previously described, a value of 70.

EP 0 067 556 B1

That data item's initial bias entry written into a corresponding address space of BIASM 23910 will contain a bias value of 32. That initial bias value of 32 indicates that at least the first read operation required to transfer that data item from MEM 10112 will transfer 32 bits of data.

When a data item having a length of less than 32 bits, for example 24 bits, is to be read from MEM 10112, that data item's bias value of 24 is read from BIASM 23910 and provided to LENGTH Bus 20226 as length field of logical descriptor for that read operation. Concurrently, that bias value of 24 is subtracted from that data item's length field read from LENGRF 20236. Subtracting that bias value from that length value will yield a result of zero, indicating that no further read operations are required to complete transfer of that data item.

If a data item having, for example, a length of 70 bits is to be read from MEM 10112, that data item's initial bias value of 32 is read from BIASM 23910 to LENGTH Bus 20226 as length field of first logical descriptor of a string transfer. Concurrently, that data item's initial length field is read from LENGRF 20236. That data item's initial bias value, 32, is subtracted from that data item's initial length value, 70, and LENALU 20252. The result of that subtraction operation is the remaining length of data item to be transferred in one or more subsequent read operations. In this example, subtracting initial bias value from initial length value indicates that 38 bits of that data item remain to be transferred. LENALU 20252's output representing results of this subtraction, for example 38, are transferred to LENSEL 20250's third input to LENGRF 20236 and written into address location from which that data item's initial length value was read. This new length field entry then represents remaining length of that data item. Concurrently, LDET 23912 examines that residual length value being written into LENGRF 20236 to determine whether remaining length of that data item is greater than 32 bits or is equal to or less than 32 bits. If remaining length is greater than 32 bits, LDET 23912 generates a next bias value of 32, which is written into BIASM 23910 and same address location that held initial bias value. If remaining data item length is less than 32 bits, LDET 23912 generates a 6 bit binary number representing actual remaining length of data item to be transferred. Actual remaining length would then, again, be written into BIASM 23910 address location originally containing initial bias value. These operations are also performed by LDET 23912 in examining initial length field and generating a corresponding initial bias value. These read operations are continued as described above until LDET 23912 detects that remaining length field is 32 bits or less, and thus that transfer of that data item will be completed upon next read operation. When this event is detected, LDET 23912 generates a seventh bit input into BIASM 23910, which is written into BIASM 23910 together with last bias value of that string transfer, indicating that remaining length will be zero after next read operation. When a final bias value is read from BIASM 23910 at start of next read operation of that string transfer, that seventh bit is examined by NXTZRO 23914 which subsequently generates a test condition output, Last Read (LSTRD) to FUCTL 20214. FUCTL 20214 may then terminate execution of that string transfer after that last read operation, if the transfer has been successfully completed.

As previously described, the basic unit of length of a data item in CS 10110 is 32 bits. Accordingly, data items of 32 or fewer bits may be transferred directly while data items of more than 32 bits require a string transfer. In addition, transfer of a data item through a string transfer requires tracking of the transferred length, and remaining length to be transferred, of both the data item itself and the data storage space of the location the data item is being transferred to. As such, BIAS 20246 will store, and operate with, in the manner described above, length and bias fields of the logical descriptors of both the data item and the location the data item is being transferred to. FUCTL 20214 will receive an LSTRD test condition if bias field of source descriptor becomes zero before or concurrently with that of the destination, that is a completed transfer, or if bias field of destination becomes zero before that of the source, and may provide an appropriate microcode control response. It should be noted that if source bias field becomes zero before that of the destination, the remainder of the location that this data item is being transferred to will be filled and padded with zeros. If the data item is larger than the destination storage capacity, the destination location will be filled to capacity and FUCTL 20214 notified to initiate appropriate action.

In addition to allowing data item transfers which are insensitive to data item length, BIAS 20246 allows string transfers to be accomplished by short, tight microcode loops which are insensitive to data item length. A string transfer, for example, from location A to location B is encoded as:

- (1) Fetch from A, subtract length from bias A, and update offset and length of a; and,
- (2) Store to B, subtract length from bias B, and branch to (1) if length of B does not go to zero or fall through (end transfer) if length of B goes to zero. Source (A) length need not be tested as the microcode loop continues until length of B goes to zero; as described above, B will be filled and padded with zeros if length of A is less than length of B, or B will be filled and the string transfer ended if length of A is greater than or equal to length of B.

LDET 23912 and NXTZRO 23914 thereby allow FUCTL 20214 to automatically initiate a string transfer upon occurrence of a single microinstruction from FUCTL 20214 initiating a read operation by DESP 20210. That microinstruction initiating a read operation will then be automatically repeated until LSTRD to FUCTL 20214 from NXTZRO 23914 indicates that the string transfer is completed. LDET 23912 and NXTZRO 23914 may, respectively, be comprised for example of S74S260s, SN74S133s, SN74S51s, SN74S00s, SN74S00s, SN74S04s, SN74S02s, and SN74S32s.

Referring finally to SBIAS 23916, SBIAS 23916 is a multiplexer comprised, for example, of SN74S288s, SN74S374s, and SN74S244s. SBIAS 23916, under microinstruction control from FUCTL 20214, selects BIAS

EP 0 067 556 B1

20246's output to be one of a bias value from BIASM 23910, L8, or L. SBIAS 23916's first input, from BIASM 23910, has been described above. SBIAS 23916's second input, L8, is provided from FUCTL 20214 and is 8 bits of a microinstruction provided from FUSITT 11012. SBIAS 23916's second input allows microcode selection of bias values to be used in manipulation of length and offset fields of logical descriptors by LENALU 20252 and OFFALU 20242, and for generating entries to MC 10226. SBIAS 23916's third input, L, is similarly provided from FUCTL 20214 and is a decoded length value derived from portions of microinstructions in FUSITT 11012. These microcode length values represent certain commonly occurring data item lengths, for example length of 1, 2, 4, 8, 16, 32, and 64 bits. An L input representing a length of 8 bits, may be used for example in reading data from MEM 10112 on a byte by byte basis.

Having described operation of LERP 20220, operation of AONP 20216 will be described next below.

f. AON Processor 20216

a.a. AONGRF 20232

As described above, AONP 20216 includes AONSEL 20248 and AONGRF 20232. AONGRF 20232 is a 28 bit wide vertical section of GRF 10354 and stores AON fields of AON pointers and logical descriptors. AONSEL 20248 is a multiplexer for selecting inputs to be written into AONGRF 20232. AONSEL 20248 may be comprised, for example of SN74S257s. AONGRF 20232 may be comprised of, for example, Fairchild 93422s.

As previously described, AONGRF 20232's output is connected onto AON Bus 20230 to allow AON fields of AON pointers and logical descriptors to be transferred onto AON Bus 20230 from AONGRF 20232. AONGRF 20232's output, together with a bi-directional input from AON Bus 20230, is connected to a second input of AONSEL 20248 and to a fourth input of AONSEL 20238. This data path allows AON fields, either from AONGRF 20232 or from AON Bus 20230, to be written into AONGRF 20232 or AONGRF 20234, or provided as an input to OFFMUX 20240.

b.b. AON Selector 20248

AONSEL 20248's first input is, as previously described, connected from output of OFFALU 20242 and is used, for example, to allow AON fields generated or manipulated by OFFP 20218 to be written into AONGRF 20232. AONSEL 20248's third input is a 28 bit word wherein each bit is a logical zero. AONSEL 20248's third input allows AON fields of all zeros to be written into AONGRF 20232. An AON field of all zeros is reserved to indicate that corresponding entries in OFFGRF 20234 and LENGRF 20236 are neither AON pointers nor logical descriptors. AON fields of all zeros are thereby reserved to indicate that corresponding entries in OFFGRF 20234 and LENGRF 20236 contain data.

In summary, as described above, DESP 20210 includes AONP 20216, OFFP 20218, and LERP 20220. OFFP 20218 contains a vertical section of GRF 10354, OFFGRF 20234, for storing offset fields of AON pointers and logical descriptors, and for containing data to be operated upon by DESP 20210. OFFP 20218 is principal path for transfer of data from MEM 10112 to JP 10114 and is a general purpose 32 bit arithmetic and logic unit for performing all usual arithmetic and logic operations. In addition, OFFP 20218 includes circuitry, for example OFFMUX 20240, for generation and manipulation of AON, OFFSET, and LENGTH fields of logical descriptors and AON pointers. OFFP 20218 may also generate and manipulate entries for, for example, NC 10226, ATU 10228, PC 10234, AOT 10712, MHT 10716, MFT 10718, and other data and address structures residing in MEM 10112. LERP 20220 includes a vertical section of GRF 10354, LENGRF 20236, for storing length fields of logical descriptors, and for storing data. LERP 20220 further includes BIAS 20246, used in conjunction with LENGRF 20236 and LENALU 20252, for providing length fields of logical descriptors for MEM 10112 read operations and in particular automatically performing string transfers. AONP 20216 similarly includes a vertical section of GRF 10354, AONGRF 20232. A primary function AONGRF 20232 is storing and providing AON fields of AON pointers and logical descriptors.

Having described structure and operation of DESP 20210, structure and operation of Memory Interface (MEMINT) 20212 will be described next below.

2. Memory Interface 20212 (Figs. 106, 240)

MEMINT 20212 comprises FU 10120's interface to MEM 10112. As described above, MEMINT 20212 includes Name Cache (NC) 10226, Address Translation Unit (ATU) 10228, and Protection Cache (PC) 10234, all of which have been previously briefly described. MEMINT 20212 further includes Descriptor Trap (DEST) 20256 and Data Trap (DAT) 20258. Functions performed by MEMINT 20212 includes (1) resolution of Names to logical descriptors, by NC 10226; (2) translation of logical descriptors to physical descriptors, by ATU 10228; and (3) confirmation of access writes to objects, by PC 10234.

As shown in Fig. 202, NC 10226 address input (ADR) is connected from NAME Bus 20224. NC 10226 Write Length Field Input (WL) is connected from LENGRF 20236's output. NC 10226's Write Offset Field Input (WO) and Write AON Field Input (WA) are connected, respectively, from OFFSET Bus 20228 and AON Bus 20230. NC 10226 Read AON Field (RA), Read Offset Field (RO), and Read Length Field (RL) outputs are connected, respectively, to AON Bus 20230, OFFSET Bus 20228, and LENGTH Bus 20226.

DEST 20256's bi-directional AON (AON), Offset (OFF), and Length (LEN) ports are connected by bi-directional buses to and from, respectively, AON Bus 20230, OFFSET Bus 20228, and LENGTH Bus 20226.

PC 10234 has AON (AON) and Offset (OFF) inputs connected from, respectively, AON Bus 20230 and

EP 0 067 556 B1

OFFSET Bus 20228. PC 10234 has a Write Entry (WEN) input connected from JPD Bus 10142. ATU 10228 has AON (AON), Offset (OFF), and Length (LEN) inputs connected from, respectively, AON Bus 20230, OFFSET Bus 20228, and LENGTH Bus 20226. ATU 10228's output is connected to physical Descriptor (PD) Bus 10146.

5 Finally, DAT 20258 has a bi-directional port connected to and from JPD Bus 10142.

a.a. Description Trap 20256 and Data Trap 20258

Referring first to DST 20256 and DAT 20258, DST 20256 is a register for receiving and capturing logical descriptors appearing on AON Bus 20230, OFFSET Bus 20228, and Length Bus 20226. Similarly, DAT 20258 is a register for receiving and capturing data words appearing on JPD Bus 10142. DST 20256 and DAT 20258 may subsequently return captured logical descriptors or data words to, respectively, AON Bus 20230, OFFSET Bus 20228, and LENGTH Bus 20226, and to JPD Bus 10142.

As previously described, many CS 10110 operations, in particular MEM 10112 and JP 10114 operations, are pipelined. That is, operations are overlapped with certain sets within two or more operations being executed concurrently. For example, FU 10120 may submit read request to MEM 10112 and, while MEM 10112 is accepting and servicing that request, submit a second read request. DEST 20256 and DAT 20258 assist in execution of overlapping operations by providing a temporary record of these operations. For example, a part of a read or write request to MEM 10112 by FU 10120 is a logical descriptor provided to ATU 10228. If, for example the first read request just referred to results in a ATU 10228 cache miss or a protection violation, the logical descriptor of that first request must be recovered for subsequent action by CS 10110 as previously described. That logical descriptor will have been captured and stored in DEST 20256 and thus is immediately available, so that DESP 20210 is not required to regenerate that descriptor. DAT 20258 serves a similar purpose with regard to data being written into MEM 10112 from JP 10114. That is, DAT 20258 receives and captures a copy of each 32 bit word transferred onto JPD Bus 10142 by JP 10114. In event of MEM 10112 being unable to accept a write request, that data may be subsequently reprovided from DAT 20258.

b.b. Name Cache 10226, Address Translation Unit 10228, and Protection Cache 10234 (Fig. 106)

Referring to NC 10226, ATU 10228, and PC 10234, these elements of MEMINT 20212 are primarily cache mechanisms to enhance the speed of FU 10120's interface to MEM 10112, and consequently of CS 10110's operation. As described previously, NC 10226 contains a set of logical descriptors corresponding to certain operand names currently appearing in a process being executed by CS 10110. NC 10226 thus effectively provides high speed resolution of certain operand names to corresponding logical descriptors. As described above with reference to string transfers, NC 10226 will generally contain logical descriptors only for data items of less than 256 bits length. NC 10226 read and write addresses are names provided on NAME Bus 20224. Name read and write addresses may be provided from DESP 20210, and in particular from OFFP 20218 as previously described, or from FUCTL 20214 as will be described in a following description of FUCTL 20214. Logical descriptors comprising NC 10226 entries, each entry comprising an AON field, an Offset field, a Length field, are written into NC 10226 through NC 10226 inputs WA, WO, and WL from, respectively, AON Bus 20230, OFFSET Bus 20228, and LENGRF 20236's output. Logical descriptors read from NC 10226 in response to names provided to NC 10226 ADR input are provided to AON Bus 20230, OFFSET Bus 20228, and LENGTH Bus 20226 from, respectively, NC 10226 outputs RA, RO, and RL.

ATU 10228 is similarly a cache mechanism for providing high speed translation of logical to physical descriptors. In general, ATU 10228 will contain, at any given time, a set of logical to physical page number mappings for MEM 10112 read and write requests which are currently being made, or anticipated to be made, to MEM 10112 by JP 10114. As previously described, each physical descriptor is comprised of a Frame Number (FN) field, and Offset Within Frame (O) fields, and a Length field. As discussed with reference to string transfers, a physical descriptor length field, as in a logical descriptor length field, specify a data item of less than or equal to 32 bits length. Referring to Fig. 106C, as previously discussed a logical descriptor comprised of a 14 bit AON field, a 32 bit Offset field, and Length field, wherein 32 bit logical descriptor Offset field is divided into a 18 bit Page Number (P) field and a 14 bit Offset within Page (O) field. In translating a logical into a physical descriptor, logical descriptor Length and O fields are used directly, as respectively, physical descriptor length and O fields. Logical descriptor AON and P fields are translated into physical descriptor FN field. Because no actual translation is required, ATU 10228 may provide logical descriptor L field and corresponding O field directly, that is without delay, to MEM 10112 as corresponding physical descriptor O and Length fields. ATU 10228 cache entries are thereby comprised of physical descriptor FN fields corresponding to AON and P fields of those logical descriptors for which ATU 10228 has corresponding entries. Because physical descriptor FN fields are provided from ATU 10228's cache, rather than directly as in physical descriptor O and Length fields, a physical descriptor's FN field will be provided to MEM 10112, for example, one clock cycle later than that physical descriptors O and Length fields, as has been previously discussed.

Referring to Fig. 202, physical descriptor FN fields to be written into ATU 10228 are, in general, generated by DESP 20210. FN fields to be written into ATU 10228 are provided to ATU 10228 Data Input (DI) through JPD Bus 10142. ATU 10228 read and write addresses are comprised of AON and P fields of logical

descriptors and are provided to ATU 10228's AON and OFF inputs from, respectively, AON Bus 20230 and OFFSET Bus 20228. ATU 10228 read and write addresses may be provided from DESP 20210 or, as described further below, from FUCTL 20214. ATU 10228 FN outputs, together with O and Length fields comprising a physical descriptor, are provided to PD Bus 10146.

5 PC 10234 is a cache mechanism for confirming active procedure's access rights to objects identified by logical descriptors generated as a part of JP 10114 read or write requests to MEM 10112. As previously described access rights to objects are arbitrated on the basis of subjects. A subject has been defined as a particular combination of a principal, process, and domain. A principal, process, and domain are each identified by corresponding UIDs. Each subject having access rights to an object is assigned an Active
10 Subject Number (ASN) as described in a previous description of CS 10110's Protection Mechanism. The ASN of a subject currently active in CS 10110 is stored in ASN Register 10916 in FU 10120. Access rights of a currently active subject to currently active objects are read from those objects Access Control Lists (ACL) 10918 and stored in PC 10234. If the current ASN changes, PC 10234 is flushed of corresponding access right entries and new entries, corresponding to the new ASN, are written into PC 10234. The access rights
15 of a particular current ASN to a particular object may be determined by indexing, or addressing, PC 10234 with the AON identifying that object. Addresses to write entries into or read entries from PC 10234 are provided to PC 10234 AON input from AON Bus 20230. Entries to be written into PC 10234 are provided to PC 10234's WEN input from JPD Bus 10142. PC 10234 is also provided with inputs, not shown in Fig. 202 for purposes of clarity, from FUCTL 20214 indicating the current operation to be performed by JP 10114 with
20 respect to an object being presently addressed by FU 10120. Whenever FU 10120 submits a read or write request concerning a particular object to MEM 10112, AON field of that request is provided as an address to PC 10234. Access rights of the current active subject to that object are read from corresponding PC 10234 entry and compared to FUCTL 20214 inputs indicating the particular operation to be performed by JP 10114 with respect to that object. The operation to be performed by JP 10114 is then compared to that active
25 subject's access rights to that object and PC 10234 provides an output indicating whether that active subject possesses the rights required to perform the intended operation. Indexing of PC 10234 and comparison of access rights to intended operation is performed concurrently with translation of the memory request logical descriptor to a corresponding physical descriptor by ATU 10228. If PC 10234 indicates that that active subject has the required access rights, the intended operation is executed by JP 10114. If PC 10234
30 indicates that that active subject does not have the required access rights, PC 10234 indicates that a protection mechanism violation has occurred and Interrupts execution of the intended operation.

c.c Structure and Operation of a Generalized Cache and NC 10226 (Fig. 240)

35 Having described overall structure and operation of NC 10226, ATU 10228, and PC 10234, structure and operation of these caches will be described in further detail below. Structure and operation of NC 10226, ATU 10228, and PC 10234 are similar, except that NC 10226 is a four-way set associative cache, ATU 10228 is a three-way set associative cache and PC 10234 is a two-way set associative cache.

40 As such, the structure and operation of NC 10226, ATU 10228, and PC 10234 will be described by reference to and description of a generalized cache similar but not necessarily identical to each of NC 10226, ATU 10228, and PC 10234. Reference will be made to NC 10226 in the description of a generalized cache next below, both to further illustrate structure and operation of the generalized cache, and to describe differences between the generalized cache and NC 10226. ATU 10228 and PC 10234 will then be described by description of differences between ATU 10228 and PC 10234 and the generalized cache.

45 Referring to Fig. 240, a partial block diagram of a generalized four-way, set associative cache is shown. Tag Store (TS) 24010 is comprised of Tag Store A (TSA) 24012, Tag Store B (TSB) 24014, Tag Store C (TSC) 24016, and Tag Store D (TSD) 24018. Each of the cache's sets, represented by TSA 24012 to TSD 24018, may contain, for example as in NC 10226, up to 16 entries, so that TSA 24012 to TSD 24018 are each 16 words long.

50 Address inputs to a cache are divided into a tag field and an index field. Tag fields are stored in the cache's tag store and indexed, that is addressed to be read or written from or to tag store by index field of the address. A tag read from tag store in response to index field of an address is then compared to tag field of that address to indicate whether the cache contains an entry corresponding to that address, that is, whether a cache hit occurs. In, for example, NC 10226, a Name syllable may be comprised of an 8, 12, or 16 bit binary word, as previously described. The four least significant bits of these words, or Names, comprise
55 NC 10226's index field while the remaining 4, 8, or 12 most significant bits comprise NC 10226's tag field. TSA 24012 to TDS 24018 may each, therefore, be 12 entry wide memories to store the 12 bit tag fields of 16 bit names. Index (IND) or address inputs of TSA 24012 to TSD 24018, would in NC 10226, be connected from four least significant bits of NAME Bus 20224 while Tag Inputs (TAGI) of TSA 24012 to TSD 24018 would be connected from the 12 most significant bits of NAME Bus 20224.

60 As described above, tag outputs of TS 24010 are compared to tag fields of addresses presented to the cache to determine whether the cache contains an entry corresponding to that address. Using NC 10226 as an example 12 bit Tag Outputs (TAGOs) of TSA 24012 to TSD 24018 are connected to first inputs of Tag Store Comparators (TSC) 24019, respectively to inputs of Tag Store Comparitor A (TSCA) 24020, Tag Store Comparitor B (TSCB) 24022, Tag Store Comparitor D (TSCD) 24024, and Tag Store Comparitor E (TSCE) 24026. Second 12 bit inputs of TSCA 24020 to TSCE 24026 may be connected from the 12 most significant
65

bits of NAME Bus 20224 to receive tag fields of NC 10226 addresses. TAS 24020 to TSCE 24026 compare tag field of an address to tag outputs read from TSA 24012 to TSE 24018 in response to index field of that address, and provide four bit outputs indicating which, if any, of the possible 16 entries and their associated tag store correspond to that address tag field. TSCA 24020 to TSCE 24026 may be comprised, for example, of Fairchild 93S46s.

Four bit outputs of TSCA 24012 to TSCE 24026 are connected in the generalized cache to inputs of Tag Store Pipeline Registers (TSPR) 24027; respectively to inputs of Tag Store Pipeline Register A (TSPRA) 24028, Tag Store Pipeline Register B (TSPRB) 24030, Tag Store Pipeline Register C (TSPRC) 24032, and Tag Store Pipeline Register D (TSPRD) 24034. ATU 10228 and PC 10234 is pipelined with a single cache access operation being executed in two clock cycles. During first clock cycle tag store is addressed and tags store therein compared to tag field of address to provide indication of whether a cache hit has occurred, that is whether cache contains an entry corresponding to a particular address. During second clock cycle, as will be described below, a detected cache hit is encoded to obtain access to a corresponding entry in cache data store. Pipeline operation over two clock cycles is provided by cache pipeline registers which include, in part, TSPRA 24028 to TSPRD 24034. NC 10226 is not pipelined and does not include TSPRA 24028 to TSPRD 24034. In NC 10226, outputs of TSCA 24012 to TSCD 24024 are connected directly to inputs of TSHEA 24036 to TSHED 24042, described below.

Outputs of TSPRA 24028 to TSPRD 24034 are connected to inputs of Tag Store Hit Encoders (TSHE) 24035, respectively to Tag Store Hit Encoder A (TSHEA) 24036, Tag Store Hit Encoder B (TSHEB) 24038, Tag Store Hit Encoder C (TSHEC) 24040, and Tag Store Hit Encoder D (TSHED) 24042. TSHEA 24036 to TSHED 24042 encode, respectively, bit inputs from TSPRA 24028 to TSPRD 24034 to provide single bit outputs indicating which, if any, set of the cache's four sets includes an entry corresponding to the address input.

Single bit outputs of TSHEA 24036 to TSHED 24042 are connected to inputs of Hit Encoder (HE) 24044. HE 24044 encodes single bit inputs from TSHEA 24036 to TSHED 24042 to provide two sets of outputs. First outputs of HE 24044 are provided to Cache Usage Store (CUS) 24046 and indicate in which of the cache's four sets, corresponding to TSA 24012 to TSD 24018, a cache hit has occurred. As described previously with reference to MC 20116, and will be described further below, CUS 24046 is a memory containing information for tracking usage of cache entries. That is, CUS 24046 contains entries indicating whether, for a particular Index, Set A, Set B, Set C or Set D of the cache's four sets has been most recently used and which has been least recently used. CUS 24046 entries regarding Sets A, B, C, and D are stored in, respectively, memories CUSA 24088, CUSB 24090, CUSC 24092, and CUSD 24094. Second output of HE 24044, as described further below, is connected to selection input of Data Store Selection Multiplexer (DSSMUX) 24048 to select an output from Data Store (DS) 24050 to be provided as output of the cache when a cache hit occurs.

Referring to DS 24050, as previously described a cache's data store contains the information, or entries, stored in that cache. For example, each entry in NC 10226's DS 24050 is a logical descriptor comprised of an AON, and Offset, and Length. A cache's data store parallels, in structure and organization, that cache's tag store and entries therein are identified and located through that cache's tag store and associated tag store comparison and decoding logic. In NC 10226, for example, for each Name having an entry in NC 10226 there will be an entry, the tag field of that name, stored in TS 24010 and a corresponding entry, a logical descriptor corresponding to that Name, in DS 24050. As described above, NC 10226 is a four-way, set associative cache so that TS 24010 and DS 24050 will each contain four sets of data. Each set was previously described as containing up to 16 entries. DS 24050 is therefore comprised of four 16 word memories. Each memory is 65 bits wide, accommodating 28 bits of AON, 32 bits of offset, and 5 bits of length. These four component data store memories of DS 24050 are indicated in Fig. 240 as Data Store A (DSA) 24052, Data Store B (DSB) 24054, Data Store C (DSC) 24056, and Data Store D (DSD) 24058. DSA 24052, DSB 24054, DSC 24056 and DSD 24058 correspond, respectively, in structure, contents, and operation to TSA 24012, TSB 24014, TSC 24016 and TSD 24018.

Data Inputs (DIs) of DSA 24052 to DSD 24058 are, in NC 10226 for example, connected from AON Bus 20230, OFFSET Bus 20228, LENGTH Bus 20226 and comprise inputs WA, WO, WL respectively of NC 10226. DSA 24052 to DSD 24058 DIs are, in NC 10226 as previously described, utilized in writing NC 10226 entries into DSA 24052 to DSD 24058. Address inputs of DSA 24052 to DSD 24058 are connected from address outputs of Address Pipeline Register (ADRPR) 24060. As will be described momentarily, except during cache flush operations, DSA 24052 to DSD 24058 address inputs are comprised of the same index fields of cache addresses as are provided as address inputs to TS 24010, but are delayed by one clock cycle and ADRPR 24060 for pipelining purposes. As described above, NC 10226 is not pipelined and does not have the one clock cycle delay. An address input to the cache will thereby result in corresponding entries, selected by index field of that address, being read from TSA 24012 to TSD 24018 and DSA 24052 to DSD 24058. The four outputs of DSA 24052 to DSD 24058 selected by a particular index field of a particular address are provided as inputs to DSSMUX 24048. DSSMUX 24048 is concurrently provided with selection control input from HE 24044. As previously described, this selection input to DSSMUX 24048 is derived from TS 24010 tag entries and indicates which of DSA 24052 to DSD 24058 entries corresponds to an address provided to the cache. In response to that selection control input, DSSMUX 24048 selects one of DS 24050's four logical descriptor outputs as the cache's output corresponding to that address. DSSMUX 24048's output is then provided, through Buffer Driver (BD) 24062 as the cache's output, for example in NC 10226 to AON Bus 20230, OFFSET Bus 20228, and LENGTH Bus 20226.

Referring to ADRMUX 24062, ADRMUX 24062 selects one of two sources to provide address inputs to DS 24050, that is to index to DS 24050. As described above, a first ADRMUX 24062 input is comprised of the cache's address index fields and, for example in NC 10226, is connected from the four least significant bits of NAME Bus 20224. During cache flush operations, DS 24050 address inputs are provided from Flush Counter (FLUSHCTR) 24066, which in the example is a four bit counter. During cache flush operations, FLUSHCTR 24066 generates sequential bit addresses which are used to sequentially address DSA 24052 to DSD 24058. Selection between ADRMUX 24062 first and second inputs, respectively the address index fields and from FLUSHCTR 24066, is controlled by Address Multiplexer Select (ADMUXS) from FUCTL 20214.

Validity Store (VALS) 24068 and Dirty Store (DIRTYS) 24070 are memories operating in parallel with, and addressed in parallel with TS 24010. VALS 24068 contains entries indicating validity of corresponding TS 24010 and DS 24050 entries. That is, VALS 24068 entries indicate whether corresponding entries have been written into corresponding locations in TS 24010 and DS 24050. In the example, VALS 24068 may thereby be a 16 word by 4 bit wide memory. Each bit of a VALS 24068 word indicates validity of a corresponding location in TSA 24012 and DSA 24052, TSB 24014 and DSB 24054, TSC 24016 and DSC 24056, and TSD 24018 and DSD 24058. DIRTYS 24070 similarly indicates whether corresponding entries in corresponding locations of TS 24010 and DS 24050 have been written over, or modified. Again, DIRTYS 24070 will be a sixteen word by four bit wide memory.

Address inputs of VALS 24068 and DIRTYS 24070 are, for example in NC 10226, connected from least significant bits of NAME Bus 20224 and are thus addressed by index fields of NC 10226 addresses in parallel with TS 24010. Outputs of VALS 24068 are provided to TSCA 24020 to TSEE 24026 to inhibit outputs of TSCA 24020 through TSCE 24026 upon occurrence of an invalid concurrence between a TS 24010 entry and a NC 10226 address input. Similar outputs of DIRTYS 24070 are provided to FUCTL 20214 for use in cache flush operations to indicate which NC 10226 entries are dirty and must be written back into an MT 10350 rather than discarded.

Outputs of VALS 24068 and DIRTYS 24070 are also connected, respectively, to inputs of Validity Pipeline Register (VALPR) 24072 and Dirty Pipeline Register (DIRTYPR) 24074. VALPR 24072 and DIRTYPR 24074 are pipeline registers similar to TSPRA 24028 to TSPRD 24034 and are provided for timing purposes as will be described momentarily. Outputs of VALPR 24072 and DIRTYPR 24074 are connected to inputs of, respectively, Validity Write Logic (VWL) 24076 and Dirty Write Logic (DWL) 24078. As described above, NC 10226 is not a pipelined cache and does not include VALPR 24072 and DIRTYPR 24074; outputs of VALS 24068 and DIRTYS 24070 are connected directly to inputs of VWL 24076 and DWL 24078. Outputs of VWL 24076 and DWL 24078 are connected, respectively, to data inputs of VALS 24068 and DIRTYS 24070. Upon occurrence of a write operation to TS 24010 and DS 24050, that is writing in or modifying a cache entry, corresponding validity and dirty word entries are read from VALS 24068 and DIRTYS 24070 by index field of the caches input address. Outputs to VALS 24068 DIRTYS 24070 are received and stored in, respectively, VALPR 24072 and DIRTYPR 24074. At start of next clock cycle, validity and dirty words in VALPR 24072 and DIRTYPR 24074 are read into, respectively, VWL 24076 and DWL 24078. VWL 24076 and DWL 24078 respectively modify those validity or dirty word entries from VALS 24068 and DIRTYS 24070 in accordance to whether the corresponding entries in TS 24010 and DS 24050 are written into or modified. These modified validity and dirty words are then written, during second clock cycle, from VWL 24076 and DWL 24078 into, respectively, VALS 24068 and DIRTYS 24070. Control inputs of VWL 24076 and DWL 24078 are provided from FUCTL 20214.

Referring finally to Least Recent Used Logic (LRUL) 24080, LRUL 24080 tracks usage of cache entries. As previously described, the generalized cache of Fig. 240 is a four way, set associative cache with, for example, up to 16 entries in each of NC 10226's sets. Entries within a particular set are identified, as described above, by indexing the cache's TS 24010 and DS 24050 may contain, concurrently, up to four individual entries identified by the same index but distinguished by having different tags. In this case, one entry would reside in Set A, comprising TSA 24012 and DSA 24052, one in Set B, comprising TSB 24014 and DSB 24054, and so on. Since the possible number of individual entries having a common tag is greater than the number of cache sets, it may be necessary to delete a particular cache entry when another entry having the same tag is to be written into the cache. In general, the cache's least recently used entry would be deleted to provide a location in TS 24010 and DS 24050 for writing in the new entry. LRUL 24080 assists in determining which cache entries are to be deleted when necessary in writing in a new entry by tracking and indicating relative usage of the cache's entries. LRUL 24080 is primarily comprised of a memory, LRU Memory (MLRU) 24081, containing a word for each cache set. As described above, NC 10226, for example, includes 16 sets of 4 frames each, so that LRUL 24080's memory may correspondingly be, for example, 16 words long. Each word indicates relative usage of the 4 frames in a set and is a 6 bit word.

Words are generated and written into LRUL 24080's MLRU 24081, through Input Register A, B, C, D (RABCD) 24083, according to a write only algorithm executed by HE 24044, as described momentarily. Each bit of each six word pertains to a pair of frames within a particular cache set and indicates which of those two frames was more recently used than the other. For example, Bit 0 will contain logic 1 if Frame A was used more recently than Frame B and a logic zero if Frame B was used more recently than Frame A. Similarly, Bit 1 pertains to Frames A and C, Bit 2 to Frames A and D, Bit 3 to Frames B and C, Bit 4 to Frames B and D, and Bit 5 to Frames C and D. Initially, all bits of a particular LRUL 24080 word are set to zero.

Assuming, for example, that the frames of a particular set are used in the sequence Frame A, Frame D, Frame B; Bits 0 to 5 of that LRUL 24080 word will initially contain all zeros. Upon a reference to Frame A, Bits 0, 1, and 2, referring respectively to Frames A and B, Frames A and C, and Frames A and D, will be written as logic 1's. Bits 3, 4, and 5, referring respectively to Frames B and C, Frames B and D, and Frames C and D, will remain logic 0. Upon reference to Frame D, Bits 0 and 1, referring respectively to Frames A and B and Frames A and C, will remain logic 1's. Bit 2, referring to Frames A and D, will be changed from logic 1 to logic 0 to indicate that Frame D has been referred to more recently than Frame A. Bit 3, referring to Frames B and C, will remain logic 0. Bits 4 and 5, referring respectively to Frames B and D and Frames C and D, will be written as logic 0, although they are already logic zeros, to indicate respectively that Frame D has been used more recently than Frame B or Frame C. Upon reference to Frame B, Bit 0, referring to Frames A and B, will be written to logic 0 to indicate that Frame B has been used more recently than Frame A. Bits 1 and 2, referring respectively to Frames A and C and Frames A and D, will remain respectively as logic 1 and logic 0. Bits three and four, referring respectively to Frames B and C and Frames B and D, will be written as logics 1's to indicate respectively that Frame B has been used more recently than Frame C or Frame D. Bit five will remain logic 0.

When it is necessary to replace a cache entry in a particular frame, the LRUL 24080 word referring to the cache set containing that frame will be read from LRUL 24080's MLRL 24081 through LRU Register (RLRU) 24085 and decoded by LRU Decode Logic (LRUD) 24087 to indicate which is least recently used frame. This decoding is executed by means of a Read Only Memory operating as a set of decoding gating.

Having described the structure and operation of a generalized cache as shown in Fig. 240, with references to NC 10226 for illustration and to point out differences between the generalized cache and NC 10226, structure and operation of ATU 10228 and PC 10234 will be described next below. ATU 10228 and PC 10234 will be described by describing the differences between ATU 10228 and PC 10234 and the generalized cache and NC 10226. ATU 10228 will be described first, followed by PC 10234.

d.d. Address Translation Unit 10228 and Protection Cache 10234

ATU 10228 is a three-way, set associative cache of 16 sets, that is contains 3 frames for each set. Structure and operation of ATU 10228 is similar to the generalized cache described above. Having 3 rather than 4 frames per set, ATU 10228 does not include a STD 24018, ATSCE 24026, ATSPRD 24034, ATSHED 24042, or ADSD 24058. As previously described ATU 10228 address inputs comprise AON and O fields of logical descriptors. AON fields are each 28 bits and O fields comprise the 18 most significant bits of logical descriptor offset fields, so that ATU 10228 address inputs are 48 bits wide. Four least significant bits of O fields are used as index. AON fields and the 14 most significant bits of O field comprise ATU 10228's tags. ATU 10228 tags are thereby each 42 bits in width. Accordingly, TSA 24012, TSB 24014, and TSC 24016 of ATU 10228's TS 24010 are each 16 words long by 42 bits wide.

DSA 24052, DSB 24054, and DSC 24056 of ATU 10228 are each 16 bits long. ATU 10228 outputs are, as previously described, physical descriptor Frame Number (FN) fields of 13 bits each. ATU 10228's DSA 24052, DSB 24054, DSC 24056 are thereby each 13 bits wide.

ATU 10228's LRUL 24080 is similar in structure and operation to that of the generalized cache. ATU 10228's LRUL 24080 words, each corresponding to an ATU 10228 set, are each 3 bits in width as 3 bits are sufficient to indicate relative usage of frames within a 3 frame set. In ATU 10228, Bit 1 of an LRUL 24080 word indicates whether Frame A was used more recently than Frame B, Bit 2 whether Frame A was used more recently than Frame C, and Bit 3 whether Frame B was used more recently than Frame C. In all other respects, other than as stated above, ATU 10228 is similar in structure and operation to the generalized cache.

Referring to PC 10234, PC 10234 is a two-way, set associative cache of 8 sets, that is has two frames per set. Having 2 rather than 4 frames, PC 10234 will not include a TSC 24016, a TSD 24018, a TSCE 24024, a TSCD 24026, a TSPRC 24032, a TSPRD 24034, a TSHED 24042, a DSC 24056, or a DSD 24058.

Address inputs of PC 10234 are the 28 bit AON fields of logical descriptors. The 3 least significant bits of those AON fields are utilized as indexes for addressing PC 10234's TS 24010 and DS 24050. The 25 most significant bits of those AON field address inputs are utilized as PC 10234's tags, so that PC 10234's TSA 24012 and TSB 24014 are each 8 word by 25 bit memories.

Referring to PC 10234's LRUL 24080, a single bit is sufficient to indicate which of the two frames in each of PC 10234's sets was most recently accessed. PC 10234's LRUL 24080's memory is thereby 8 words, or sets long, one bit wide.

As previously described, PC 10234 entries comprise information regarding access rights of certain active subjects to certain active objects. Each PC 10234 entry contains 35 bits of information. Three bits of this information indicate whether a particular subject was read, write, or execute rights relative to a particular object. The remaining 32 bits effectively comprise a length field indicating the volume or portion, that is the number of data bits, of that object to which those access rights pertain.

Referring again to Fig. 240, PC 10234 differs from the generalized cache and from NC 10226 and ATU 10228 in further including Extent Check Logic (EXTCHK) 24082 and Operation Check Logic (OPRCHK) 24084. PC 10234 entries include, as described above, 3 bits identifying type of access rights a particular subject has to a particular object. These 3 bits, representing a Read (R), Write (W), or Execute (E) right, are provided to a

first input of OPRCHK 24084. A second input of OPRCHK 24084 is provided from FUCTL 20214 and specifies whether JP 10114 intends to perform a Read (RI), a Write (WI), or Execute (EI), operation with respect to that object. OPRCHK 24084 compares OPRCHK 24084 access right inputs from DS 24050 to OPRCHK 24084's intended operation input from FUCTL 20214. If that subject does not possess the rights to that object which are required to perform the operation intended by JP 10114, OPRCHK 24084 generates an Operation Violation (OPRV) indicating that a protection violation has occurred.

Similarly, the 32 bits of a PC 10234 entry regarding extent rights is provided as an input (EXTENT) to EXTCHK 24082. As stated above, EXTENT field of PC 10234 entry indicates the length or number of data bits, within an object, to which those access rights pertain. EXTENT field from PC 10234 entry is compared, by EXTCHK 24082, to offset field of the logical descriptor of the current JP 10114 request to MEM 10112 for which a current protection mechanism check is being made. If comparison of extent rights and offset field indicate that the current memory request goes beyond the object length to which the corresponding rights read from DS 24050 apply, EXTCHK 24082 generates an Extent Violation (EXTV) output. EXTV indicates that a current memory request by JP 10114 refers to a portion of an object to which the PC 10234 entry read from BS 24050 does not apply. As described previously, each read from or write to MEM 10112, even as part of a string transfer, is a 32 bit word. As such, EXTCHK 24082 will generate an EXTV output when OFFSET field of a current logical descriptor describes a segment of an object less than 32 bits from the limit defined by EXTENT field of the PC 10234 entry provided in response to that logical descriptor. EXTV and OPRV are gated together, by Protection Violation Gate (PVG) 24086 to generate Protection Violation (PROTV) output indicating that either an extent or an operation violation has occurred.

Having described the structure and operation of MEMINT 20212, and previously the structure and operation of DESP 20210, structure and operation of FUCTL 20214 will be described next below.

3. Fetch Unit Control Logic 20214 (Fig. 202)

The following descriptions will provide a detailed description of FU 10120's structure and operation. Overall operation of FU 10120 will be described first, followed by description of FU 10120's structure, and finally by a detailed description of FU 10120 operation.

As previously described, FUCTL 20214 directs operation of JP 10114 in executing procedures of user's processes. Among the functions performed by FUCTL 20214 are, first, maintenance and operation of CS 10110's Name Space, UID, and AON based addressing system, previously described; second, interpretation of SOPs of user's processes to provide corresponding sequences of microinstructions to FU 10120 and EU 10122 to control operation of JP 10114 in execution of user's processes, previously described; and, third, control of operation of CS 10110's internal mechanisms, for example CS 10110's stack mechanisms.

As will be described in further detail below, FUCTL 20214 includes prefetcher (PREF) 20260 which generates a sequence of logical addresses, each logical address comprising an AON and an offset field, for reading S-Instructions (SINs) of a user's program from MEM 10112. As previously described, each SIN may be comprised of an S-Operation (SOP) and one or more operand Names and may occupy one or more 32 bit words. SINs are read from MEM 10112 as a sequence of single 32 bit words, so that PREF 20260 need not specify a length field in a MEM 10112 read request for an SIN. SINs are read from MEM 10112 through MOD Bus 10144 and are captured and stored in Instruction Buffer (INSTB) 20262. PARSER 20264 extracts, or parses, SOPs and operand Names from INSTB 20262. PARSER 20264 provides operand Names to NC 10226 and SOPs to FUS Interpreter Dispatch Table (FUSDT) 11010 and to EU Dispatch Table (EUSDT) 20266 through Op-Code Register (OPCODEREG) 20268. Operation of INSTB 20262 and PARSER 20264 is controlled by Current program Counter (CPC) 20270, Initial Program Counter (IPC) 20272, and Executed program Counter (EPC) 20274.

As previously described, FUSDT 11010 provides, for each SOP received from OPCODEREG 20268, a corresponding S-Interpreter Dispatch (SD) Pointer, or address, to FUSITT 11012 to select a corresponding sequence of microinstructions to direct operation of JP 10114, in particular FU 10120. As previously described, FUSITT 11012 also contains sequences of microinstructions for controlling and directing operation of CS 10110's internal mechanisms, for example those mechanisms such as RCWS 10358 which are involved in swapping of processes. EUSDT 20266 performs an analogous function with respect to EU 10122 and provides SD Pointers to EU S-Interpreter Tables (EUSITTs) residing in EU 10122.

Micro-Program Counter (mPC) 20276 provides sequential addresses to FUSITT 11012 to select individual microinstructions of sequences of microinstructions. Branch and Case Logic (BRCASE) 20278 provides addresses to FUSITT 11012 to select microinstructions sequences for microinstructions branches and and cases. Repeat Counter (REPCTR) 20280 and Page Number Register (PNREG) 20282 provide addresses to FUSITT 11012 during FUSITT 11012 load operations.

As previously described, FUSITT 11012 is a writable microinstruction control store which is loaded with selected S-Interpreters (SINTs) from MEM 10112.

FUSITT 11012 addresses are also provided by Event Logic (EVENT) 20284 and by JAM input from NC 10226. As will be described further below, EVENT 20284 is part of FUCTL 20214's circuitry primarily concerned with operation of CS 10110's internal mechanisms. Input JAM from NC 10226 initiates certain FUCTL 20214 control functions for CS 10110's Name Space addressing mechanisms, and in particular NC 10226. Selection between the above discussed address inputs to FUSITT 11012 is controlled by S-

Interpreter Table Next Address Generator Logic (SITTNAG) 20286.

Other portions of FUCTL 20214's circuitry are concerned with operation of CS 10110's internal mechanisms. For example, FUCTL 20214 includes Return Control Word Stack (RCWS) 10358, previously described with reference to CS 10110's Stack Mechanisms. Register Address Generator (RAG) 20288 provides pointers for addressing of GRF 10354 and RCWS 10358 and includes Micro-Stack Pointer Registers (MISPR) 10356.

As previously described, MISPR 10356 mechanism provides pointers for addressing Micro-Stack (MIS) 10368. As will be described further below, actual MIS 10368 Pointers pointing to current, previous, and bottom frames of MIS 10368 reside in Micro-Control Word Register 1 (MCW1) 20290. MCW1 20290 and Micro-Control Word Zero Register (MCWO) 20292 together contain certain information indicating the current execution environment of a microinstruction sequence currently being executed by FU 10120. This execution information is used in aide of execution of these microinstruction sequences. State Registers (STATE) 20294 capture and store certain information regarding state of operation of FU 10120. As described further below, this information, referred to as state vectors, is used to enable and direct operation of FU 10120.

Timers (TIMERS) 20296 monitor elapsed time since occurrence of the events requiring servicing by FU 10120. If waiting time for these events exceeds certain limits, TIMERS 20296 indicate that these limits have been exceeded so that service of those events may be initiated.

Finally, Fetch Unit to E Unit Interface Logic (FUEUINT) 20298 comprises the FU 10120 portion of the interface between FU 10120 an EU 10122. FUEUINT 20298 is primary path through which operation of FU 10120 and EU 10122 is coordinated.

Having described overall operation of FU 10120, structure of FU 10120 will be described next below with aide of Fig. 202, description of FU 10120's structure will be followed by a detailed description of FU 10120 wherein further, more detailed, diagrams of certain portions of FU 10120 will be introduced as required to enhance clarity of presentation.

a.a. Fetch Unit Control Logic 20214 Overall Structure

Referring again to Fig. 202, as previously described Fig. 202 includes a partial block diagram of FUCTL 20214. Following the same sequence of description as above, PREF 20260 has a 28 bit bi-directional port connected to AON Bus 20230 and 32 bit bi-directional port directed from OFFSET Bus 20228. A control input of PREF 20260 is connected from control output of INSTB 20262.

INSTB 20262 32 bit data input (DI) is connected from MOD Bus 10144. INSTB 20262's 16 bit output (DO) is connected to 16 bit bi-directional input of OPCODEREG 20268 and to 16 bit NAME Bus 20224. OPCODEREG 20268's input comprises 8 bits of SINT and 3 bits of dialect selection. As previously described, NAME Bus 20224 is connected to 16 bit bi-directional port of Name Trap (NT) 20254, to address input ADR of NC 10226, and to inputs and outputs of OFFP 20228. Control inputs of INSTB 20262 and PARSER 20264 are connected from a control output of CPC 20270.

Thirty-two bit input of CPC 20270 is connected from JPD Bus 10142 and CPC 20270's 32 bit output is connected to 32 bit input of IPC 20272. Thirty-two bit output of IPC 20272 is connected to 32 bit input of EPC 20274 and to JPD Bus 10142. EPC 20274's 32 bit output is similarly connected to JPD Bus 10142.

Eleven bit outputs of OPCODEREG 20268 are connected to 11 bit address inputs of FUSDT 11010 and EUSDT 20266. These 11 bit address inputs to FUSDT 11010 and EUSDT 20266 each comprise 3 bits of dialect selection code and 8 bits of SINT code. Twelve bit SDT outputs of EUSDT 20266 is connected to inputs of Microinstruction Control Store in EU 10122, as will be described in a following description of EU 10122. FUSDT 11010 has, as described further below, two outputs connected to address (ADR) Bus 20298. First output of FUSDT 11010 are six bit SDT pointers, or addresses, corresponding to generic SINTs as will be described further below. Second output of FUSDT 11010 are 15 bit SDT pointers, or addresses, for algorithm microinstruction sequences, again as will be described further below.

Referring to RCWS 10358, RCWS 10358 has a first bidirectional port connected from JPD Bus 10142. Second, third, and fourth bi-directional ports of RCWS 10358 are connected from, respectively, a bi-directional port of MCW1 20290, a first bi-directional port EVENT 20284, and a bi-directional port of mPC 20276. An output of RCWS 10358 is connected to ADR Bus 20298.

An input of mPC 20276 is connected from ADR Bus 20298 and first and second outputs of mPC 20276 are connected to, respectively, an input of BRCASE 20278 and to ADR Bus 20298. An output of BRCASE 20278 is connected to ADR Bus 20298.

As described above, a first bi-directional port of EVENT 20284 is connected to RCWS 10358. A second bidirectional port of EVENT 20284 is connected from MCWO 20292. An output of EVENT 20284 is connected to ADR Bus 20298.

Inputs of RPCTR 20280 and PNREG 20282 are connected from JPD Bus 10142. Outputs of RPCTR 20280 and PNREG 20282 are connected to ADR Bus 20298.

ADR Bus 20298, and an input from a first output of FUSITT 11012, are connected to inputs of SITTNAG 20286.

Output of SITTNAG 20286 is connected, through Control Store Address (CSADR) Bus 20299, to address input of FUSITT 11012. Data input of FUSITT 11012 is connected from JPD Bus 10142. Control outputs of FUSITT 11012 are connected to almost all elements of JP 10114 and thus, for clarity of presentation, are not

shown in detail by drawn physical connections but are described in following descriptions.

As described above, MCW0 20292 and MCW1 20290 have bi-directional ports connected to, respectively, bidirectional ports of EVENT 20284 and to a second bidirectional port of RCWS 10358. Outputs of MCW0 20292 and MCW1 20290 are connected to JPD Bus 10142. Other inputs of MCW0 20292 and MCW1 20290, as will be described further below, are connected from several other elements of JP 10114 and, for clarity of presentation, are not shown herein in detail but are described in the following text. STATE 20294 similarly has a large number of inputs and outputs connected from and to other elements of JP 10114, and in particular FU 10120. Inputs and outputs of STATE 20294 are not indicated here for clarity of presentation and will be described in detail below.

RAG 20288 has an input connected from JPD Bus 10142 and other inputs connected, for example, from MCW1 20290. RAG 20288, including MISPR 10356, provides outputs, for example, as address inputs to RCWS 10358 and GRF 10354. Again, for clarity of presentation, inputs and outputs of RAG 20288 are not shown in detail in Fig. 202 but will be described in detail further below.

TIMERS 20296 receive inputs from EVENT 20284 and FUSITT 11012 and provide outputs to EVENT 20284. For clarity of presentation, these indications are not shown in detail in Fig. 202 but will be described further below.

FUINT 20298 receives control inputs from FUSITT 11012 and EU 10122. FUINT 20298 provides outputs to EU 10122 and to other elements of FUCTL 20214. For clarity of presentation, connections to and from FUINT 20298 are not shown in detail in Fig. 202 but will be described in further detail below.

Having described the overall operation, and structure, of FUCTL 20214, operation of FUCTL 20214 will be described next below. During the following descriptions further diagrams of certain portions of FUCTL 20214 will be introduced as required to disclose structure and operation of FUCTL 20214 to one of ordinary skill in the art. FUCTL 20214's operation with regard to fetching and interpretation of SINS, that is SOPs and operand Names, will be described first, followed by description of FUCTL 20214's operation with regard to CS 10110's internal mechanisms.

b.b. Fetch Unit Control Logic 20214 Operations

Referring first to those elements of FUCTL 20214 directly concerned with control of JP 10114 in response to SOPs and Name syllables, those elements include: (1) PREF 20260; (2) INSTB 20262; (3) PARSER 20264; (4) CPC 20270, IPC 20272, and EPC 20274; (5) OPCODEREG 20268; (6) FUSDT 11010 and EUSDT 20266; (7) mPC 20276; (8) BRCASE 20278; (9) REPCTR 20280 and PNREG 20282; (10) a part of RCWS 10358; (11) SITTNAG 20286; (12) FUSITT 11012; and, (13) NT 20254. These FUCTL 20214 elements will be described below in the order named.

a.a.a. Prefetcher 20260, Instruction Buffer 20262, Parser 20264, Operation Code Register 20268, CPC 20270, IPC 20272, and EPC 20274 (Fig. 241)

As described above, PREF 20260 generates a series of addresses to MEM 10112 to read SINS of user's programs from MEM 10112 to FUCTL 20214, and in particular to INSTB 20262. Each PREF 20260 read request transfers one 32 bit word from MEM 10112. Each SIN may be comprised of an SOP and one or more Name syllables. Each SOP may comprise, for example, 8 bits of information while each Name syllable may comprise, for example, 8, 12, or 16 bits of data. In general, and as will be described in further detail in a following description of STATE 20294, PREF 20260 obtains access to MEM 10112 on alternate 110 nanosecond system clock cycles. PREF 20260's access to MEM 10112 is conditional upon INSTB 20262 indicating that INSTB 20262 is ready to receive an SIN read from MEM 10112. In particular, INSTB 20262 generates control output Quiry Prefetch (QPF) to PREF 20260 to enable PREF 20260 to submit a request to MEM 10112 when, as described further below, INSTB 20262 is ready to receive an SIN read from MEM 10112.

PREF 20260 is a counter register comprised, for example of SN74S163s.

Bi-directional inputs and outputs of PREF 20260 are connected to AON Bus 20230 and OFFSET Bus 20228. As PREF 20260 reads only single 32 bit words, PREF 20260 is not required to specify a LENGTH field as part of an SIN read request, that is an AON and an OFFSET field are sufficient to define a single 32 bit word. At start of read of a sequence of SINS from MEM 10112, address (AON and OFFSET fields) of first 32 bit word of that SIN sequence are provided to MEM 10112 by DESP 20210 and concurrently loaded, from AON Bus 20230 and OFFSET Bus 20228, into PREF 20260. Thereafter, as each successive thirty-two bit word of the SIN's sequence is read from MEM 10112, the address residing in PREF 20260 is incremented to specify successive 32 bit words of that SIN's sequence. The successive single word addresses are, for all words after first word of a sequence, provided to MEM 10112 from PREF 20260.

As described above, INSTB 20262 receives SINS from MEM 10112 through MOD Bus 10144 and, with PARSER 20264 and operating under control of CPC 20270, provides Name syllables to NAME Bus 20224 and SINS to OPCODEREG 20268. INSTB 20262 is provided, together with PREF 20260 to increase execution speed of SINS.

Referring to Fig. 241, a more detailed block diagram of INSTB 20262, PARSER 20264, CPC 20270, IPC 20272, EPC 20274 as shown. INSTB 20262 is shown as comprising two 32 bit registers having parallel 32 bit inputs from MOD Bus 10144. INSTB 20262 also receives two Write Clock (WC) inputs, one for each 32 bit register of INSTB 20262, from Instruction Buffer Write Control (INSTBWC) 24110. INSTB 20262's outputs

are structured as eight, eight bit Basic Syllables (BSs), indicated as BSO to BS7. BSO, BS2, BS4, and BS6 are ORed to comprise eight bit Basic Syllable, Even (BSE) of INSTB 20262 while BSO, BS3, BS5, and BS7 are similarly ORed to comprise Basic Syllable, Odd (BSO) of INSTB 20262. BSO and BSE are provided as inputs of PARSER 20264.

5 PARSER 20264 receives a first control input from Current Syllable Size Register (CSSR) 24112, associated with CPC 20270. A second control input of PARSER 20264 is provided from Instruction Buffer Syllable Decode Register (IBSDECR) 24114, also associated with CPC 20270. PARSER 20264 provides an eight bit output to NAME Bus 20224 and to input of OPCODEREG 20268.

10 Referring to INSTBWC 24110, INSTBWC 24110 provides, as described further below, control signals pertaining to writing of SINs into INSTB 20262 from MOD Bus 10144. INSTBWC 24110 also provides control signals pertaining to operation of PREF 20260. In addition to WC outputs to INSTB 20262, INSTBWC 24110 provides control output QPF to PREF 20260, control output Instruction Buffer Hung (IBHUNG) to EVENT 20284, and control signal Instruction Buffer Wait (IBWAIT) to STATE 20294. INSTBWC 24110 also receives a control input BRANCH from BRCASE 20278 and an error input from TIMERS 20296.

15 Referring to CPC 20270, IPC 20272, and EPC 20274, IPC 20272 and EPC 20274 are represented in Fig. 241 as in Fig. 202. Further FUCTL 20214 circuitry is shown as associated with CPC 20270. CPC 20270 is a twenty-nine bit register receiving bits one to twenty-five (CPC(1—25)) from bits one to twenty-five of JPD Bus 10142. CPC 20270 Bit 0 (CPC0) is provided from CPC0 CPC0 Select (CPCOS) 24116. Inputs of CPCOS 24116 are Bit 1 output from CPC 20270 (CPC1) and Bit 0 from JPD Bus 10142. Bits twenty-six, twenty-seven, and twenty-eight of CPC 20270 (CPC(2628)) are provided from CPC Multiplexer (CPCMUX) 24118. CPCMUX 24118 also provides an input to IBSDECR 24114. Inputs of CPCMUX 24118 are bits twenty-five, twenty-six, and twenty-eight from JPD Bus 10142 and a three bit output of CPC Arithmetic and Logic Unit (CPCALU) 24120. A first input of CPCALU 24120 is connected from output bits 26, 27, and 28 of CPC 20270. Second input of CPCALU 24120 is connected from CSSR 24112. CSSR 24112's input is connected from JPD Bus 10142.

25 As described above, INSTB 20262 is implemented as a sixty-four bit wide register. INSTB 20262 is organized as two thirty-two bit words, referred to as Instruction Buffer Word 0 (IB0) and Instruction Buffer Word 1 (IB1), and operates as a two word, first-in-first-out buffer memory. PREF 20260 loads one of IB0 or IB1 on each memory reference by PREF 20260. Only PREF 20260 may load INSTB 20262, and INSTB 20262 may be loaded only from MOD Bus 10144. Separate clocks, respectively Instruction Buffer Write Clock 0 (IBWC0) and Instruction Buffer Write Clock 1 (IBWC1), are provided from INSTBWC 24110 to load, respectively, IBW0 and IBW1 into INSTB 20262. IBWC0 and IBWC1 are each a gated 110 nano-second clock. An IBW0 or an IBW1 is written into INSTB 20262 when, respectively, IBWC0 or IBWC1 is enabled by INSTBWC 24110. IBWC0 and IBWC1 will be enabled only when MEM 10112 indicates that data for INSTB 20262 is available by asserting interface control signal DAVI as previously discussed.

35 INSTBWC 24110 is primarily concerned with control of FU 10120 with respect to writing of SINs into INSTB 20262. As described above, INSTBWC 24110 provides IBWC0 and IBWC1 to INSTB 20262. IBWC0 and IBWC1 are enabled by INSTBWC 24110's input DAVI from MEM 10112. Selection between IBWC0 and IBWC1 is controlled by INSTBWC 24110's input from CPC 20270. In particular, and as will be described further below, Bit 26 (CPC 26) of CPC 20270's twenty-nine bit word indicates whether IBW0 or IBW1 is written into INSTB 20262.

40 In addition to controlling writing of IBW0 and IBW1 into INSTB 20262, INSTBWC 24110 provides control signals to elements of FU 10120 to control reading of SINs from MEM 10112 to INSTB 20262. In this regard, INSTBWC 24110 detects certain conditions regarding status of SIN words in INSTB 20262 and provides corresponding control signals, described momentarily, to other elements of FU 10120 so that INSTB 20262 would generally always contain at least one valid SOP or Name syllable. First, if INSTB 20262 is not full, that is either IBW0 or IBW1 or both is invalid, for example because IBW0 has been read from INSTB 20262 and executed, INSTBWC 24110 detects this condition and provides control signal QPF to PREF 20262 to initiate a read from MEM 10112. INSTBWC 24110 currently enables either IBW0 or IBW1 portion of INSTB 20262 to receive the word read from MEM 10112 in response to PREF 20260's request. As stated above, this operation will be initiated when INSTBWC 24110 detects and indicates, by generating a validity flag, that either IBW0 or IBW1 is invalid. In this case, IBW0 or IBW1 will be indicated as invalid when read from INSTB 20262 by PARSER 20264. As will be described further below, INSTBWC 24110 validity flags for IBW0 and IBW1 are generated by INSTBWC 24110 control inputs comprising Bits 26 to 28 (CPC 26—28) from CPC 20270 and by current syllable size or value, flag (K) input from CSSR 24112. Secondly, INSTBWC 24110 will detect when INSTB 20262 is empty, that is when both IBW0 and IBW1 are invalid, as just described, or when only a half of a sixteen bit Name syllable is present in INSTB 20262. In response to either condition, INSTBWC 24110 will generate control signal IBWAIT to STATE 20294. As will be described further below, IBWAIT will result in suspension of execution of microinstructions referencing INSTB 20262. PREF 20260 requests to MEM 10112 will already have been initiated, as described above unless certain other conditions, described momentarily, occur. Thirdly, INSTBWC 24110 will detect when INSTB 20262 is empty and PREF 20262 is hung, that is unable to submit requests to MEM 10112, and a current microinstruction is attempting to parse a syllable from INSTB 20262. In this case, INSTBWC 24110 will generate control signal Instruction Buffer Hung (IBHUNG) to EVENT 20284. As will be described further below, IBHUNG will result in initiation of a microinstruction sequence to restore flow of words to INSTB 20262. Fourthly, INSTBWC 24110 will detect, through microinstruction control signals provided from FUSITT 11012, when a branch in

EP 0 067 556 B1

a microinstruction sequence provided by FUSITT 11012 in response to an SOP occurs. In this case, both IBW0 and IBW1 will be flagged as invalid. INSTBWC 24110 will then ignore SIN words being read from MEM 10112 in response to a previously submitted PREF 20260 request, but not yet received at the time the branch occurs. This prevents INSTB 20260 from receiving invalid SIN words; PREF 20260 and INSTB 20262 will then proceed to request and receive valid SIN words of the branch.

As described above, PARSER 20264, operating under control of CPC 20270 and CPC 20270 associated circuitry, reads Name syllables and SOPs from INSTB 20262 to, respectively, NAME Bus 20224 and OPCODEREG 20268. PARSER 20264 operates as a multiplexer with associated control logic.

As previously described, INSTB 20262 is internally structured as eight, eight bit words, BS0 to BS7. IBW0 comprises BS0 to B3 while IBW1 comprises BS4 to BS7. Each SOP is comprised of eight bits of data and thus comprises one Basic Syllable while each Name syllable comprises 8, 12, or 16 bits of data and thus comprises either one or two Basic Syllables. Name syllable size, as previously stated, is indicated by Current Syllable Size Value K stored in CSSR 24112.

BS0 and BS4 are loaded into INSTB 20262 from MOD Bus 10144 bits zero to seven while BS2 and BS6 are loaded from MOD Bus 10144 bits sixteen to twenty-three. BS1 and BS5 are loaded from MOD Bus 10144 bits eight to fifteen while BS3 and BS7 are loaded from MOD Bus 10144 bits twenty-four to thirty-one. Odd numbered Basic Syllable outputs BS1, BS3, BS5, and BS7 are ORed to comprise eight bit Basic Syllable, Odd output BSO of INSTB 20262. Even numbered Basic Syllable outputs BSo, BS2, BS4 and BS6 of INSTB 20262 are similarly ORed to comprise eight bit Basic Syllable, Even output BSE. At any time, one odd numbered Basic Syllable output and one even numbered Basic Syllable output of INSTB 20262 are selected as inputs to PARSER 20264 by Instruction Buffer Read Enable (IBORE) enable and selection signals provided to INSTB 20262 by IBSDECR 24114. IBSDECR 24114 includes decoding circuitry. Input to IBSDECR 24114's decoding logic is comprised of three bits (RCPC(26-28)) provided from CPCMUX 24118. As indicated in Fig. 241, CPC (26-28) may be provided from JPD Bus 10142 bits 25 to 28 or from output of CPCALU 24120. One input CPCALU 24120 is CPC (26-28) from CPC 20270. Operation of CPC 20270 and CPC 20270's associated circuitry will be described further below. RCPC (26-28) is decoded by IBSDECR 24114 to generate IBORE (0-7) to INSTB 20262. RCPC 28 and RCPC 27 are decoded to select one of the four odd numbered Basic Syllable outputs (that is BS1, BS3, BS5 or BS7) of INSTB 20262 as the odd numbered basic syllable input to PARSER 20264. RCPC 28 selects either the preceding or the following even numbered Basic Syllable output of INSTB 20262 as the even numbered Basic Syllable input to PARSER 20264. The eight decoded bits of IBORE (0-7) generated by IBSDECR 24114 decoding logic are loaded into IBSDECR 24114 eight bit register and subsequently provided to INSTB 20262 as IBORE (0-7).

PARSER 20264 selects BSO, or BSE, or both BSO and BSE, as PARSER 20264's output to NAME Bus 20224 or to OPCODEREG 20268. In the case of an SOP or an eight bit Name syllable, either BSO or BSE will be selected as PARSER 20264's output. In the case of a twelve or sixteen bit Name syllable, both BSO and BSE may be selected as PARSER 20264's output. PARSER 20264 operation is controlled by microinstruction control outputs from FUSITT 11012.

Program counters IPC 20272, EPC 20274, and CPC 20270 are associated with control of fetching and parsing of SINs. In general, IPC 20272, EPC 20274, and CPC 20270 operate under microinstruction control from FUSITT 11012.

CPC 20270 is Current Program Counter and contains 28 bits pointing to the current syllable in INSTB 20272. Bits 29 to 31 of CPC 20270 are not provided, so the bits 29 to 31 of CPC 20270's output are zero, which guarantees byte boundaries for SOPs. Contents of CPC 20270 are thereby also a pointer which is a byte align offset into a current procedure object. Initial Program Counter (IPC) 20272 is a buffer register connected from output of CPC 20270 and provided for timing overlap. IPC 20272 may be loaded only from CPC 20270 which, as previously described, is 29 bits wide, that does not contain bits 29, 30, and 31 which are forced to zero in IPC 20272. IPC 20272 may be read onto JPD Bus 10142 as a start value in an unconditional branch.

EPC 20274 is a thirty-two bit register usually containing a pointer to the current SOP being executed. Upon occurrence of an SOP branch, the pointer in EPC 20274 will point to the SOP from which the branch was executed. The pointer residing in EPC 20274 is an offset into a current procedure object. EPC 20274 may be loaded only from IPC 20272, and may be read onto JPD Bus 10142.

Referring again to CPC 20270, as described above CPC 20270 is a current syllable counter. CPC 20270 contains a pointer to the next SOP syllable, or Base Syllable, to be parsed by PARSER 20264. As SOPs are always on byte boundaries, CPC 20270 pointer is 29 bits wide, CPC (0-28). The three low order bits of CPC 20270's pointer, that is CPC (29-31), do not physically exist and are assumed to be always zero. CPC 20270's pointer to next instruction syllable to be parsed thereby always points to byte boundaries.

CPC 20270 bits 26 to 28, CPC (26-28), indicate, as described above, a particular Base Syllable in INSTB 20262. Bits 0-25 (CPC(0-25)) of CPC 20270 indicate 32 bit words, read into INSTB 20262 as IBW0 and IBW1, of a sequence of SINs. CPC 20270 pointer is updated each time a parse operation reading a Base Syllable from INSTB 20262 is executed. As previously described, these parsing operations are performed under microinstruction control from FUSITT 11012.

Conceptually, CPC 20270 is organized as a twenty-six bit counter, containing CPC (0-25), with a three bit register appended on the low order side, as CPC (26-28). This organization is used because CPC (26-28) counts INSTB 20262 Base Syllables parsed and must be incremented dependant upon current

Name Syllable Size K stored CSSR 24112. CPC (0—25), however, counts successive thirty-two bit words of a sequence of SINS and may thereby be implemented as a binary counter. As shown in Fig. 241, CPC (26—28) is loaded from output of CPCMUX 24118. A first input of CPCMUX 24118 is connected from bits 29 to 31 of JPD Bus 10142. This input to CPC (26—28) from JPD Bus 10142 is provided to allow CPC 20270 to be loaded from JPD Bus 10142, for example when loading CPC 20270 with an initial pointer value. Second input of CPCMUX 24118 is from output of CPCALU 24120 and is the path by which CPC (26—28) is incremented as successive Base Syllables are parsed from INSTB 20262. A first input of CPCALU 24120 is CPC (26—28) from CPC 20270. Second input of CPCALU 24120 is a dual input from CSSR 24112. First input from CSSR 24112 is logic 1 in the least significant bit position, that is in position corresponding to CPC (28). This input is used when single Base Syllables are parsed from INSTB 20262, for example in an eight bit SOP or an eight bit Name syllable. CSSR 24112's first input to CPCALU 24120 increments CPC (0—32) by eight, that is one to CPC (26—28), each time a single Base Syllable is parsed from INSTB 20262. Second input to CPCALU 24120 from CSSR 24112 is K, that is current Name Syllable size. As previously described, K may be eight, twelve, or sixteen. CPC (26—28) is thereby incremented by one when K equals eight and is incremented by two when K equals twelve or sixteen. As shown in Fig. 241, K is loaded into CSSR 24112 from JPD Bus 10142.

CPC (0—25), as described above, operates as a twenty-six bit counter which is incremented each time CPC (26—28) overflows. CPC (0—25) is incremented by carry output of CPCALU 24120. In actual implementation, CPC 20270 is organized to reduce the number of integrated circuits required. CPC (1—25) is constructed as a counter and inputs of CPC (1—25) counter are connected from bits 1 to 24 of JPD Bus 10142 to allow loading of an initial value of CPC 20270 pointer. CPC (0) and CPC (26—28) are implemented as a four bit register. Operation of CPC (26—28) portions of this register have been described above. Input of CPC (0) portion of this register is connected from output of CPCOS 24116. CPCOS 24116 is a multiplexer having a first input connected from bit 0 of JPD Bus 10142. This input from JPD Bus 10142 is used, for example, when loading CPC 20272 with an initial pointer value. Second input of CPCOS 24116 is overflow output of CPC (1—25) counter and allows CPC (0) portion of the four bit register and CPC (1—25) counter to operate as a twenty-six bit counter.

Finally, as shown in Fig 241, output of CPC 20270 may be loaded into IPC 20272. An initial CPC 20270 pointer value may therefore be written into CPC 20270 from JPD Bus 10142 and subsequently copied into IPC 20272.

Referring again to PARSER 20264, as described above PARSER 20264 reads, or parses, basic syllables from INSTB 20262 to NAME Bus 20224. Input of PARSER 20264 is a sixteen bit word comprised of an eight bit odd numbered Base Syllable, BSO, and an eight bit even numbered Base Syllable, BSE. Depending upon whether PARSER 20264 is parsing an eight bit SOP, an eight bit Name syllable, a twelve bit Name syllable, or sixteen bit Name syllable, PARSER 20264 may select BSO, BSE, or both BSO and BSE, as output onto NAME Bus 20224.

If PARSER 20264 is parsing Name syllables and K is not equal to eight, that is equal to twelve or sixteen, PARSER 20264 transfers both BSO and BSE onto NAME Bus 20224 and determines which of BSO or BSE is most significant. The decision as to whether BSO or BSE is most significant is determined by CPC (28). If CPC (28) indicates BSO is most significant, BSO is transferred onto NAME Bus 20224 bits 0 to 7 (NAME(0—7)) and BSE onto NAME Bus 20224 bits eight to fifteen (NAME(8—15)). If CPC (28) indicates BSE is most significant, BSE is transferred onto NAME (0—7) and BSO onto NAME (8—15). This operation insures that Name syllables are parsed onto NAME Bus 20224 in the order in which occur in the SIN stream.

If PARSER 20264 is parsing Name syllables of Syllable Size K = 8, PARSER 20264 will select either BSO or BSE, as indicated by CPC (28), as output to NAME (0—7). PARSER 20264 will place 0's on NAME (8—15).

If PARSER 20264 is parsing SOPs of eight bits, PARSER 20264 will select BSO or BSE as output to NAME (0—7) as selected by CPC (28). PARSER 20264 will place 0's onto NAME (8—15). Concurrently, PARSER 20264 will generate OPREG to OPCODEREG 20268 to enable transfer of NAME (0—7) into OPCODEREG 20268. OPCODEREG 20268 is not loaded when PARSER 20264 is parsing Name syllables. The microinstruction input from FUSITT 11012 which controls PARSER 20264 operation also determines whether PARSER 20264 is parsing an SOP or a Name syllable and controls generation of OPREG.

Operation of NC 10226, which receives Name syllables as address inputs from NAME Bus 20224, has been discussed previously with reference to MEMINT 20212. Name Trap (NT) 20254 is connected from NAME Bus 20224 to receive and capture Name syllables parsed onto NAME Bus 20224 by PARSER 20264. Operation of NT 20254 has been also previously discussed with reference to MEMINT.

b.b.b. Fetch Unit Dispatch Table 11010, Execute Unit Dispatch Table 20266 and Operation Code Register 20268 (Fig. 242)

As previously described, CS 10110 is a multiple language machine. Each program written in a high level user language is compiled into a corresponding S-Language program containing S-Language instructions referred to as SOPs. CS 10110 provides a set or dialect, of microcode instructions, referred to as S-Interpreters (SINTs) for each S-Language. SINTs interpret SOPs to provide corresponding sequences of microinstructions for detailed control of CS 10110 operations. CS 10110's SINTs for FU 10120 and EU 10122 operations are stored, respectively, in FUSITT 11012 and in a corresponding control store memory in EU 10122, described in a following description of EU 10122. Each SINT comprises one or more sequences

of microinstructions, each sequence of microinstructions corresponding to a particular SOP in a particular S-Language dialect. Fetch Unit S-Interpreter Dispatch Table (FUSDT) 11010 and Execute Unit S-Interpreter Dispatch Table (EUSDT) 20266 contain an S-Interpreter Dispatcher (SD) for each S-Language dialect. Each SD is comprised of a set of SD Pointers (SDPs) wherein each SDP in a particular SD corresponds to a particular SOP of that SD dialect. Each SDP is an address pointing to a location, in FUSITT 11012 or EUSITT, of the start of the corresponding sequence of microinstructions for interpreting the SOP corresponding to that SDP. As will be described further below, SOPs received and stored in OPCODEREG 20268 are used to generate addresses into FUSDT 11010 and EUSDT 20266 to select corresponding SDPs. Those SDPs are then provided to FUSITT 11012 through ADR 20202, or to EUSITT through EUDIS Bus 20206, to select corresponding sequences of microinstructions from FUSITT 11012 and EUSITT.

Referring to Fig. 242, a more detailed block diagram of OPCODEREG 20268, FUSDT 11010, and EUSDT 20266 is shown. As shown therein, OPCODEREG 20268 is comprised of OP-Code Latch (LOPCODE) 24210, Dialect Register (RDIAL) 24212, Load Address Register (LADDR) 24214, and Fetch Unit Dispatch Encoder (FUDISENC) 24216. Data inputs of LOPCODE 24010 are connected from NAME Bus 20224 to receive SOPs parsed from INSTB 20262. Load inputs of RDIAL 24212 are connected from Bits 28 to 31 of JPD Bus 10142. Outputs of LOPCODE 24210, RDIAL 24212 and LADDR 24214 are connected to inputs of FUDISENC 24216. Outputs of FUDISENC 24216 are connected to address inputs of FUSDT 11010 and EUSDT 20266.

FUSDT 11010 is comprised of Fetch Unit Dispatch File (FUDISF) 24218 and Algorithm File (AF) 24220. Address inputs of FUDISF 24218 and AF 24220 are connected, as previously described, from address outputs of FUDISENC 24216. Data load inputs of FUDISF 24218 and AF 24220 are connected from, respectively, Bits 10 to 15 and Bits 16 to 31 of JPD Bus 10142. SDP outputs of FUDISF 24218 and AF 24220 are connected to ADR Buses 20202.

EUSDT 20266 is comprised of Execute Unit Dispatch File (EUDISF) 24222 and Execute Unit Dispatch Selector (EUDISS) 24224. Address inputs of EUDISF 24222 are, as described above, connected from outputs of FUDISENC 24216. Data load inputs of EUDISF 24222 are connected from Bits 20 to 31 of JPD Bus 10142. Inputs of EUDISS 24224 are connected from SDP output of EUDISF 24222, from Bits 20 to 31 of JPD Bus 10142, and from Microcode Literal (mLIT) output of FUSITT 11012. SDP outputs of EUDISS 24224 are connected to EUDIS Bus 20206.

As previously described, OPCODEREG 20268 provides addresses, generated from SOPs loaded into OPCODEREG 20268, to FUSDT 11010 and EUSDT 20266 to select SDPs to be provided as address inputs to FUSITT 11012 and EUSITT. LOPCODE 24210 receives and stores eight bit SOPs parsed from INSTB 20262 as described above. OPCODEREG 20268 also provides addresses to FUSDT 11010 and EUSDT 20266 to load FUSDT 11010 and EUSDT 20266 with SDs for S-Language dialects currently being utilized by CS 10110. LOPCODE 24210 and RDIAL 24212, as described below, provide addresses to FUSDT 11010 and EUSDT 20266 when translating SOPs to SDPs and ADDR 24214 provides addresses when FUSDT 11010 and EUSDT 20266 are being loaded with SDs.

Referring first to LADDR 24214, LADDR 24214 has an eight bit counter. Addresses are provided to FUSDT 11010 and EUSDT 20266 from LADDR 24214 only when FUSDT 11010 and EUSDT 20266 are being loaded with SDs, that is groups of SDPs for S-Language dialects currently being utilized by CS 10110. During this operation, output of LADDR 24214 is enabled to FUSDT 11010 and EUSDT 20266 by microcode control signals (not shown for clarity of presentation) from FUSITT 11012. Selection between FUDISF 24218, AF 24220, and EUDISF 24222 to receive addresses is similarly provided by microinstruction enable signals (also not shown for clarity of presentation) provided from FUSITT 11012. These FUSDT 11010 and EUSDT 20266 address enable inputs may select, at any time, any or all of FUDISF 24218, AF 24220, or EUDISF 24222 to receive address inputs. SDPs to be loaded into FUDISF 24218, AF 24220, and EUDISF 24222 are provided, respectively, from Bits 10 to 15 (JPD(10-15)), Bits 16 to 31 (JPD(16-31)), and Bits 20 to 31 (JPD(20-31)) of JPD Bus 10142. Address contents of LADDR 24214 are successively incremented by one as successive SDPs are loaded into FUSDT 11010 and EUSDT 20266. Incrementing of LADDR 24214 is, again, controlled by microinstruction control inputs from FUSITT 11012.

Address inputs to FUSDT 11010 and EUSDT 20266 during interpretation of SOPs are provided from LOPCODE 24210 and RDIAL 24212. LOPCODE 24210 is a register counter having, as described above, data inputs connected from NAME Bus 20224 to receive SOPs from PARSER 20264. In a first mode, LOPCODE 24210 may operate as a latch, loaded with one SOP at a time from output of PARSER 20264. In a second mode, LOPCODE 24210 operates as a clock register to receive successive eight bit inputs from low order eight bits of NAME Bus 20224 (NAME(8-15)). Loading of LOPCODE 24210 is controlled by microinstruction control outputs (not shown for clarity of presentation) from FUSITT 11012.

As will be described further below, eight bit SOPs stored in LOPCODE 24210 are concatenated with the output of RDIAL 24212 to provide addresses to FUSDT 11010 and EUSDT 20266 to select SDPs corresponding to particular SOPs. That portion of these addresses provided from LOPCODE 24210, that is the eight bit SOPs, selects particular SDPs within a particular SD. Particular SDs are selected by that portion of these addresses which is provided from the contents of RDIAL 24212.

RDIAL 24212 receives and stores four bit Dialect Codes indicating the particular S-Language dialect currently being used by CS 10110 and executing the SOPs of a user's program. These four bit Dialect Codes are provided from JPD Bus 10142, as JPD (28-31). Loading of RDIAL 24212 with four bit Dialect Codes is controlled by microinstruction control signals provided from FUSITT 11012 (not shown for clarity of

presentation).

Four bit Dialect Codes in RDIAL 24212 define partitions in FUDISF 24218, AF 24220 and EUDISF 24222. Each partition contains SDPs for a different S-Language dialect, that is contains a different SD. FUDISF 24218, AF 24220 and EUDISF 24222 may contain, for example, eight 128 word partitions or four 256 word partitions. A single bit of Dialect Code, for example Bit 3, defines whether FUDISF 24218, AF 24220, and EUDISF 24222 contain four or eight partitions. If FUSDT 11010 and EUSDT 20266 contain four partitions, the two most significant bits of address into FUSDT 11010 and EUSDT 20266 are provided from Dialect Code Bits 1 and 2 and determine which partition is addressed. The lower order eight bits of address are provided from LOPCODE 24210 and determine which word in a selected partition is addressed. If FUSDT 11010 and EUSDT 20266 contain eight partitions, the three most significant bits of address into FUSDT 11010 and EUSDT 20266 are provided from Bits 0 to 2 of Dialect Code, to select a particular partition, and the lower seven bits of address are provided from LOPCODE 24210 to select a particular word in the selected partition.

As described above, LOPCODE 24210 eight bit output and RDIAL 24212's four bit output are concatenated together, through FUDISENC 24216, to provide a ten bit address input to FUSDT 11010 and EUSDT 20266. FUDISENC 24216 is an encoding circuit and will be described further below with reference to FUDISF 24218. As previously described, selection of FUDISF 24218, AF 24220, and EUDISF 24222 to receive address inputs from RDIAL 24212 and LOPCODE 24210 is controlled by microinstruction control enable inputs provided from FUSITT 11012 (not shown for clarity of presentation).

Referring to FUSDT 11010, both FUDISF 24218 and AF 24220 provide SDPs to FUSITT 11012, but do so for differing purposes. In general, microinstruction control operations may be regarded as falling into two classes. First, there are those microinstruction operations which are generic, that is general in nature and used by or applying to a broad variety of SOPs of a particular dialect or even of many dialects. An example of this class of microinstruction operation is fetches of operand values. FUDISF 24218 provides SDPs for this class of microinstruction operations. As described below, FUDISF 24218 is a fast access memory allowing a single microinstruction control output of FUSITT 11012 to parse an SOP from INSTB 20262 into LOPCODE 24210, and a corresponding SDP to be provided from FUDISF 24218. That is, an SOP of this generic class may be parsed from INSTB 20262 and a corresponding SDP provided from FUDISF 24218 during a single system clock cycle. Operation of FUDISF 24218 thereby enhances speed of operation of JP 10114, in particular at the beginning of execution of new SOPs.

The second class of microinstruction operations are those specific to particular SINTs or to particular groups of SINTs. These groups of SINTs may reside entirely within a particular dialect, for example FORTRAN, or may exist within one or more dialects. SDPs for this class of microinstruction operation are provided by AF 24220. As described further below, AF 24220 is slower than FUDISF 24218, but is larger. In general, AF 24220 contains SDPs of microinstruction sequences specific to particular SINTs. In general, generic microinstruction operations are performed before those operations specific to particular SINTs, so that SDPs are required from AF 24220 at a later time than those from FUDISF 24218. SDPs for specific SINT operations may therefore be provided from lower speed AF 24220 without a penalty in speed of execution of SOPs.

Referring again to FUDISF 24218, FUDISF 24218 is a 1,024 word by 6 bit fast access by polar memory. Each word contained therein, as described above, is an SDP, or address to start of a corresponding sequence of microinstructions in FUSITT 11012. As will be described further below, FUSITT is an 8K (8192) word memory. SDPs provided by FUDISF 24218 are each, as described above, 6 bits wide and may thus address a limited, 32 word area of FUSITT 11012's address space. FUDISF 24218 is enabled to provide SDPs to FUSITT 11012 by microinstruction control signals (not shown for clarity of presentation) from FUSITT 11012. FUDISF 24218 six bit SDPs are encoded by FUDISENC 24219 to address FUSITT 11012 address space in increments of 4 microinstructions, that is in increments of 4 address locations. FUDISF 24218 SDPs thereby address 4 microinstructions at a time from FUSITT 11012's microinstruction sequences. As will be described further below, mPC 20276 generates successive microinstruction addresses to FUSITT 11012 to select successive microinstructions of a sequence following an initial microinstruction selected by an SDP from FUSDT 11010. An FUDISF 24218 SDP will thereby select the first microinstruction of a 4 microinstruction block, and mPC 20276 will select the following 3 microinstructions of that 4 microinstruction sequence. A 4 microinstruction sequence may therefore be executed in line, or sequentially, for each FUDISF 24218 SDP provided in response to a generic SOP. FUDISENC 24219 encodes FUDISF 24218 six bit SDPs to select these 4 microinstruction sequences so that the least significant bit of these SDPs occupies the 24 bit of FUSITT 11012 address inputs, and so on. The two least significant bits of an FUSITT 11012 address, or SDP, provided from FUDISF 24218 are forced to 0 while the ninth and higher bits may be hard-wired to define any particular block of 128 addresses in FUSITT 11012. This hard-wiring of the most significant bits of FUSITT 11012 addresses from FUDISF 24218 allows a set of generic microinstruction sequences selected by FUDISF 24218 to be located as desired within FUSITT 11012's address space. FUDISENC 24219 is comprised of a set of driver gates.

As previously described, SDPs for generic microinstructions currently being utilized by CS 10110 in executing user's programs are written into FUDISF 24218 from Bits 10 to 15 of JPD Bus 10142 (JPD(10—15)). Addresses for loading SDPs into FUDISF 24218 are provided, as previously described, from LADDR 24214. LADDR 24214 is enabled to provide load addresses, and FUDISF 24218 is enabled to be

written into, by microinstruction control signals (not shown for clarity of presentation) provided from FUSITT 11012.

Referring to AF 24220, as previously described AF 24220 is of larger capacity than FUDISF 24218, but has slower access time. AF 24220 is a 1,024 word by 15 bit memory. In general, 2 clock cycles are required to obtain a DSP from AF 24220. During first clock cycle, an SOP is loaded into LOPCODE 24210 and, during second clock cycle, AF 24220 is addressed to provide a corresponding SDP. SDPs provided by AF 24220 are each 15 bits in width and thus capable of addressing a larger address space than that of FUSITT 11012. As previously described, FUSITT 11012 is an 8K word memory. If FUSITT 11012 is addressed by an AF 24220 SDP referring to an address location outside of FUSITT 11012's address space, FUSITT 11012 will generate a microinstruction Not In Control Store output to EVENT 20284 as described further below. An AF 24220 SDP resulting in this event will then be used to address certain microinstruction sequences stored in MEM 10112. These microinstructions will then be executed from MEM 10112, rather than from FUSDT 11010. This operation allows certain microinstruction sequences, for example rarely used microinstruction sequences, to remain in MEM 10112, thus freeing AF 24220 and FUSITT 11012's address spaces from more frequently used SOPs.

As previously described AF 24220 is loaded, with SDPs, for SINTs currently being used by CS 10110 in executing user's programs, from Bits 16—31 of JPD Bus 10142 (JPD(16—31)). Also as previously discussed, addresses to load SDPs into AF 24220 are provided from LADDR 24214. LADDR 24214 is enabled to provide load addresses and AF 24220 to receive SDPs, by microinstruction control signals (not shown for clarity of presentation) provided from FUSITT 11012.

Referring finally to EUSDT 20266, SDPs may be provided to EU 10122 from 3 sources. EU 10122 SDPs may be provided from EUDISF 24222, from JPD Bus 10142 or from literal fields of microinstructions provided from FUSITT 11012. EUDISF 24222's SDPs are each 12 bits in width and comprise 9 bits of address into EUSITT and 3 bits of operand format information.

EUDISF 24222 is 1,024 word by 12 bit memory. As previously described addresses to read SDPs from EUDISF 24222 are provided from OPCODEREG 20268 by concatenating a 4 bit Dialect Code from RDIALL 24212 and an 8 bit SOP from LOPCODE 24210. SDPs provided by EUDISF 24222 are provided as a first input to EUDISS 24224.

EUDISS 24224 is a multiplexer. As just described, a first input of EUDISS 24224 are SDPs from EUDISF 24222. A second 12 bit input of EUDISS 24224 is provided from Bits 20 to 31 of JPD Bus 10142 (JPD(20—31)). A third input of EUDISS 24224 is a 12 bit input provided from a literal field of an FUSITT 11012 microinstruction output. EUDISS 24224 selects one of these 3 inputs to be transferred on EUDIS Bus 20206 to be provided as an execute unit SDP to EUSITT. Selection between EUDISS 24224's inputs is provided by microinstruction control signals (not shown for clarity of presentation) provided from FUSITT 11012.

As previously described, EUDISF 24222 is loaded, with SDPs for S-Language dialects currently being used by CS 10110, from Bits 20 to 31 of JPD Bus 10142 (JPD(20—31)). Addresses to load SDPs into EUDISF 24222 are provided, as previously described, from LADDR 24214. FUSITT 11012 provides enable signals (not shown for clarity of presentation) to LADDR 24214 and EUDISF 24222 to enable writing of SDPs into EUDISF 24222.

The structure and operation of FUCTL 20214 circuitry for fetching and parsing SINTs from MEM 10112 to provide Name syllables and SOPs, and for interpreting SOP to provide SDPs to FUSITT 11012 and EUSITT from FUSDT 11010 and EUSDT 20266, have been described above. As described above, SDPs provided by FUSDT 11010 and EUSDT 20266 are initial, or starting, addresses pointing to first microinstructions of sequences of microinstructions. Addresses for microinstructions following those initial microinstructions are provided by FUCTL 20214's next address generator circuitry which may include mPC 20276, BRCASE 20278, REPCTR 20280 and PNREG 20282, EVENT 20284 and SITTNAG 20286. mPC 20276, BRCASE 20278, REPCTR 20280 and PNREG 20282, and SITTNAG 20286 are primarily concerned with generation of next addresses during execution of microinstruction sequences in response to SOPs and will be described next below. EVENT 20284 and other portions of FUCTL 20214's circuitry are more concerned with generation of microinstruction sequences with regard to CS 10110's internal mechanisms operations and will be described in a later description. EU 10122 also includes next address generation circuitry and this circuitry will be described in a following description of EU 10122.

c.c.c. Next Address Generator 24310 (Fig. 243)

As stated above, in FU 10120 first, or initial, microinstructions of microinstruction sequences for interpreting SOPs are provided by FUSDT 11010. Subsequent addresses of microinstructions within these sequences are, in general, provided by mPC 20276 and BRCASE 20278. mPC 20276, as described further below, provides sequential addresses for selecting sequential microinstructions of microinstruction sequences. BRCASE 20278 provides addresses for selecting microinstructions when a microinstruction Branch or microinstruction Case operation is required. REPCTR 20280 and PNREG 20282 provide addresses for writing, or loading, of microinstruction sequences into FUSITT 11012. Other portions of FUCTL 20214 circuitry, for example EVENT 20284, provides microinstruction sequence selection addresses to select microinstruction sequences for controlling operation of CS 10110's internal mechanisms. SITTNAS 20286 selects between these microinstruction address sources to provide to FUSITT 11012 those addresses

required to select microinstructions of the operation to be currently executed by CS 10110.

Referring to Fig. 243, a partial block diagram of FU 10120's Next Address Generator (NAG) 24310 is shown. In addition to FUSDT 11010, NAG 24310 includes mPC 20276, BRCASE 20278, EVENT 20284, REPCTR 20280 and PNREG 20282, a part of RCWS 10358, and SITTNAS 20286. EVENT 20284 is, as described above, primarily concerned with execution of microinstruction sequences for controlling CS 10110 internal mechanisms. EVENT 20284 as shown herein only to illustrate its relationships to other portions of NAG 24310. EVENT 20284 will be described further in a following description of FUCTL 20214's circuitry controlling CS 10110's internal mechanisms. Similarly, operation of RCWS 10358 will be described in part in the present description of NAG 24310, and in part in a following description of control of CS 10110's internal mechanisms.

Referring first to NAG 24310's structure, interconnections of FUSDT 11010, RCWS 10358, mPC 20276, BRCASE 20278, REPCTR 20280, PNREG 20282, EVENT 20284, and SITTNAS 20286 have been previously described with reference to Fig. 202. NAG 24310's structure will be described below only wherein Fig. 243 differs from Fig. 202.

Referring first to SITTNAS 20286, SITTNAS 20286 is shown as comprised of EVENT Gate (EVNTGT) 24310 and Next Address Select Multiplexer (NASMUX) 24312. NASMUX 24312 is comprised of NAS Multiplexer A (NASMUXA) 24314, NASMUXB 24316, NASMUXC 24318, and NASMUXD 24320. Outputs of EVNTGT 24310 and NASMUXA 24314 to NASMUXD 24320 are ORed to CSADR 20204 to provide microinstruction selection addresses to FUSITT 11012.

ADR 20202 is shown in Fig. 243 as comprised of nine buses, Address A (ADRA) Bus 24322 to Address I (ADRI) Bus 24338. Output of EVENT 20284 is connected to input of EVNTGT 24310 by ADRA Bus 24322. Outputs of REPCTR 20280 and PNREG 20282 and output of AF 24220 are connected to inputs of NASMUXA 24314 by, respectively, ADRB Bus 24324 and ADC Bus 24326. Outputs of RCWS 10358 and FUDISENC 24219 are connected to inputs of NASMUXB 24316 by, respectively, ADRD Bus 24328 and ADRE Bus 24330. Outputs of BRCASE 20278 and second output of mPC 20276 are connected to inputs of NASMUXC 24318 by, respectively, ADRF Bus 24332 and ADRG Bus 24334. Second output of mPC 20276 and JAM output of NC 10226 are connected to inputs of NASMUXD 24320 by, respectively, ADRH Bus 24336 and ADRI Bus 24338. ADR 20202 thus comprises a set of buses connecting microinstruction address sources to inputs of SITTNAS 20286.

Referring to mPC 20276, mPC 20276 is comprised of Micro-Program Counter Counter (mPCC) 24340 and Micro-Program Counter Arithmetic and Logic Unit (mPCALU) 24342. Data input of mPCC 24340 is connected from CSADR Bus 20204. Output of mPCC 24340 is connected to a first input of mPCALU 24342 and is mPC 20276's third output to BRCASE 20278. Second input of mPCALU 24342 is a fifteen binary number set, for example by hard-wiring, to be binary one. Output of mPCALU 24342 comprises mPC 20276's first output, to RCWS 10358, and mPC 20276's second output, to inputs of NASMUXC 24318 and NASMUXD 24320.

BRCASE 20278 is shown in Fig. 243 as comprising Mask and Shift Multiplexer (MSMUX) 24344, Case Mask and Shift Logic (CASEMS) 24346, Branch and Case Multiplexer (BCMUX) 24348 and Branch and Case Arithmetic and Logic Unit (BCALU) 24350. A first input of MSMUX 24344 (AONBC, not previously shown) is connected from output of AONGRF 20232. A second input of MSMUX 24344 (OFFMUXR, not previously shown) is connected from output of OFFMUXR 23812. Output of MSMUX 24344 is connected to input CASEMS 24346, and output of CASEMS 24346 is connected to a first input of BCMUX 24348. A second input of BCMUX 24348, BLIT is connected from a literal field output of FUSITT 11012's microinstruction output. Output of BCMUX 24348 and third output of mPC 20276, from output of mPCC 24340, are connected, respectively, to first and second inputs of BCALU 24350. Output of BCALU 24350 comprises BRCASE 20278 outputs to NASMUXC 24318.

An address to select a next microinstruction may be provided to FUSITT 11012 by SITTNAS 20286 from any of eight sources. First source is output of mPC 20276. Output of mPC 20276 is referred to as Micro-Program Count Plus 1 (mPC+1) and is fifteen bits of address. Second source is from EVENT 20284 and is comprised of five bits of address. Third source is output of FUDISP 24218 and FUDISENC 24219 and, as previously described, is comprised of six bits of address. Fourth source is output of AF 24220 and, as previously described, is comprised of fifteen bits of address. Fifth source is output of BRCASE 20278. Output of BRCASE 20278 is referred to as Branch and Case Address (BRCASEADR) and comprises fifteen bits of address. Sixth source is an output of RCWS 10358. Output of RCWS 10358 is referred to as RCWS Address (RCWSADR) and is comprised of fifteen bits of address. Seventh source is REPCTR 20280 and PNREG 20282 whose outputs (REPPN) together comprise fifteen bits of address. Finally, eighth source is JAM input from NC 10226, which comprises five bits of address. These address sources differ in number of bits of address that they provide, but a microinstruction address gated onto CSADR Bus 20202 by SITTNAS 20286 always comprises fifteen bits of address. If a particular source applies fewer than fifteen bits, that address is extended to fifteen bits by SITTNAS 20286. In general, extension of address bits may be performed by hard-wiring of additional address input bits to SITTNAS 20286 from each of these sources and will be described further below.

Referring to mPC 20276, mPCC 24340 is a fifteen bit register and mPCALU 24342 is a fifteen bit ALU. mPCC 24340 is, as described above, connected from CSADR Bus 20204 and is sequentially loaded with a microinstruction address currently being presented to FUSITT 11012. mPCC 24340 will thus contain the

address of the currently executing microinstruction. mPCALU 24342 is dedicated to incrementing the address contained in mPCC 24340 by one. mPC+1 output of mPCALU 24342 will thereby always be address of next sequential microinstruction. mPC+1 is, as described above, a fifteen bit address and is thus not extended in SITTNAS 20286.

5 Referring to BRCASE 20278, as described above BRCASE 20278 provides next microinstruction addresses for mPC 20276 Relative Branches and for Case Branches. Next microinstruction addresses for microprogram Relative Branches and for Case Branches are both generated as addresses relative to address of currently executing microinstruction as stored in mPCC 24340, but differ in the manner in which these relative addresses are generated. Considering first Case Branches, Case Branch addresses relative to
 10 a currently executing microinstruction address are generated, in part, by MSMUX 24344 and CASEMS 24346. As described above, MSMUX 24344 which is a multiplexer receives two inputs. First input is AONBC from output of AONGRF 20232 and second input is OFFMUXR from output of OFFMUXR 23812. Each of these inputs is eight bits, or one byte, in width. Acting under control of microinstruction output from FUSITT 11012, MSMUX 24344 selects either input AONBC or input OFFMUXR as an eight bit output to input
 15 of CASEMS 24346. CASEMS 24346 is a Mask and Shift circuit, similar in structure and operation to that of FIU 20116 but operating upon bytes rather than thirty-two bit words. CASEMS 24346, operating under microinstruction control from FUSITT 11012, manipulates eight bit input from MSMUX 24344 by masking and shifting to provide eight bit Case Value (CASEVAL) output to BCMUX 24348. CASEVAL represents a microinstruction address displacement relative to address of a currently executing microinstruction and,
 20 being an eight bit number, may express a displacement of 0 to 255 address locations in FUSITT 11012.

BCMUX 24348 is an eight bit multiplexer, similar in structure and operation to MSMUX 24344, and is controlled by microinstruction inputs provided from FUSITT 11012. In executing a case operation, BCMUX 24348 selects CASEVAL input to MCMUX 24348's output to first input of BCALU 24350. BCALU 24350 is a
 25 sixteen bit arithmetic and logic unit. Second input of BCALU 24350 is fifteen bit address of currently executing microinstruction from mPCC 24340. BCALU 24350 operates under microinstruction control provided from FUSITT 11012 and, in executing a Case operation, adds CASEVAL to the address of a currently executing microinstruction. During a Case operation, carry input of BCALU 24350 is forced, by microinstruction control from FUSITT 11012, to one so that BCALU 24350's second input is effectively
 30 mPC+1, or address of currently executing microinstruction plus 1. Output BRCASEADR of BCALU 24350 will thereby be fifteen bit Case address which is between one and 256 FUSITT 11012 address locations higher than the address location of the currently executing microinstruction. The actual case value address displacement from the address of the currently executing microinstruction is determined by either input AONBC or input OFFMUXR to MSMUX 24344, and these mask and shift operations are performed by CASEMS 24346.

35 Case operations as described above may be used, for example, in interpreting and manipulating CS 10110 table entries. For example, Name Table Entries of Name Tables 10350 contain flag fields carrying information regarding certain operations to be performed in resolving and evaluating those Name Table Entries. These operations may be implemented as Case Branches in microinstruction sequences for resolving and evaluating those Name Table Entries. In the present example, during resolve of a Name
 40 Table Entry the microinstruction sequence for performing that resolve may direct a byte of that Name Table Entry's flag field to be read from AONGRF 20232, or OFFMUXR 23812, and through MSMUX 24344 to CASEMS 24346. That microinstruction sequence will then direct CASEMS 24346 to shift and mask that flag field byte to provide a CASEVAL. That CASEVAL will have a value dependent upon the flags within that flag field byte and, when added to mPC+1, will provide a FUSITT 11012 microinstruction address for a
 45 microinstruction sequence for handling that Name Table Entry in accordance with those flag bits.

As described above, BRCASE 20278 may also generate microinstruction addresses for Branches occurring within execution of a given microinstruction sequence. In this case, microinstruction control signals from FUSITT 11012 direct BCMUX 24348 to select BCMUX 24348's second input as output to BCALU 24350. BCMUX 24348's second input is Branch Literal (BLIT). As described above, BLIT is provided from a
 50 literal field of a microinstruction word from FUSITT 11012's microinstruction output. BLIT output of BCMUX 24348 is added to address of currently executing microinstruction from mPCC 24340, and BCALU 24350, to provide fifteen bit BRCASEADR of a microinstruction address branched to from the address of the currently executing microinstruction. BRCASEADR may represent, for example, any of four Branch Operations. Possible Branch Operations are: first, a Conditional Short Branch; second, a Conditional Short Call; third, a
 55 Long Go To; and, fourth, a Long Call. In each of these possible Branch Operations, BLIT is treated as the two's complement of the desired branch value, that is the microinstruction address offset relative to the address of the currently executing microinstruction. BLIT field may therefore be, effectively, added to or subtracted from the address of the currently executing microinstruction, to provide a microinstruction address having a positive or negative displacement from the address of the currently executing
 60 microinstruction. In a Conditional Short Branch or a Conditional Short Call, the fourteen bit literal field is a sign extended eight bit number. Both Conditional Short Branch and Conditional Short Call microinstruction addresses may therefore point to an address within a range of +127 to -128 FUSITT 11012 address locations of the address of the currently executing microinstruction. In the case of a Long Go To or Long Call, the BLIT field is a fourteen bit number representing displacement relative to the address of the
 65 currently executing microinstruction. BRCASEADR may, in these cases, represent a FUSITT 11012

microinstruction address within a range of +8191 to -8192 FUSITT 11012 address locations of the address of the currently executing microinstruction. BRCASE 20278 thereby provides FU 10120 with capability of executing a full range of microinstruction sequence Case and Branch operations.

5 Referring to RCWS 10358, as previously described RCWS 10358 stores information regarding microinstruction sequences whose execution has been halted. RCWS 10358 allows execution of those microinstruction sequences to be resumed at a later time. A return control word (RCW) may be written onto RCWS 10358 during any microinstruction sequence that issues a Call to another microinstruction sequence. The calling microinstruction sequence may, for example, be aborted to service an event, as described further in a following description, or may result in a Jam. A Jam is a call for a microinstruction sequence which is forced by operation of CS 10110 hardware, rather than by a microinstruction sequence. 10 RCWS 10358 operation with regard to CS 10110's internal mechanisms will be described in a following description of EVENT 20284, STATE 20294, and MCW1 20290 and MCWO 20292. For purposes of the present discussion, that portion of a RCW concerned with interpretation of SOPs contains, first, certain state information from FUSITT 11012 and, second, a return address into FUSITT 11012. State that FUSITT 15 11012 state is provided from STATE 20294, as described below, and that portion of a RCW containing FUSITT 11012 state information will be described in a following description. Microinstruction address portions of RCWs are provided from output of mPCALU 24342. This microinstruction address is the address of the microinstruction to which FU 10120 is to return upon return from a Call, Event, or Jam. Upon occurrence of a Call or Jam, the microinstruction return address is mPC+1, that is the address of the microinstruction after the microinstruction issuing the Call or Return. For aborted microinstruction sequences, the microinstruction return address is mPC, that is the address of the microinstruction 20 executing at the time abort occurs.

Upon return from a call, service of an event, or service of a jam, FU 10120 state flag portion of RCW is loaded into STATE 20294. Microinstruction return address is provided by RCWS 10358 as fifteen bit 25 RCWSADR to SITTNAS 20286 and is gated onto CSADR 20204. RCWSADR is provided to FUSITT 11012 to select the next microinstruction and is loaded into mPCC 24340 from CSADR 20204.

As previously described, RCWS 10358 is connected to JPD Bus 10142 by a bi-directional bus. RCWs may be written into RCWS 10358 from JPD Bus 10142, or read from RCWS 10358 to JPD Bus 10142. The fifteen bit next microinstruction address portion, and the single bit FUSITT 11012 state portion of RCW is 30 written from or read to Bits 16 to 31 of JPD Bus 10142. FU 10120 may write Present Bottom RCW or Previous RCW into RCWS 10358 from JPD Bus 10142 and may read Present Bottom RCW, or Previous RCW, or another selected RCW, onto JPD Bus 10142. RCWS 10358 thereby provides a means for storing and returning microinstruction addresses of microinstruction sequences whose execution has been suspended, and a means for writing and reading microinstruction address, and FUSITT 11012 state flags, 35 from and to JPD Bus 10142.

As previously described, REPCTR 20280 and PNREG 20282 provide microinstruction addresses for writing of microinstructions into FUSITT 11012. REPCTR 20280 is an eight bit counter and PNREG 20282 is a seven bit register. Eight bit output of REPCTR 20280 is left concatenated with seven bit output of PNREG 20282 to provide fifteen bit microinstruction addresses REPPN. That is, REPCTR 20280 provides the eight 40 low order bits of microinstruction address while PNREG 20282 provides the seven most significant bits of address.

REPCTR may be loaded from Bits 24-31 of JPD Bus 10142, and may be read to Bits 24-31 of JPD Bus 10142. In addition, the eight bits of microinstruction address in REPCTR 20280 may be incremented or decremented as microinstructions are written into FUSITT 11012.

45 As described above, PNREG 20282 contains the seven most significant bits of microinstruction address. These address bits may be written into PNREG 20282 from Bits 17-23 of JPD Bus 10142. Contents of PNREG 20282 may not, in general, be read to JPD Bus 10142 and may not be incremented or decremented.

Referring to JAM input to SITTNAS 20286 from NC 10226, certain Name evaluate or resolve operations 50 may result in jams. A Jam functions as a call to microinstruction sequences for servicing Jams and are forced by FU 10120 hardware circuitry involved in Name syllable evaluates and resolves.

JAM input to SITTNAS 20286 is comprised of six Jam address bits. Three bits are provided by NC 10226 and three bits are provided from FUSITT 11012's microinstruction output as part of microinstruction sequences for correcting Name syllable evaluates and resolves. The three bits of address from NC 10226 55 form the most significant three bits of JAM address. One of these bits gates JAM address onto CSADR Bus 20204 and is thus not a true address bit. Output of FUSITT 11012 provides the three least significant bits of JAM address and specifies the particular microinstruction sequence required to service the particular Jam which has occurred. Therefore, during Name evaluate or resolves, the microinstruction sequences provided by FUSITT 11012 to perform Name evaluates or resolves specifies what microinstruction sequences are to be initiated if a Jam occurs. The three bits of JAM address provided by NC 10226 60 determine, first, that a Jam has occurred and, second, provide two bits of address which, in combination with the three bits of address from FUSITT 11012, specify the particular microinstruction sequence for handling that Jam. JAM address inputs from NC 10226 and from FUSITT 11012 thereby provide six of the fifteen bits of JAM address. The remaining nine bits of JAM address are provided, for example, by hard-wired inputs to NASMUXD 24320. These hard-wired address bits force JAM address to address FUSITT 65

EP 0 067 556 B1

11012 in blocks of 4 microinstruction addresses, in a manner similar to address inputs to FUDISF 24218 and FUDISENC 24219.

Address inputs provided to SITTNAS 20286 from FUSDT 11010 have been previously described with respect to description of FUCTL 20214 fetch, parse, and dispatch operations. Address inputs provided by
5 EVENT 20284 will be described in a following description of FUCTL 20214's operations with regard to CS 10110's internal mechanisms.

Referring finally to SITTNAS 20286, as previously described SITTNAS 20286 is comprised of EVNTGT 24310 and NASMUX 24312. Inputs are provided to NASMUX 24312, as described above, from FUSDT 11010, mPC 20276, BRCASE 20278, RCWS 10358, REPCTR 20280 and PNREG 20282, and by JAM input.
10 These inputs are, in general, provided with regard to FUCTL 20214's operations in fetching, parsing, and interpreting SOPs and Name syllables. These operations are thereby primarily directly concerned with execution of user's programs, that is the execution of sequences of SINS. NASMUX 24312 selects between these inputs and transfers selected address inputs onto CSADR 20204 as microinstruction addresses to FUSITT 11012 under microinstruction control from microinstruction outputs of FUSITT 11012.
15 Microinstruction address outputs are provided to SITTNAS 20286 from EVENT 20284 in response to Events, described further below, occurring in CS 10110's operations in executing user's programs. These microinstruction addresses from EVENT 20284 are gated onto CSADR 20204, to select appropriate microinstruction sequences, by EVNTGT 24310. EVNTGT 24310 is separated from NASMUX 24312 to allow EVNTGT 24310 to over-ride NASMUX 24312 and provide microinstruction address to EVENT 20284 while
20 NASMUX 24312 is inhibited due to occurrence of certain Events. These Events are, in general, associated with operation of CS 10110's internal mechanisms and structure and operation of EVENT 20284, together with STATE 20294, MCW1 20290, and MCWO 20292, and other portions of RCWS 10358, will be described next below.

c.c. FUCTL 20214 Control Circuitry for CS 10110 Internal Mechanisms (Figs. 244—249)

Certain portions of FUCTL 20214's Control Circuitry are more directly concerned with operation of CS 10110's internal mechanisms, for example CS 10110 Stack Mechanisms. This circuitry may include STATE 20294, EVENT 20284, MCW1 20290 and MCWO 20292, portions of RCWS 10358, REG 20288, and Timers 20296. These FUCTL 20214 control elements will be described next below, beginning with STATE 20294.
30

a.a.a. State Logic 20294 (Figs. 244A—244Z)

In general, all CS 10110 operations, including execution of microinstructions, are controlled by CS 10110's Operating State. CS 10110 has a number of Operating States, hereafter referred to as States, each State being defined by certain operations which may be performed in that State. Each of these States will
35 be described further below. Current State of CS 10110 is indicated by a set of State Flags stored in a set of registers in STATE 20294. Each State is entered from previous State and is exited to a following State. Next State of CS 10110 is detected by random logic gating distributed throughout CS 10110 to detect certain conditions indicating which State CS 10110 will enter next. Outputs of these Next State Detection gates are provided as inputs to STATE 20294's registers. A particular State register is set and provides a State Flag output when CS 10110 enters the State associated with that particular register. State Flag outputs of STATE
40 20294's state registers are provided as enable signals throughout CS 10110 to enable initiation of operations allowed within CS 10110's current State, and to inhibit initiation of operations which are not allowed within CS 10110's current State.

Certain of CS 10110's States, and associated STATE 20294 State Registers and State Flag outputs, are:

(1) MO: the initial State of any microinstruction.

State MO is always entered as first data cycle of every microinstruction. During MO, CS 10110's State may not be changed, thus allowing a microinstruction to be arbitrarily aborted and restarted from State MO. In normal execution of microinstructions, State MO is followed by State M1, described below, that is, State MO is exited to State M1. State MO may be entered from State M0 and from State M1, State AB, State
50 LR, State NR, or State MS, each of which will be described below.

(2) EP: Enable Pause State. State EP is entered when State MO is entered for the first time in a microinstruction. If that microinstruction requests a pause, that microinstruction will force State MO to be re-entered for one clock cycle. If State MO lasts more than one clock cycle, State EP is entered on each extension of State MO unless the extension is a result of a pause request.

(3) SR: Source GRF State. SR State is active for one clock cycle wherein SR State register enables loading of a GRF 10354 output register. State SR is re-entered on every State MO cycle except a State MO cycle generated by a microinstruction requesting extension of State MO. When all STATE 20294 State Registers are cleared, DP 20218 may set state SR register alone, for purposes of reading from GRF 10354.

(4) M1: Final state of normal microinstruction execution. State M1 is the exit State of normal microinstruction execution. FUSITT 11012 microinstruction register, described below, is loaded with a next microinstruction upon exit from State M1. In addition, State M1 Flag output of STATE 20294 enables all CS 10110 registers to receive data on their inputs, that is data on inputs of these registers are clocked to outputs of these registers. State M1 may be entered from State M1, or from State M0, State MW, State
60 MWA, or State WB.

(5) LA: Load Accumulator Enable State. State LA is entered, upon exit from State M1, by
65

EP 0 067 556 B1

microinstructions which read data from MEM 10112 to OFFMUXR 23812. As previously described, OFFMUXR 23812 serves as a general purpose accumulator for DESP 20210. STATE LA overlaps into execution of next microinstruction, and persists until data is returned from MEM 10112 in response to a request to MEM 10112. When MEM 10112 signals data is available, by asserting DAVFA, LA State Flag enables loading of data into OFFMUXR 23812. If the next microinstruction references OFFMUXR 23812, that microinstruction execution is deferred until a read to OFFMUXR 23812 is completed, as indicated by CS 10110 exiting from State LA.

(6) RW: Load GRF 10354 Wait State. State RW is entered from State M1 of microinstructions which read data from MEM 10112 to GRF 10354. RW Flag inhibits initiation of a next microinstruction, that is prevents entry to State M0, and persists through the CS 10110 clock cycle during which data is returned from MEM 10112 in response to a request. State RW initiates Load GRF Enable State, described below.

(7) LR: Load GRF Enable State. State LR is entered in parallel with State RW, on last clock cycle of RW, and persists for one CS 10110 clock cycle. LR Flag enables writing of MEM 10112 output data into GRF 10354.

(8) MR: Memory Reference Trailer State. State MR is entered on transition to State M0 whenever a previous microinstruction makes a logical or physical address reference to MEM 10112. MR Flag enables recognition of any MEM 10112 reference Events, described below, which may occur. State MR persists for one clock cycle. If a MEM 10112 memory reference Event occurs, that Event forces exit from State MR to States AB and MA, otherwise State MR has no effect upon selection next state.

(9) SB: Store Back Enable State. State SB is entered during State M0 of a microinstruction following a microinstruction which generated a store back of a result of a EU 10122 operation. SB Flag gates that result to be written into MEM 10112 through JPD Bus 10142.

(10) AB: Microinstruction Abort State. State AB is entered from first M0 State after an Event request is recognized, as described in a following description.

State AB may be entered from State M0 or from State AB and suppresses an entry into State M1. If there has been an uncompleted reference to MEM 10112, that is, the reference has not been aborted and data has not returned from MEM 10112, JP 10114 remains in State AB until the MEM 10112 reference is completed. Should an abort have occurred due to a MEM 10112 reference Event, State AB lasts two clock cycles only. As will be described in a following description of EVENT 20284, State M0 of a first microinstruction of a Handler for an Event causing an abort is entered from State AB. AB Flag gates the Handler address of the highest priority recognized Event onto CSADR Bus 20204 to select a corresponding Event Handler microinstruction sequence. EVENT 20284 is granted control of CSADR Bus 20204 during all State AB clock cycles.

(11) AR: Microinstruction Abort Reset State. State AR is entered in parallel with first clock cycle of State AB and persists for one clock cycle. AR Flag resets various STATE 20294 State Registers when an abort occurs. If there are no uncompleted MEM 10112 references, next State AB clock cycle is the last. On uncompleted MEM 10112 references, State AR is entered, but State AB remains active until reference is complete. Should a higher priority Event request service and be recognized while JP 10114 is in State AB, State AR is reentered. State AB will thereby be active for two clock cycles during all honored Event requests.

(12) MA: MEM 10112 Reference Abort. State MA is entered in parallel with State AB if a MEM 10112 reference is aborted, as indicated by asserted ABORT control signal output from MEM 10112. State MA persists for one clock cycle and State AB flag generates a MEM 10112 Reference Abort Flag which, as described below, results in a repeat of the MEM 10112 reference. AB Flag also resets MEM 10112 Trailer States, described below.

(13) NW: Nano-interrupt Wait State. State NW is entered from State M0 of a microinstruction which issues a Nano-interrupt Request to EU 10122 for an EU 10122 operation. FU 10120 remains in State NW until EU 10122 acknowledges that interrupt. Various EU 10122 Events may make requests at this time. State NW is exited into State AB or State M1.

(14) FM: First Microinstruction of a SIN. State FM is entered in parallel with State M0 on first microinstruction of each SIN and persists for one clock cycle. FM Flag inhibits premature use of AF 24220 and enables recognition of SIN Entry Events. State FM is re-entered upon return from all aborts taken during State M0 of the first microinstruction of an SIN.

(15) SOP: Original Entry to First SIN. State SOP is entered upon entry to State M0 of the first microinstruction of an SOP and is exited from upon any exit from that microinstruction. State SOP is entered only once for each SOP. SOP Flag may be used, for example, for monitoring performance of JP 10114.

(16) EU: EU 10122 Operand Buffer Unavailable. State EU is entered from State M0 of a microinstruction which attempts to read data to EU 10122 Operand Buffer, described in a following description, wherein EU 10122 Operand Buffer is full. When a new SOP is entered, three fetches of data from MEM 10112 may be performed before EU 10122 Operand Buffer is full; two fetches will fill EU 10122 Operand Buffer but EU 10122 may take one operand during a second fetch, thereby clearing EU 10122 Operand Buffer space for a third operand.

(17) NR: Long Pipeline Read. Entry into State NR disables overlap of MEM 10112 reads and disables execution of the next microinstruction. A following microinstruction does not enter State M0 until

requested data is returned from MEM 10112. State NR is entered from State NR or from State M1.

(18) NS: Nonpipeline Store Back. State NS is entered in parallel with State SB whenever a microinstruction requesting a pipeline store back, or a write to MEM 10112, occurs. State NS flag generates entry into State M0 of a following microinstruction upon exit from State SB.

5 (19) WA: Load Control Store State A. State WA is entered from State M0 of a microinstruction which directs loading of microinstruction into FUSITT 11012. WA State Flag controls selection of addresses to CSADR Bus 20204 for writing into FUSITT 11012, and generates a write enable pulse to FUSITT 11012 to write microinstructions into FUSITT 11012.

10 (20) WB: Load Control Store State B. State WB is entered from State WA and is used to generate an appropriate timing interval for writing into FUSITT 11012. State WB also extends State M1 to 2 clock cycles to ensure a valid address input to FUSITT 11012 when a next microinstruction is to be read from FUSITT 11012.

15 Having described certain CS 10110 states, and operations which may be performed within those states, state sequences for certain CS 10110 operations will be described next below with aid of Figs. 244A to 244Z. Fig. 244A to Fig. 244Z represent those state timing sequences necessary to indicate major features of CS 10110 state timing. All state timing shown in Figs. 244A to 244V assumes full pipelining of CS 10110 operations, for example pipelining of reads from and writes to MEM 10112 by JP 10114. Pipelining is not assumed in Figs. 244W to 244Z. Referring to Figs. 244A to 244Z, these figures are drawn in the form of timing diagrams, with time increasing from left to right. Successive horizontally positioned "boxes" 20 represents successive CS 10110 states during successive CS 10110 110 nano-second clock cycles. Vertically aligned "boxes" represent alternate CS 10110 states which may occur during a particular clock cycle. Horizontally extended dotted lines connecting certain states represented in Fig. 244A to 244Z represent an indeterminate time interval which is an integral multiple of 110 nano-second CS 10110 clock cycles.

Referring to Fig. 244A to 244Z in sequence, State Timing Sequences shown therein represent:

25 (1) Fig. 244A; state timing for execution of a normal microinstruction with no Events occurring and no MEM 10112 references.

(2) Fig. 244B execution of a normal microinstruction, with no Events occurring, no MEM 10112 references, and a hold in State M0 for one clock cycle.

30 (3) Fig. 244C; a microinstruction requests an extension of State M0 for one clock cycle, with no Events occurring and no MEM 10112 references.

(4) Fig. 244D; a write to MEM 10112 from DESP 20210, for example from GRF 10354 or from OFFALU 20242. MEM 10112 port is available and MEM 10112 reference is made during first sequential occurrence of States M0 and M1.

35 (5) Fig. 244E; a write to MEM 10112 from DESP 20210 as described above. MEM 10112 port is unavailable for an indeterminate number of clock cycles. A MEM 10112 reference is made during first sequential occurrence of States M0 and M1.

(6) Fig. 244F; writing of an EU 10122 result back into MEM 10112. MEM 10112 is available and a write operation is initiated during first sequential occurrence of States M0 and M1.

40 (7) Fig. 244G; writing back of an EU 10122 result to MEM 10112 as described above. MEM 10112 port is unavailable for an undetermined number of clock cycles, or EU 10122 does not have a result ready to be written into MEM 10112. Write operation is initiated during first sequential occurrence of States M0 and M1.

(8) Fig. 244H; a read of an EU 10122 result into FU 10120. EU 10122 result is not available for an undetermined number of clock cycles.

45 (9) Fig. 244I; a read from MEM 10112 to OFFMUXR 23812, with no delays. The microinstruction following the microinstruction initiating a read from MEM 10112 does not reference OFFMUXR 23812.

(10) Fig. 244J; a read from MEM 10112 to OFFMUXR 23812 with data from MEM 10112 being delayed by an indeterminate number of clock cycles. The next following microinstruction from that initiating the read from MEM 10112 does not reference OFFMUXR 23812.

50 (11) Fig. 244K; a read from MEM 10112 to OFFMUXR 23812. The next microinstruction following the microinstruction initiating the read from MEM 10112 references OFFMUXR 23812.

(12) Fig. 244L; a read from MEM 10112 to GRF 10354. The read to GRF 10354 is initiated by the first sequentially occurring States M0 and M1.

(13) Fig. 244M; a read from MEM 10112 to GRF 10354 and to OFFMUXR 23812. In this case, read operations may not be overlapped.

55 (14) Fig. 244N; JP 10114 honors an Event request and initiates a corresponding Event Handler microinstruction sequence, no MEM 10112 references occur.

(15) Fig. 244O; JP 10114 honors an Event request as stated above. MEM 10112 references are made during the first sequential occurrence of States M0 and M1 and a MEM 10112 reference Event occurs. In case of an MEM 10112 reference event, State MA is entered from one clock cycle. This occurs only if a MEM 10112 reference is made and aborted.

60 (16) Fig. 244P; an Event occurs in a MEM 10112 reference made during the first sequential occurrence of States M0 and M1. The MEM 10112 reference does not result in a memory reference Event. CS 10110 remains in State AB until the MEM 10112 reference is completed by return of data from MEM 10112.

65 (17) Fig. 244Q; a read of data from MEM 10112 or JPD Bus 10114 to EU 10122 Operand Queue. EU 10122 Operand Queue is not full.

(18) Fig. 244R; a read of MEM 10112 or JPD Bus 10142 data to EU 10122 Operand Queue. EU 10122 Operand Queue is full when the microinstruction initiating the read is issued.

(19) Fig. 244S; a request for a "nano-interrupt" to EU 10122 by FU 10120 with no Events occurring.

5 (20) Fig. 244T; FU 10120 submits a "nano-interrupt" request to EU 10122 and an EU 10122 State Overflow, described further in a following description, occurs. No other Events are recognized, as described in a following description of EVENT 20284.

(21) Fig. 244U; FU 10120 submits a "nano-interrupt" request to EU 10122. Another Event is recognized during State M0 and an abort results. First abort state is entered for the non-EU 10122 event. All aborts recognized in State M0 are taken or acknowledged, before entrance into State M0. Therefore, on retry at 10 State M0 of the original microinstruction entered from State M0, next abort recognized is for EU 10122 Stack Overflow Event since EU 10122 Stack Overflow has higher priority.

(22) Fig. 244V; a load of a 27 bit microinstruction segment into FUSITT 11012.

In Figs. 244A to 244V, pipelining MEM 10112 reads and writes, and of JP 10114 operations, has been assumed. In Figs. 244W to 244Z, non-overlapping operation of JP 10114 is assumed.

15 (23) Fig. 244W; a read of data from MEM 10112 to OFFMUXR 23812.

(24) Fig. 244X; a read of data from MEM 10112 to EU 10122 Operand Queue.

(25) Fig. 244Y; a write of an EU 10122 result into MEM 10112.

(26) Fig. 244Z; a read of a 32 bit SIN word from MEM 10112 in response to a prefetch or conditional 20 prefetch request.

Having described the general structure and operation of STATE 20294, and the operating states and operations of CS 10110, structure and operation of EVENT 20284 will be described next below.

b.b.b. Event Logic 20284 (Figs. 245, 246, 247, 248)

25 An Event is a request for a change in sequence of execution of microinstructions which is generated by CS 10110 circuitry, rather than by currently executing microinstructions. Occurrence of an Event will result in provision of a microinstruction sequence, referred to as an Event Handler, by FUSITT 11012 which modifies CS 10110's operations in accordance with the needs of that Event. Event request signals may be generated by CS 10110 circuitry internal to JP 10114, that is from FU 10120 or EU 10122 or CS 10110 circuitry external to JP 10114, for example from IOP 10116 or from MEM 10112. Event request signals are 30 provided as inputs to EVENT 20284. As will be described further below, EVENT 20284 masks Event Requests to determine which Events will be recognized during a particular CS 10110 Operating State, assigns priorities for servicing multiple Event Requests, and fabricates Handler addresses to FUSITT 11012 for microinstruction sequences for servicing requests. EVENT 20284 then provides those Handler microinstruction addresses to FUSITT 11012 through EVNTGT 24310, to initiate execution of selected Event 35 Handler microinstruction sequences.

Certain terms and expressions are used throughout the following description. The following paragraphs define these usages and provide examples illustrating these terms. An Event "makes a request" when a condition in CS 10110 hardware operation results in a Event Request signal being provided to EVENT 20284. As will be described further below, these Event Request signals are provided to 40 EVENT 20284 combinatorial logic which determines the validity of those "requests".

An Event Request "is recognized" if it is not masked, that is inhibited from being acted upon. Masking may be explicit, using masks generated by FUSITT 11012, or may be implicit, resulting from an improper CS 10110 State or invalid due to other considerations. That is, certain Events are recognized only during certain CS 10110 States even though those requests may be recognized during certain other states. Any 45 number of requests, for example up to 31, may be simultaneously recognized.

An Event Request is "honored" if it is the highest priority Event Request occurring. When a request is honored, a corresponding address, of a corresponding microinstruction sequence in FUSITT 11012, for its Handler microinstruction sequence is gated onto CSADR Bus 20204 by EVENT 20284. A request is honored when CS 10110 enters State AB. State AB gates the selected Event Handler microinstruction address on 50 CSADR Bus 20284.

To summarize, a number of Events may request service by JP 10114. Of these Events, all, some, or none, may be recognized. Only one Event Request, the highest priority Event Request, will be honored when JP 10114 enters State AB. Microinstruction control of CS 10110 will then transfer to that Event's Handler microinstruction sequence. A necessary condition for entering State AB is that an Event Request 65 has been made and recognized.

A microinstruction sequence "completes", "is completed", or reaches "completion" when CS 10110 exits State M1 while that microinstruction sequence is active. A microinstruction sequence may, as described above, be aborted in State M0 an indefinite number of times before, if ever, reaching completion.

60 A MEM 10112 reference "completes", "is completed", or reaches "completion" when requested data is returned to the specified destination, that is read from MEM 10112 to the requestor, or MEM 10112 accepts data to be written into MEM 10112.

"Trace Traps" are an inherent feature of microinstructions being executed. Trace Traps occur on every microinstruction of a given type (if not masked), for example during a sequence of microinstructions to perform a Name evaluate or resolve, and occur on each microinstruction of the sequence. In general, a 65 Trace Trap Event must be serviced before execution of the next microinstruction. Trace Traps are distinct

from Interrupts in that an Interrupt, described below, does not occur on execution of each microinstruction of a microinstruction sequence, but only on those microinstructions where certain other conditions must be considered.

"Interrupts" are the largest class of events in JP 10114. Occurrence of an Interrupt may not, in general, be predicted for a particular execution of a particular microinstruction in a particular instance. Interrupts may require service before execution of the next microinstruction, before execution of the current microinstruction can complete, or before beginning of the next SIN. An Interrupt may be unrelated to execution of any microinstruction, and is serviced before beginning of the next microinstruction.

A "Machine Check" is an Event that JP 10114 may not handle alone, or whose occurrence makes further actions by JP 10114 suspect. These events are captured in EVENT 20284 Registers and result in a request to DP 10118 to stop operation of JP 10114 for subsequent handling.

In summary, three major classes of Events in CS 10110 are Trace Traps, Interrupts, and Machine Checks. Each of these class of events will be described in further detail below, beginning with Trace Traps.

The State of all possible Trace Trap Event Requests, whether requesting or not requesting, is loaded into EVENT 20284 Registers at completion of State M1 and at completion of State AB. That is, since Trap Requests are a function of the currently executing microinstruction, the State of a Trap Request will be loaded into EVENT 20284 Trace Trap Registers at end of State M1 of each currently executing microinstruction. Similarly, if any Trap Requests are recognized, State AB will be entered at the end of the first clock cycle of the next following State M0 and their State loaded at end of the State AB.

Recognized, or unmasked, Trap Requests may be pushed onto RCWS 10358 as Pending Requests. Unrecognized, or masked, Trace Trap Requests may be pushed onto RCWS 10358 as Not Pending Requests and are subsequently disregarded. Subsequently, when a microinstruction sequence ends in a return to a calling microinstruction sequence, the Trace Trap Request bits in an RCWS 10358 may be used to generate Trace Trap Event Requests.

Upon exit from State AB, all Trace Trap Requests, except Micro-Break-Point and Microinstruction Trace Traps, described below, are loaded into corresponding EVENT 20284 Trace Trap Request Registers as not requesting. Micro-Break-Point and Microinstruction Trace Traps, are, in general, always latched as requesting at completion of State AB. Trace Traps may be explicitly masked by a Trace Mode Mask, an Indivisibility Mode Mask, and by a Trace Enable input, all generated by FUSITT 11012 as described below. Micro-Break-Point Trap may also be masked by clearing a Trace Enable bit in a Trace Enable field of certain microinstructions containing Trace Traps. In general, masking is effective from State M0 of the microinstruction which generates the mask, through completion of a microinstruction which clears the mask Trace Traps generated by a microinstruction which clears a mask are taken so as to abort a following microinstruction during its M0 State.

Referring to Fig. 245, CS 10110 state timing for a typical Trap Request, and generation of a microinstruction address to a corresponding Trace Trap Handler microinstruction sequence by EVENT 20284 is shown. Fig. 245 is drawn using the same conventions as described above with reference to Fig. 244A to 244Z. In Fig. 245, a microinstruction executing in States M0 and M1 causes a Trace Trap Request but does not generate an MR (Memory Reference) Trailer State. Trace Trap Request to EVENT 20284 is signaled by Time A. This Trace Trap Request is latched into EVENT 20284 Trace Trap Event Registers, and an Abort Request is provided to STATE 20294. At Time B, FU 10120 enters States AB and AR. The microinstruction address for a Handler microinstruction sequence of the highest priority Event present in EVENT 20284 is presented to FUSITT 11012 and execution of the addressed microinstruction sequence begins. At Time C, FU 10120 exits States AB and AR and enters State AB. State AB will be exited at end of the next 110 nanosecond clock cycle. Address of the selected Event Handler microinstruction sequence will remain on CSADR Bus 20204 for duration of State AB. At Time D, a pointer into RCWS 10358, described in a following description, is incremented, thereby effectively pushing the first microinstruction's return control word, that is the microinstruction executing at first State M0, onto RCWS 10358. First microinstruction of the Trace Trap Event Handler microinstruction sequence is provided by FUSITT 11012. Execution of Handler microinstruction sequence will begin at start of the third State M0 of the state timing sequence shown in Fig. 245. EVENT 20284's Trace Trap Register for this event is now latched in nonrequesting state and will remain so until transition out of second State M1 shown in Fig. 245. At this time, EVENT 20284 Registers will latch new Trap Requests. Finally, at Time E, Trace Trap Event Registers of EVENT 20284 are latched with new Trap Requests arising from execution of the microinstruction being executed in States M0 and M1 occurring between Times D and E. Traps due to the microinstruction that was executed in States M0 and M1 before Time A, but were not serviced, are requested again when the previously pushed RCW described above is returned from RCWS 10358 upon return from the Trace Trap Event Handler microinstruction sequence initiated at Time D. All Trace Trap Requests which have been serviced are explicitly cleared in RCWS 10358 RCWs by their Event Handler microinstruction sequences to prevent recurrence of those Trap Requests. Since Trace Trap Event Requests arising from reads or writes to MEM 10112 will recur if those requests are repeated, EVENT 20284 generates memory repeat Interrupts after all aborted MEM 10112 read and write requests to insure that these Traps will eventually be serviced. Event Handler microinstruction sequences for these read and write Trace Trap Events explicitly disable serviced Trace Trap Event Requests by clearing bits in the logical descriptor of the aborted memory read and write requests.

Having described overall structure and operation of Trace Trap Events, certain specific Trace Trap Events will be described in greater detail below. Trace Trap Events occurring in CS 10112 may include Name Trace Traps, SOP Trace Traps, Microinstruction Trace Traps, Micro-Break-Point Trace Traps, Logical Write Trace Traps, Logical Read Trace Traps, UID Read Trace Traps, and UID Write Trace Traps. These

5 Trace Traps will be described below in the order named.

A Name Trace Trap is requested upon every microinstruction sequence that contains an evaluate or resolve of a Name syllable. Name Trace Traps are provided by decoding certain microinstruction fields of those microinstruction sequences. Name Trace Trap field is masked by either Trace Mask, Indivisibility Mask, or Trace Enable, as described above. All of these masks are set and cleared by microinstruction

10 control signals provided during microinstruction sequences calling for resolves or evaluates of Name syllables.

A SOP Trace Trap may be requested whenever FU 10120 enters State FM (First Microinstruction of an SOP). SOP Trace Traps may be masked by Trace Mask, Indivisibility Mask, or Trace Trap Enable, again provided by microinstruction control outputs of FUSITT 11012. In general, the first microinstruction of such

15 a microinstruction sequence interrupting such SOPs is not completed before a Trace Trap is taken.

Microinstruction Trace Traps may be requested upon completion of microinstructions which do not contain a Return Command, that is those microinstructions which do not return microinstruction control of CS 10110 to the calling microinstruction sequence. For microinstruction sequences containing Return Commands, state of microinstruction Trace Trap Request in a corresponding RCW is used. Every

20 microinstruction for which a Microinstruction Trace Trap is not masked is aborted during State M0 of execution of that microinstruction. Microinstruction Trace Traps may be masked by Trace Mask, Indivisibility Mask, or Trace Enable from FUSITT 11012. A Micro-Break-Point Trap may be requested upon

execution of microinstructions which do not contain Return Commands, but in which a Trace Enable bit in a microinstruction is asserted. A Micro-Break-Point Trap may be masked by Trace Mask, Indivisibility Mask, or Trace Enable. In addition, a Trace Enable bit of a microinstruction field in these microinstruction

25 sequences controls recognition of Micro-Break-point Traps. Micro-Break-Point Traps are thereby requested whenever a microinstruction Trace Trap is requested, but have additional enabling conditions expressed in the microinstructions. Since only recognized Traps are pushed onto RCWS 10358 in a RCW, a Microinstruction Trace Trap and a Micro-Break-Point Trap having different request states may be present in

30 RCWS 10358 concurrently.

Logical Write Trace Traps may be requested when enabled by a bit set in a logical descriptor during a microinstruction sequence submitting a write request to MEM 10112 and using logical descriptors to do so. Logical Write Trace Traps are recognized only if they occur during a state which will be immediately

35 followed by State MR (Memory Reference Trailer). A Logical Write Trace Trap will result in the MEM 10112 write request being aborted. Logical Write Trace Traps may be masked by Trace Masks, Indivisibility Mask, or Trace Trap Enable. A further condition for recognition of a Logical Write Trace Trap is determined by the state of certain bits in a logical descriptor of the memory write request. Logical Write Trace Traps are, in general, not pushed onto RCWS 10358 as part of a RCW since aborted MEM 10112 requests are re-generated so that Logical Write Trace Traps may be repeated.

Logical Read Trace Traps are similar in all respects to the Logical Write Trace Traps, but occur during MEM 10112 read requests. Generation of Logical Read Trace Traps is controlled again in part by certain bits

40 in logical descriptors of MEM 10112 read requests.

In certain implementations of CS 10110, UID Trace Traps may be requested when FU 10120 requests an MEM 10112 read operation based upon a UID address or pointer. UID Read Trace Traps are recognized if

45 requested and there is, in general, no explicit masking of UID Read Trace Traps. Generation of UID Read Trace Traps is controlled by certain bits in MEM 10112 read request logical descriptors. UID Read Trace Trap Requests result in the MEM 10112 read requests being aborted and CS 10110 entering State AB. Handler microinstruction sequences for UID Read Trace Traps will, in general, reset the trapped enable bit in the MEM 10112 read request logical descriptor before re-issuing the MEM 10112 read request.

UID Write Trace Traps are similar to UID Read Trace Traps, and are controlled by bits in the logical

50 descriptor in MEM 10112 write request based upon UID addresses or pointers.

Having described above structure and operation of Trace Trap Events, CS 10110 Interrupt Events will be described next below.

As previously described, Interrupts form the largest class of CS 10110 Events. Interrupts may be

55 regarded as falling into one or more of several classes. First, Memory Reference Repeat Interrupts are those Interrupt Events associated, in general, with read and write requests to MEM 10112 in which a read or write request is submitted to MEM 10112, and an Interrupt Event results. That Interrupt Event is handled, and the MEM 10112 request repeated. Second, Deferred Service Interrupts are those Interrupts wherein CS 10110 defers service of an Interrupt until entry to a new SIN. Fourth, Microinstruction Service Interrupts occur

60 when a currently executing microinstruction requires assistance of an Event Handler microinstruction sequence to be completed. Finally, Asynchronous Interrupt Events may occur at any time and must be serviced before CS 10110 may exit State M0 of the next microinstruction. These Interrupt Events will be described next below in the order named.

A Memory Reference Repeat Interrupt is requested, for example, if a microinstruction executes a

65 command, and a corresponding RCW read from RCWS 10358 indicates that a memory reference was

5 aborted before entrance to the microinstruction sequence from which return was executed. This type of Interrupt Event occurs for all aborted memory references. If an event is honored, that is abort state is entered, for any event and there is a memory reference outstanding, not aborted, the memory reference completes before State AB is exited. No Memory Repeat Interrupt Request will be written into the RCW written onto RCWS 10358. Conversely, if a memory reference is aborted, even if the event honored is not that event which aborted the memory reference, a Memory Repeat Interrupt Request will be written into a RCW pushed onto a RCWS 10358.

10 There are two state timing sequences for execution of Memory Repeat Interrupts. In the first case, there are no MEM 10112 references in the microinstruction executing a Return Command. In the second case, a microinstruction executing a Return Command executes a return and also makes a MEM 10112 reference. Referring to Fig. 246, a CS 10110 State Timing Diagram for the first case is shown. Fig. 246 is drawn using the same conventions as used in Fig. 244 and 245. As described above, in the first case a microinstruction executing a Return Command is executed in States M0 and M1 following Time D. An aborted MEM 10112 reference was made in States M0 and M1 preceding Time A. An MEM 10112 Reference Abort Request is made upon CS 10110's entry into State MR following Time A. Since a Memory Repeat Interrupt is requested only from a RCW provided by RCWS 10358, a Memory Repeat Interrupt is indicated only if a microinstruction executes a Return Command resulting in RCWS 10358 providing such an RCW. Therefore, a Memory Repeat Interrupt Request Register of EVENT 20284 is loaded with "not requesting" at this time. At Time B, CS 10110 enters State AB, State AR, and State MA. At this time, a Memory Reference Abort Request is asserted and written into an RCW when State AB is exited just before Time D. At Time D, CS 10110 exits State AR and State MA. As just described, CS 10110 will remain in State B until Time D. At Time D, Memory Reference Abort Request is written into RCWS 10358 as part of an RCW and, as described further below, various RCWS 10358 Stack Pointers are incremented to load that RCW into RCWS 10358. At this time, EVENT 20284's Interrupt Request Register receives "no request" as state of Memory Repeat Interrupt. First microinstruction of Memory Repeat Interrupt Handler microinstruction sequence is provided by FUSITT 11012. At Time E, the last microinstruction of the Memory Repeat Interrupt Handler microinstruction sequence is provided by FUSITT 11012 and a Return Command is decoded. RCWS 10358 Previous Stack Pointer, previously described, is selected to address RCWS 10358 to provide the previously written RCW as output to EVENT 20284's Memory Repeat Interrupt Event Register. At Time F, EVENT 20284's Memory Repeat Interrupt Register is loaded from output of RCWS 10358 and RCWS 10358's Stack Register Pointers are decremented. At this time, Memory Repeat Interrupt Request is made and, as described below, is written into the current Return Control Word, whether honored or not. JP 10114 then repeats the aborted MEM 10112 reference.

35 In the second case, a State Timing Sequence wherein the microinstruction executing a return also makes a MEM 10112 reference, CS 10110 State Timing is identical up to Time F. At Time F, MEM 10112 Repeat request is not recognized and the state of Memory Repeat Interrupt written into the current Return Control Word is "not requesting" unless a current MEM 10112 reference is aborted. The previous MEM 10112 Repeat Interrupt Request is disregarded as it is assumed that it is no longer required. Thus, there are two ways to avoid, or cancel a Memory Repeat Interrupt Request. First, that portion of a RCW receiving a MEM 10112 Repeat Interrupt Request may be rewritten as "not requesting". Second, an aborted MEM 10112 reference may be made in the same microinstruction that returns from a Handler servicing the aborted MEM 10112 reference.

Certain CS 10110 Events result in aborting a MEM 10112 read or write references and may result in repeat of MEM 10112 references. These events may include:

- 45 (1) Logical read and write Traps and, in certain implementations of CS 10110, UID read and write Traps, previously discussed;
- (2) A PC 10234 miss;
- (3) Detection of a protection Violation by PC 10234;
- (4) A Page Crossing in a MEM 10112 read or write request;
- 50 (5) A Long Address Translation, that is an ATU 10228 miss requiring JP 10114 to evaluate a logical descriptor to provide a corresponding physical descriptor;
- (6) Detection of a reset dirty bit flag from ATU 10228 upon a MEM 10112 write request as previously described;
- (7) An FU 10122 stack overflow;
- 55 (8) An FU 10122 Illegal Dispatch;
- (9) A Name Trace Trap event as previously described;
- (10) A Store Back Exception, as will be described below;
- (11) EU 10122 Events resulting in aborting of a Store Back, that is a write request to MEM 10112 from EU 10122;
- 60 (12) A read request to a non-accelerated Stack Frame, that is a Stack Frame presently residing in MEM 10112 rather than accelerated to JP 10114 Stack Mechanisms; and,
- (13) Conditional Branches in SIN sequences resulting outstanding MEM 10112 read reference from PREF 20260; and,

65 Of these Events, Logical Read and Write Traps, UID Read and Write Traps, and Name Trace Traps have been previously described. Other Events listed above will be described next below in further detail.

A PC 10234 Miss Interrupt may be requested upon a logical MEM 10112 reference, that is when a logical descriptor is provided as input to ATU 10228 and a protection state is not encached in PC 10234. PC 10234 will, as previously described, indicate that a corresponding PC 10234 entry is not present by providing a Event Protection Violation (EVENTPVIOL) output to EVENT 20284. PC 10234 will concurrently assert an Abort output (ABORT) to force CS 10110 into State AB and thus abort that MEM 10112 reference.

A Page Crossing MEM 10112 Reference Interrupt is requested if a logical MEM 10112 reference, that is a logical descriptor, specifies an operand residing on two logical pages of MEM 10112. An output of ATU 10228 will abort such MEM 10112 references by asserting an Abort output (ABORT).

A Protection Violation Interrupt is requested if a logical MEM 10112 reference does not possess proper access rights, a mode violation, or if that reference appears to refer to an illegal portion of that object, an extent violation. Again, PC 10234 will indicate occurrence of a Protection Violation Event, which may be disabled by a microinstruction control output of FUSITT 11012.

A Long Address Translation Event may be requested upon a logical MEM 10112 reference for which ATU 10228 does not have an encached entry. ATU 10228 will abort that MEM 10112 reference by asserting outputs ABORT and Long Address Translation Event (EVENTLAT).

A Dirty Bit Reset Event Interrupt may be requested when JP 10114 attempts to write to an MEM 10112 page having an encached entry in ATU 10228 whose dirty bit is not set. ATU 10228 will abort that MEM 10112 write request by asserting outputs ABORT and Write Long Address Translation Event (EVENTWLAT).

An FU 10120 User Stack Overflow Event may be requested if the distance between a Current Frame Pointer and a Bottom Frame Pointer, previously described with reference to CS 10110 Stack Mechanisms, is greater than a given value. As previously described, in CS 10110 this value is eight. A User Stack Overflow Event will continue to be requested until either Current Frame Pointer or Bottom Frame Pointer changes value so that the difference limit defined above is no longer violated. A User Stack Overflow Event may be masked by a Trace Mask, an Indivisibility Mask, or by enable outputs of a microinstruction from FUSITT 11012. A Handler microinstruction sequence for User Stack Overflow Events must be executed with one or more of these masks set to prevent recursion of these events. CS 10110 is defined to be running on Monitor Stack (MOS) 10370 when User Stack Overflow Events are masked. User Stack Overflow Events are not loaded into any of EVENT 20284's Event Registers, nor are these events written into a RCW to be written onto RCWS 10358.

Illegal EU 10122 Dispatch Events are requested by EUSDT 20266 if FU 10120 attempts to dispatch, or provide an initial microinstruction sequence address, to EU 10122 to a EUSITT address which is not accessible to a user's program. Illegal EU 10122 Dispatch Events are, in general, not masked. Illegal EU 10122 Dispatch Event Requests are cleared upon CS 10110 exits from State AB. The Handler microinstruction sequence for Illegal EU 10122 Dispatch Events should, in general, reset Illegal EU 10122 Dispatch Event entries in RCWs to prevent recursion of these events.

EU 10122 will indicate a Store Back Exception Event if any one of a number of exceptional conditions arise during arithmetic operations. These events are recognized when CS 10110 enters State SB and are ignored except during Store Back to MEM 10112 of EU 10122 results. These Events may be disabled by microinstruction output of FUSITT 11012 but are, in general, not masked. Store Back Exception Events may be written into RCWs, to be stored in RCWS 10358, and are cleared upon CS 10110's exit from State AB. Again, a Store Back Exception Event Handler microinstruction sequence should reset Store Back Exception Events written into RCWs to prevent recursion of these events.

As described above, the next major class of Interrupt Events are Deferred Service Interrupts. CS 10110 defers service of Deferred Service Interrupts until entry of a new SOP Deferred Service Interrupts which have been recognized will be serviced before completion of execution of the first microinstruction of that new SOP. Deferred Service Interrupts include Nonfatal MEM 10112 Errors, Interval Timer Overflows, and Interrupts from IOS 10116. These Interrupts will be described below, in the order named.

A Nonfatal MEM 10112 Interrupt is signaled by MEM 10112 upon occurrence of a correctable (single bit) MEM 10112 error. Nonfatal Memory Error Interrupts are recognized only during State M0 of the first microinstruction of an SOP. MEM 10112 will continue to assert Nonfatal Memory Error Interrupt until JP 10114 issues an acknowledgement to read MEM 10112's Error Log.

An Interval Timer Overflow Interrupt is indicated by TIMERS 20296 when, as described below, an Interval Timer increments to zero, thus indicating lapse of an allowed time limit for execution of an operation. Interval Timer Overflow Interrupts are recognized during State M0 of the first microinstruction of a SOP. TIMERS 20296 will continue to request such interrupts until cleared by a microinstruction output of FUSITT 11012.

IOS 10116 will indicate an IOS 10116 Interrupt to indicate that an inter-processor message from IOS 10116 to JP 10114 is pending. IOS 10116 will continue to assert an IOS 10116 Interrupt Request, which is stored in a register, until cleared by a microinstruction control output of FUSITT 11012. IOS 10116 Interrupts are recognized during State M0 of the first microinstruction of an SOP.

The next major class of CS 10110 events are Interrupts due to the requirement by microinstruction sequences to be serviced in order to complete execution. These Interrupts must be serviced before a microinstruction sequence may be completed. Microinstruction Service Interrupts include Illegal SOP Events, Microinstructions Not Present in FUSITT 11012 Events, an attempted parse of a hung INSTB 20262, underflow of an FU 10120 Stack, an NC 10226 Cache Miss, or an EU 10122 Stack Overflow. Each of these

events will be described below, in the order named.

An Illegal SOP Event is indicated by FUSDT 11010 to indicate that a current SOP Code is a Long Code, that is greater than eight bits, while the current dialect (S-Language) expects only Short Operation Codes, that is eight bit SOPs. An Illegal SOP interrupt is not detected for unimplemented SOPs within the proper code length range. Illegal SOP Events are, in general, not masked. FUSDT 11010 continues to indicate an Illegal SOP Event until a new SOP is loaded into OPCODEREG 20268. Illegal SOP Events are recognized during the first microinstruction of an SOP, that is during State FM. Should a Handler microinstruction sequence for a higher priority event change contents of OPCODEREG 20268, a previous Illegal SOP Event will be indicated again when the aborted SOP is retried.

Absence of a Microinstruction in FUSITT 11012 is indicated by FUSITT 11012 asserting a Control Store Address Invalid (CSADVALID). This FUSITT 11012 output indicates that that particular microinstruction address points outside of FUSITT 11012's address space. Output of FUSITT 11012 in such event is not determined and parity checking, described below, of microinstruction output is inhibited. The Handler microinstruction sequence for these Events will load FUSITT 11012 address zero with the required microinstruction from MEM 10112, as previously described, and return to the original microinstruction sequence.

An attempted parse of a hung INSTB 20262 is indicated by INSTBWC 24110 when a parse operation is attempted, INSTB 20262 is empty, and PREF 20260 is not currently requesting SInS from MEM 10112. In general, these Events are not masked. If a higher priority Event is serviced, these Events are indicated again when the aborted microinstruction is retried if the original conditions still apply.

An FU 10120 Stack Underflow Event is requested when a current microinstruction references a Previous Stack Frame which is not in an accelerated stack, that is, the Current Stack Pointer equals Bottom Stack Pointer. FU 10120 Underflow Events are, in general, not masked and are requested again on a retry if the microinstruction is aborted and this event has not been serviced.

An NC 10226 Miss Interrupt occurs on a MEM 10112 read or write operation when a load or read of NC 10226 is attempted and there is no valid NC 10226 block corresponding to that Name syllable. An NC 10226 Miss Event does not result in a request for a Name evaluate or resolve. In general, these Events are not masked and result in a request being issued again if the microinstruction resulting in that Event is retried and has not been serviced.

An EU 10122 Stack Overflow Event is requested from EU 10122 to indicate that EU 10122 is currently already servicing at least one level of Interrupt an FU 10122 is requesting another. As will be described in a following description of EU 10122, EU 10122 contains a one level deep stack for handling of interrupts. EU 10122 Stack Overflow Events are enabled during State NW. All previously pending events will have been serviced before EU 10122 Stack Overflow Event requests are recognized. These Events will be serviced immediately upon entry into a following State M0, being the highest priority interrupt event. EU 10122 Stack Overflow Events may, in general, not be masked and once recognized are the next honored event.

Finally, the third major class of CS 10110 Interrupt Events are Asynchronous Events. Asynchronous Events must, in general, be serviced before exiting State M0 of a microinstruction after they are recognized. Asynchronous Events include Fatal Memory Error Events, AC Power Failure Events, Egg Timer Overflow Events, and EU 10122 Stack Underflow Events. CS 10110 Egg Timer is a part of TIMERS 20296 and will be discussed as part of TIMERS 20296. These events will be described below, in the order referred to.

Fatal MEM 10112 Error Events are requested by MEM 10112 by assertion of control signal output PMODI, previously described, when last data read from MEM 10112 contains a noncorrectable error. Fatal MEM 10112 Error Events are recognized on first State M0 after occurrence. Fatal MEM 10112 Error Events are stored in an EVENT 20284 Event Register and are cleared upon entry into its service microinstruction sequence. In general, Fatal MEM 10112 Error Events may not be masked.

AC Power Failure Events are indicated by DP 10118 by assertion of output signal ACFAAIL when DP 10118 detects a failure of power to CS 10110. Recognition of AC Power Failure Events is disabled upon entry to AC Power Failure Event Handler microinstruction sequence. No further AC Power Failure Events will be recognized until DP 10118 reinitiates JP 10114 operation.

As will be described further below, FUJTL 20214's Egg Timer is a part of TIMERS 20296. Egg Timer Overflow Events are indicated by TIMERS 20296 whenever TIMERS 20296's Egg Timer indicates overflow of Egg Timer Counter. Egg Timer Overflow Events may be masked as described in a following description.

Finally, EU 10122 Stack Underflow Events are signaled by EU 10122 when directed to read a word from EU 10122 Stack Mechanism and there is no accelerated stack frame present. EU 10122 will continue to assert this Event Interrupt until acknowledged by JP 10114 by initiation of a Handler microinstruction sequence.

The above descriptions of CS 10110 events have stated that recognition of certain of those Events may be masked, that is inhibited to allow recognition of other Events having higher priority. Certain of these masking operations were briefly described in the above descriptions and will be described in further detail next below. In general, recognition of Events may be masked in five ways, four of which are properly designated as masks. These four masks are generated by microinstruction control from FUSITT 11012 and include Asynchronous Masks for, in general, Asynchronous Events. Monitor Masks are utilized for those CS 10110 operations being performed on Monitor Stack (MOS) 10370, as previously described with reference to CS 10110 Stack Mechanisms. Trace Mask is utilized with reference to Trace Trap Events. Indivisible Mask

is generated or provided by FUSITT 11012 as an integral or indivisible part of certain microinstructions and allow recognition of certain selected events during certain single microinstructions. Certain other Events, for example Logical Read and Write Traps and UID Read and Write Traps, are recognized or masked by flag bits in logical descriptors associated with those operations. Finally, certain microinstructions result in FUSITT 11012 providing microinstruction control outputs enabling or inhibiting recognition of certain events, but differ from Indivisible Masks in not being associated with single particular microinstructions.

Referring to Fig. 247, the relative priority level and applicable masks of certain CS 10110 Events are depicted therein in three vertical columns. Information regarding priority and masking of particular Events is shown in horizontal entries, each comprising an entry in each of these three vertical columns. Left hand column, titled Priority Level, states relative priority of each Event entry. Second column, titled EVENT, specifies which Event is referred to in that table entry. A particular Event will yield priority to all higher priority Events and will take precedence over all lower priority Events. Fig. 247's third column, titled Masked By, specifies for each entry which masks may be used to mask the corresponding Event. A indicates use of Asynchronous Masks, M use of Monitor Mask, T use of Trace Trap Mask, and I represents that Indivisible Mask may be used. DES indicates that an Event is enabled or masked by flag bits of logical descriptors, while MCWD indicates that a particular Event may be masked by microinstruction control signal outputs provided by FUSITT 11012. NONE indicates that a particular Event may, in general, not be masked.

The final major class of CS 10110 event was described above as Machine Check Events. In general, if any of these Events are detected by logic gating in EVENT 20284, EVENT 20284 will provide a Check Machine signal to DP 10118. DP 10118 will then stop operation of JP 10114 and Machine Check Event Handler microinstruction sequences will be initiated. Among these Machine Check Events are wherein FU 10120 is attempting to store back an EU 10122 result to MEM 10112 and EU 10122 signals a parity error in EU 10122's Control Store. These events are stored in EVENT 20284 Event Registers and recognized when FU 10120 enters State AB. EU 10122 will have previously ceased operation until a corrective microinstruction sequence may be initiated. The same Event will occur if FU 10120 attempts to use an EU 10122 arithmetic operation result or test operation result having a parity error in EU 10122's Control Store. Should MOS 10370 overflow or underflow, this event will be detected, FU 10120 operations stopped, and corrective microinstruction sequences initiated. MOS 10370 overflow or underflow occurs whenever a previous MOS 10370 Stack Frame is referenced, whenever MOS 10370 Stack Pointer equals MOS 10370 Bottom Stack Pointer, or the difference between MOS 10370 Current and Bottom Stack Pointers is greater than sixteen. Underflows result in a transfer of operation to MIS 10368, while overflows are handled by DP 10118. Finally, a Machine Check Event will be requested when a parity error is detected in a microinstruction currently being provided by FUSITT 11012 during State M0 of that microinstruction.

Having described general operation of EVENT 20284, the structure and operation of EVENT 20284 will be described briefly next below.

Referring to Fig. 248, a partial block diagram of EVENT 20284 is shown. EVENT 20284 includes Event Detector (EDET) 24810, Event Mask and Register Circuitry (EMR) 24812, and Event Handler Selection Logic (EHS) 24814. EDET 24810 is comprised of random logic gating and, as previously described, receives inputs representing event conditions from other portions of CS 10110's circuitry. EDET 24810 detects occurrences of CS 10110 operating conditions indicating that Events have occurred and provides outputs to EMR 24812 indicating what Events are requested.

EMR 24812 includes a set of registers, for example SN74S194s, comprising EVENT 20284's Event Registers. These registers are enabled by mask inputs, described momentarily, to enable masking of those Events which are latched in EVENT 20284's Event Registers. Certain Events, as previously described, are not latched and logic gating having mask enable inputs is provided to enable masking of those events which are not latched. EMR 24812 mask inputs are Asynchronous, Monitor, Trace Trap, and Indivisible Masks, respectively AMASK, MMSK, TMSK, and ISMK, provided from FUSITT 11012. Mask inputs derived from FUSITT 11012 microinstruction outputs (mWRD) are provided from microinstruction control outputs of FUSITT 11012. EMR 24812 provides outputs representing mask and unmask events which have been requested to EHS 24814.

EHS 24814 is comprised of logic gating detecting which of EHS 24814's unmasked Event Requests is of highest priority. EHS 24814 selects the highest priority unmasked Event Request input and provides a corresponding Event Handler microinstruction address to EVNTGT 24310 through ADRA Bus 24322. These address outputs of EHS 24814 are five bit addresses selecting the initial microinstruction of the Event Handler microinstruction sequence of the current highest priority unmasked Event. As previously described with reference to NASMUX 24312, certain inputs of ENTGT 24310 are hard-wired to provide a full fifteen bit address output from EVNTGT 24310. EVENT 20284 also provides, from EHS 24814, an Event Enable Select (EES) output to SITNAS 20286 to enable EVNTGT 24310 to provide microinstruction addresses to CSADR Bus 20204 when EVENT 20284 must provide a microinstruction address for handling of a current Event.

Having described the structure and operation of FUCTL 20214's circuitry providing microinstruction addresses to FUSITT 11012, FUSITT 11012 will be described next below.

c.c.c. Fetch Unit S-Interpreter Table 11012 (Fig. 249)

Referring to Fig. 249, a partial block diagram of FUSITT 11012 is shown. Address (ADR) and Data

(DATA) inputs of Micro-Instruction Control Store (mCS) 24910 are connected, respectively, from CSADR Bus 20204 through Address Driver (ADRDRV) 24912 and from JPD Bus 10142 through Data Driver (DDRV) 24194. mCS 24910 comprises a memory for storing sequences of microinstructions currently being utilized by CS 10110. mCS 24910 is an 8K (8192) word by 80 bit wide memory. That is, mCS 24910 may contain, for example, up to, 8192 80 bit wide microinstructions. Microinstructions to be written into mCS 24910 are provided, as previously described, to mCS 24910 DATA input from JPD Bus 10142 through DDRV 24914. Addresses of microinstructions to be written into or read from mCS 24910 are provided to mCS 24910 ADR input from CSADR Bus 20204 through ADRDRV 24912. ADRDRV 24912 and DDRV 24914 are buffer drivers comprised, for example, of SN74S240s and SN74S244s.

Also connected from output of ADRDRV 24912 is input of Nonpresent Micro-Instruction Logic (NPmIS) 24916. NPmIS 24916 is comprised of logic gating monitoring read addresses provided to mCS 24910. When a microinstruction read address present on CSADR Bus 20204 refers to an address location not within mCS 24910's address space, that is of a non-present microinstruction, NPmIS 24916 generates an Event Request output indicating this occurrence. As previously described FUCTL 20214 will then call, and execute, microinstructions so addressed from MEM 10112.

As indicated in Fig. 249, mCS 24910 provides three sets of outputs. These outputs are Direct Output (DO), Direct Decoded Output (DDO), and Buffered Decode Output (BDO). In general, control information within a particular microinstruction word is used on next clock cycle after the address of that particular microinstruction word has been provided to mCS 24910 ADR input. That is, during a first clock cycle a microinstruction's address is provided to mCS 24910 ADR input. That selected microinstruction appears upon mCS 24910's DO, DDO, BDO outputs during that clock cycle and are used, after decoding, during next clock cycle. Outputs DO, DDO, BDO differ in delay time before decoded microinstruction outputs are available for use.

mCS 24910 DO output provides certain bits of microinstruction words directly to particular destinations, or users, through Direct Output Buffer (DOB) 24918. These microinstructions bits are latched and decoded at their destinations as required. DOB 24918 may be comprised, for example, of SN74SO4s.

mCS 24910's DDO output provides decoded microinstruction control outputs for functions requiring the presence of fully decoded control signals at the start of the clock cycle in which those decoded control signals are utilized. As shown in Fig. 249, mCS 24910's DDO output is connected to input of Direct Decode Logic (DDL) 24920. DDL 24920 is comprised of logic gating for decoding certain microinstruction word bits during same clock cycle in which those bits are provided by mCS 24910's DDO. These microinstruction bits are provided, as described above, during the same clock cycle in which a corresponding address is provided to mCS 24910's ADR input. During this clock cycle, DDL 24920 decodes mCS 24910's DDO microinstruction bits to provide fully decoded outputs by end of this clock cycle. Outputs of DDL 24920 are connected to inputs of Direct Decode Register (DDR) 24922. DDR 24922 is a register comprised, for example, of SN74S374s. DDL 24920's fully decoded outputs are loaded into DDR 24922 at the end of the clock cycle during which, as just described, an address is provided to mCS 24910's ADR input and mCS 24910's corresponding DDO output is decoded by DDL 24920. Fully decoded microinstruction control outputs corresponding to mCS 24910's DDO outputs are thereby available at start of the second clock cycle. Microinstruction control outputs of DDR 24922 are thereby available to FU 10120 at start of the second clock cycle for those FU 10120 operations requiring immediate, that is undelayed, microinstruction control signal outputs from FUSITT 11012.

Finally, mCS 24910's BDO is provided for those FU 10120 operations not requiring microinstruction control signals immediately at the start of the second clock cycle. As shown in Fig. 249, mCS 24910's BDO is connected to inputs of Buffered Decode Register (BDR) 24924. Microinstruction word output bits from mCS 24910's BDO are provided to inputs of BDR 24924 during the clock cycle in which a corresponding address is provided to mCS 24910's ADR input. mCS 24910's BDO outputs are loaded into BDR 24924 at end of this clock cycle. BDR 24924's outputs are connected to inputs of Buffered Decode Logic (BDL) 24926. BDL 24926 is comprised of logic gating for decoding outputs of BDR 24924. BDL 24926 thereby provides decoded microinstruction control outputs to FU 10120 at some delayed time after start of the second clock cycle. Microinstruction control outputs from BDL 24926 are thereby delayed in time from the appearance of microinstruction control outputs of DDR 24922 but, as BDR 24924 stores microinstruction word bits rather than decoded microinstruction word bits, BDR 24924 is required to store proportionately fewer bits than DDR 24922.

Finally, as shown in Fig. 249 outputs of DDR 24922 and BDR 24924, are connected to inputs of Microinstruction Word Parity Checker (mWPC) 24928. mWPC 24928 is comprised of logic gating for checking parity of outputs of DDR 24922 and BDR 24924. A failure in parity of either output of DDR 24922 and BDR 24924 indicates a possible error in microinstruction output from mCS 24910. When such an error is detected by mWPC 24928, mWPC 24928 generates a corresponding Microinstruction Word Parity Error (mWPE).

d.d. CS 10110 Internal Mechanism Control

Associated with SR's 10362, the stack mechanism area of GRF 10354, are two CS 10110 control structures primarily associated with operation of CS 10110's internal mechanisms. A first of these referred to as Machine Control Block, describes current execution environment of JP 10114 microprograms, that is,

EP 0 067 556 B1

JP 10114 microinstruction sequences. Machine Control Block is comprised of two information words residing in MCW1 20290 and MCW0 20292. These Machine Control Words contain all control state information necessary to execute JP 10114's current microprogram. Second control structure is a portion of RCWS 10358, which as previously described parallels the structure of SR's 10362. Each register frame on MIS 10368 or MOS 10370 has, with exception of Top (Current) Register Frame, associated with it a Return Control Word (RCW) residing in RCWS 10358. RCWs are created when MIS 10362 or MOS 10370 register frames are pushed, that is moved onto MIS 10368 or MOS 10370 due to creation of a new Current Register Frame. A current RCW does not exist in a present embodiment of CS 10110.

RCWS 10358 will be described first next below, followed by Machine Control Block.

a.a.a. Return Control Word Stack 10358 (Fig. 251)

Referring to Fig. 251, a diagrammatic representation of a RCWS 10358 RCW is shown. As previously described, RCWS 10358 RCWs contain information necessary to reinitiate or continue execution of a microinstruction sequence if execution of that sequence has been discontinued.

Execution of a microinstruction sequence may be discontinued due to a requirement to service a CS 10110 Event, as described above, or if that microinstruction sequence has called for execution of another microinstruction sequence, as in a Branch or Case Operation.

As shown in Fig. 251, each RCW may contain, for example, 32 bits of information. RCW Bits 16 to 31 inclusive are primarily concerned with storing current microinstruction address of microinstruction sequences which have been discontinued, as described above. Bits 17 to 31 inclusive contain microinstruction sequence return address. Return address is, as previously described, address of the microinstruction currently being executed of a microinstruction sequence whose execution has been discontinued. When JP 10114 returns from servicing of an Event or execution of a called microinstruction sequence, return address is provided from RCWS 10358 to SITTNAS 20286 and through CSADR Bus 20204 to FUSITT 11012 as next microinstruction address to resume execution of that microinstruction sequence. Bit 16 of an RCW contains a state bit indicating whether the particular microinstruction referred to by return address field is the first microinstruction of a particular SOP. That is, Bit 16 of an RCW stores CS 10110 State FM.

Bits 8 to 15 inclusive of an RCW contain information pertaining to current condition code of JP 10114 and to pending Interrupt Requests. In particular, Bit 8 contains a condition code bit which, as previously described indicates whether a particular test condition has been met. RCW Bit 8 is thereby, as previously described, a means by which JP 10114 may pass results of a particular test from one microinstruction sequence to another Bits 9 to 15 inclusive of an RCW contain information regarding currently pending interrupts. These Interrupts have been previously discussed, in general, with reference to EVENT 20284. In particular, RCW Bit 9 contains pending state of Illegal EU 10122 Dispatch Interrupt Requests; RCW Bit 10 contains pending state of Name Trace Trap Request; RCW Bit 11 contains pending state of Store Back Interrupt Request; RCW Bit 12 contains pending state of Memory Repeat Interrupt Request; RCW Bit 13 contains pending state of SOP Trace Trap Request; RCW Bit 14 contains pending state of Microtrace Trap Request; and, RCW Bit 15 contains pending state of Micro-Break Point Trap Request. Interrupt Handling microinstruction sequence which require use of CS 10110 mechanisms containing information regarding pending Interrupts must, in general, save and store that information. This save and restore operation is accomplished by use of Bits 9 to 15 of RCWS 10358's RCWs. Upon entry to an Interrupt Handling microinstruction sequence, these bit flags are set to indicate Interrupts which were outstanding at time of entry to that microinstruction sequence. Because these bits are used to initiate Interrupt Request upon returns, pending Interrupts may be cancelled by resetting appropriate bits of Bits 9 to 15 upon return. This capability may be used to implement Microinstruction Trace Traps, previously described.

As indicated in Fig. 251, RCW Bits 0 to 7 are not utilized in a present embodiment of CS 10110. RCW bits 0 to 7 are not implemented in a present embodiment of CS 10110 but are reserved for future use.

As previously described, RCWs may be written into or read from RCWS 10358 from JPD Bus 10142. This allows contents of RCWS 10358 to be initially written as desired, or read from RCWS 10358 to MEM 10112 and subsequently restored as required for swapping of processes in CS 10110.

b.b.b. Machine Control Block (Fig. 252)

As described above, FUCTL 20214's Machine Control Block is comprised of a Machine Control Word 1 (MCW1) and a Machine Control Word 0 (MCW0). MCW1 and MCW0 reside, respectively, in Registers MCW1 20290 and MCW0 20292. MCW1 and MCW0 described the current execution environment of FUCTL 20214's current microprogram, that is the microinstruction sequence currently being executed by JP 10114.

Referring to Fig. 252, diagrammatic representations of MCW0 and MCW1 are shown. As indicated therein, MCW0 and MCW1 may each contain, for example, 32 bits of information regarding current microprogram execution environment.

Referring to MCW0, MCW0 includes 6 execution environment subfields. Bits 0 to 3 inclusive contain a Top Of Stack Counter (TOSCNT) subfield which is a pointer to Current Frame of accelerated Microstack (MIS) 10368. TOSCNT field is initially set to point to Frame 1 of MIS 10368. Bits 4 to 7 inclusive comprise a Top of Stack -1 Counter (TOS-1CT) subfield which is a pointer to Previous Frame of accelerated MIS 10368, that is to the MIS 10368 frame proceeding that pointed by TOSCNT subfield. TOS1CNT subfield is initially

set to Frame 0 of MIS 10368. Bits 8 to 11 inclusive comprise a Bottom of Stack Counter (BOSCNT) subfield which is a pointer to Bottom Frame of accelerated MIS 10368. BOSCNT subfield is initially set to point to Frame 1 of MIS 10368. TOSCNT, TOS-1CNT, and BOSCNT subfields of MCW0 may be read, written, incremented and decremented under microprogram control as frames are transferred between MIS 10368 and a SS 10336.

Bits 17 to 23 inclusive and Bits 24 to 31 inclusive of MCW0 comprise, respectively, Page Number Register (PNREG) and Repeat Counter (REPCTR) subfields which, together, comprise a microinstruction address pointing to a microinstruction currently being written into FUSITT 11012.

Bits 12 to 15 inclusive of MCW0 comprise an Egg Timer (EGGT) subfield which will be described further below with respect to TIMERS 20296. Bit 16 of MCW0 is not utilized in a present embodiment of CS 10110.

Referring to MCW1, MCW1 is comprised of four subfields. Of the 32 bits comprising MCW1, Bits 0 to 15 inclusive and Bits 24 and 25 are not utilized in a present embodiment of CS 10110. Bit 16 is comprised of a Condition Code (CC) subfield indicating results of certain test conditions in JP 10114. As previously described CC subfield is automatically saved and restored in RCWS 10358 RCW's.

Bits 17 to 19 inclusive of RCW1 comprise an Interrupt Mask (IM) subfield. The three bits of IM subfield are utilized to indicate a hierarchy of non-interruptible JP 10114 microinstruction control operating states. That is, a three bit code stored therein indicates relative power to interrupt between three otherwise noninterruptible JP 10114 operating states. Bits 20 to 23 inclusive comprise an Interrupt Request (IR) subfield which indicate Interrupt Request. These Interrupt Requests may include, for example, Egg Timer Overflow, Interval Timer Overflow, or Non-Fatal Memory Error, as have been previously described. Finally, Bits 26 to 31 inclusive comprise a Trace Trap Enable (TTR) subfield indicating which Trace Trap Events, previously described, are currently enabled. These enables may include Name Trace Enable, Logical Retrace Enable, Logical Write Trace Enable, SOP Trace Enable, Microinstruction Enable, and Microinstruction Break point Enable.

MCW0 and MCW1 has been described above as if residing in registers having individual, discrete existence, that is MCW1 20290 and MCW0 20292. In a present embodiment of CS 10110, MCW1 20290 and MCW0 20292 do not exist as a unified, discrete register structure but are instead comprised of individual registers having physical existence in other portions of FUCTL 20214. MCW1 20290 and MCW0 20292, and MCW1 and MCW0, have been so described to more distinctly represent the structure of information contained therein. In addition, this approach has been utilized to illustrate the manner by which current JP 10114 execution state may be controlled and monitored through JPD Bus 10142. As indicated in Fig. 202, MCW1 20290 and MCW0 20292 have outputs connected to JPD Bus 10142, thus allowing current execution state of JP 10114 to be read out of FUCTL 20214. Individual bits or subfields of MCW0 and MCW1 may, as previously described, be written by microinstruction control provided by FUSITT 11012. In a present physical embodiment of CS 10110, those registers of MCW0 20292 containing subfields TOSCNT, TOS-1CNT, and BOSCNT reside in RAG 20288. Those portions of MCW0 20292 containing subfield EGGT reside in TIMERS 20296. MCW0 20292 registers contain PNREG and REPCTR subfields are physically comprised of REPCTR 20280 and PNREG 20282. In MCW1 20290, CC subfield exists as output of FUCTL 20214 test circuits. Those MCW1 20290 registers containing IM, IR, and TTE subfields reside within EVENT 20284.

Having described FUCTL 20214 structure and operation as regards RCWS 10358, MCW1 20290 and MCW0 20292, FUCTL 20214, RAG 20288 will be described next below.

c.c.c. Register Address Generator 20228 (Fig. 253)

Referring to Fig. 253, a partial block diagram of RAG 20228, together with diagrammatic representation of GRF 10354, BIAS 20246 and RCWS 10358, is shown. As previously described, JP 10114 register and stack mechanisms include General Register File (GRF) 10354, BIAS 20246, and RCWS 10358. GRF 10354 is, in a present embodiment of CS 10110, a 256 word by 92 bit wide array of registers. GRF 10354 is divided horizontally to provide Global Registers (GRs) 10360 and Stack Registers (SRs) 10362, each of which contains 128 of GRF 10354's 256 registers. GRF 10354, that is both GRs 10360 and SRs 10362, is divided vertically into three vertical sections designated as AONGRF 20232, OFFGRF 20234, and LENGRF 20236. AONGRF 20232, OFFGRF 20234, and LENGRF 20236 are, respectively, 28 bits, 32 bits, and 32 bits wide. GRs 10360 is utilized as an array of 128 individual registers, each register containing one 92 bit word. SRs 10362 is structured and utilized as an array of 16 register frames wherein each frame contains eight registers and each register contains one 92 bit wide word. Eight of SR 10362's frames are utilized as Microstack (MIS) 10362 and the remaining eight of SR 10362's frames are utilized as Monitor Stack (MOS) 10370. For addressing purposes only, as described further below, GRs 10360 is regarded as being structured in the same manner as SRs 10362, that is as 16 frames of eight registers each.

BIAS 20246, as previously described, is a register array within BIAS 20246. BIAS 20246 contains 128 six bit wide registers, or words, and operates in parallel with and is addressed in parallel with SR 10362 portion of GRF 10354. RCWS 10358 is, as previously described, an array of 16 registers, or words, wherein each register contains one 32 bit RCW. RCWS 10358 is structured and operates in parallel with SRs 10362 with each RCWS 10358 register corresponding to a SR 10362 frame of eight registers. As described below, RCWS 10358 is addressed in parallel with SR 10362's frames.

Source and Destination Register Addresses (SDAR) for selecting a GRF 10354 register to be,

respectively, read from or written to are provided by RAG 20288. As described above BIAS 20246 operates and is addressed in parallel with SR 10362 portion of GRF 10354, that is parallel with SRs 10362. BIAS 20246 registers are thereby connected to and in parallel with address inputs of SRs 10362 and are addressed concurrently with GRs 10360. Registers RCWS 10358 also operate and are addressed in parallel with SRs 10362. Address inputs of RCWS 10358's registers are thereby connected in parallel with address inputs of SR 10362's registers.

RAG 20288's address inputs to GRF 10354, and to BIAS 20246 and RCWS 10358, may select registers therein to be either source registers, that is registers providing data, or destination registers, that is registers receiving data. RAG 20288's address outputs are designated as output Source and Destination Register Address (SDADR) of RAG 20288. RAG 20288's SDADR output is connected to address input of register comprising GRF 10354, BIAS 20246, and RCWS 10358. As described above, SRs 10362 are structured as 16 frames of 8 registers per frame and RCWS 10358 is structured as a corresponding 16 frames of one register per frame. GRF 10354 and BIAS 20246 are structured and utilized as single registers but, for addressing purposes, are regarded as being comprised of 16 frames of 8 registers per frame. Each SDADR output of RAG 20288 is an 8 bit word wherein the most significant bit indicates whether the addressed register, either a Source or a Destination Register, reside in GRs 10360 or within SRs 10362, BIAS 20246, and RCWS 10358. The four next most significant bits comprise a frame select field for selecting one of 16 frames within GRs 10360 or within SRs 10362, BIAS 20246, and RCWS 10358. The three least significant bits comprise a register select field selecting a particular register within the frame selected by frame select field.

Within a single system clock cycle, SDADR output of RAG 20288 may select a source register and data may be read from that source register, or SDADR output may select a destination register and data may be written into that destination register. As previously described, each JP 10114 microinstruction requires a minimum of two-system clock cycles for execution, that is at first clock cycle in State M0 and a second clock cycle in State M1. During a single microinstruction therefore, a source register may be selected and data read from that source register, and a destination register selected and data written into that destination register. Certain operations, however, may require more than one microinstruction for execution. For example, a read-modify-write operation wherein data is read from a particular register, modified, and written back into that register may require two or more microinstructions for execution.

Referring first to RAG 20288 structure, RAG 20288 includes MISPR 10356. MISPR 10356 includes Top Of Stack Counter (TOSCNT) 25310, Top Of Stack-1 Counter (TOS-1CNT) 25312, and Bottom Of Stack Counter (BOSCNT) 25314. Contents of TOSCNT 25310, TOS-1CNT 25312 and BOSCNT 25314 are respectively, pointers to Current, Previous, and Bottom frames of SRs 10362, that is, to MIS 10368. As will be described below, these pointers are also utilized to address MOS 10370. TOSCNT 25310, TOS-1CNT 25312, and BOSCNT 25314 are each four bit binary counters comprised, for example, of SN74S163s.

Data inputs of TOSCNT 25310 to BOSCNT 25314 are connected from JPD Bus 10142. Control inputs of TOSCNT 25310 to BOSCNT 25314 are connected from microinstruction control outputs of FUSITT 11012. Data outputs of TOSCNT 25310 to BOSCNT 25314 are connected to data inputs of Source Register Address Multiplexer (SRCADR) 25316 and to data inputs of Destination Register Address Multiplexer (DSTADR) 25318. Data outputs of TOSCNT 25310 and BOSCNT 25314 are connected to inputs of Stack Event Monitor Logic (SEM) 25320.

Source and destination frame addresses are selected, as will be described further below, by SRCADR 25316 and DSTADR 25318 respectively. In addition to data inputs from TOSCNT 25310 and BOSCNT 25314, data inputs of SRCADR 25316 and DSTADR 25318 are connected from microinstruction word CONEXT subfield output from FUSITT 11012. Control inputs of SRCADR 25316 and DSTADR 25318 are connected from, respectively, microinstruction word RS and RD subfield outputs from FUSITT 11012. Source Frame Address Field (SRCFADR) output of SRCADR 25316 and Destination Frame Address Field (DSTFADR) output of DSTADR 25318 are connected to inputs of Source and Destination Register Address Multiplexer (SDADMUX) 25322. SRCFADR and DSTFADR comprise frame select fields of RAG 20288, SDADR outputs for, respectively, source and destination registers.

In addition to SRCFADR and DSTFADR outputs of SRCADR 25316 and DSTADR 25318, SDADMUX 25322 receives microinstruction word SRC and DST subfield inputs from microinstruction outputs of FUSITT 11012. As previously described, SRC subfield is a 3 bit number designating a source register, that is, a source register within a frame selected by SRCFADR. DST is similarly a 3 bit number selecting a destination register within a frame indicated by DSTFADR. SRC subfield input to SDADMUX 25322 is concatenated with SRCADR 25316 to respectively comprise, as described above, register and frame fields of a source register SDADR output of SDADMUX 25322. Similarly, DST subfield is concatenated with DSTFADR output of DSTADR 25318 to comprise, respectively, register and frame subfields of a destination register SDADR output of SDADMUX 25322. Selection between source and destination register address inputs to SDADMUX 25322, to generate a corresponding source or destination register SDADR output of SDADMUX 25322 is controlled by microinstruction control inputs (not shown for clarity of presentation) connected to control inputs of SDADMUX 25322. RDWS 25324 is a PROM decoding MD field from microinstruction words during reads from MEM 10112 and provides register select field of destination register address and selects one of the pointers as frame select field.

An Event output of SEM 25320 is connected to an input of EVENT 20284, previously described.

SRCADR 25316, DSTADR 25318, and SDADMUX 25322, as will be described further below, operate as multiplexers and may be comprised, for example, of SN74S153s.

Having described structure and organization of GRF 10354, BIAS 20246, and RCWS 10358, and structure of RAG 20288, operation of RAG 20288 to generate Source of Destination Register Address outputs SDADR will be described next below. Addressing of JP 10114's stack mechanism, comprising SRs 10362 and RCWS 10358, will be described first, followed by addressing of GRs 10360 and BIAS 20246.

SR 10362 portion of GRF 10354, RCWS 10358, and BIAS 20246 are addressed by Current, Previous, and Bottom Frame Pointers contained, respectively, in TOSCNT 25310, TOS-1CNT 25312, and BOSCNT 25314. Current, Previous, and Bottom Pointers comprise frame select fields of SDADMUX 25322. As previously described, Current, Previous and Bottom Pointer outputs of TOSCNT 25310 to BOSCNT 25314 are provided as inputs of SRCADR 25316 and DSTADR 25318. Microinstruction word RS subfield to control input of SRCADR 25316 selects either Current, Previous or Bottom Pointer input of SRCADR 25316 to comprise SRCFADR output of SRCADR 25316, that is to be frame select field of source register address. Similarly, microinstruction word RD subfield to control input of DSTADR 25318 concurrently selects either Current, Previous, or Bottom Pointer inputs of DSTADR 25318 to comprise DSTFADR output, that is frame select field of destination register address. As described above, SRCFADR and DSTFADR are provided as inputs to SDADMUX 25322. Microinstruction word SRC and DST subfield inputs to SDADMUX 25322 concurrently determine, respectively, source and destination registers within source and destination frames specified by SRCFADR and DSTFADR. SDADMUX 25322 then, operating under microinstruction control, selects either SRCFADR and SRC to comprise SDADR output to SR 10362 as a source register address or selects DSTFADR and DST as SDADR output specifying a destination register address. By microinstruction control of SRCADR 25316, DSTADR 25318, and SDADMUX 25322, a CS 10110 microprogram may select a source frame and register within SR 10362 and simultaneously specify a possible different destination frame and register within SR 10362. All possible combinations of source frame and register and destination frame and register in GRs 10360, SRs 10362, BIAS 20246, and RCWS 10358 are valid.

Control of SRCADR 25316, DSTADR 25318, and SDADMUX 25322 in addressing SR 10362 portion of GRF 10354, and RCWS 10358, is controlled, in part, by current CS 10110 state. Pertinent CS 10110 operating states, previously described, are State M1 and State RW. When CS 10110 is in neither State RW nor State M1, SR 10362 is addressed through SRCADR 25316 and microinstruction word SRC subfield, that is SR 10362 and RCWS 10358 are provided with source register addresses when CS 10110 is in neither RW nor M1 States. When CS 10110 enters State M1, SR 10362 and RCWS 10358 is addressed through DSTADR 25318 and by microinstruction word DST subfield. That is, SR 10362 and RCWS 10358 are provided with destination register addresses during State M1. Similarly, SR 10362 and RCWS 10358 are provided with destination register addresses when CS 10110 is operating in State RW, that is when data is being read from MEM 10112 and written into SR 10362 or RCWS 10358. In this case, however, low order 3 bits of destination register address, that is register select field, are provided by RDS 25324, which decodes microinstruction word subfield MD (Memory Destination). RDS 25324 also provides a control input that DSTADR 25318 to select one of Current, Previous, or Bottom pointers from MISPR 10356 to comprise frame select field of destination register address.

As stated above, frame select field of source and destination register addresses are provided from TOSCNT 25310, TOS-1CNT 25312, and BOSCNT 25314. As described above, the most significant bit of source and destination register address are forced to logic 1 or logic 0, depending upon whether GR 10360 or SR 10362, BIAS 20246, and RCWS 10358 are being addressed. Contents of TOSCNT 25310 to BOSCNT 25314, that is Current, Previous, and Bottom Pointers, are controlled by microinstruction control outputs of FUSITT 11012. Current and Previous Pointers change as stacks are "pushed" or "popped" to and from MIS 10368 as JP 10114 performs, respectively, calls and returns. Similarly, Current, Previous and Bottom Pointers will be incremented or decremented as MIS 10368 frames are transferred between MIS 10368 and MEM 10112, as previously described with respect to CS 10110's Stack Mechanisms.

Referring first to Current and Previous Pointer operation, Current and Previous Pointers in TOSCNT 25310 and TOS-1CNT 25312 are initially set, respectively, to point to Frames 1 and 0 of MIS 10368 by being loaded from JPD Bus 10142. TOSCNT 25310 and TOS-1CNT 25312 are enabled to count when two conditions are met. First condition is dependent upon current operating state of CS 10110. TOSCNT 25310 and TOS-1CNT 25312 will be enabled to count during last system clock cycle of CS 10110 operating States M1 or AB. Second condition is dependant upon whether JP 10114 is to execute a call or return. TOSCNT 25310 and TOS-1CNT 25312 may be enabled to count if a current microinstruction indicates JP 10114 is to execute a call or return, or if CS 10110 is exiting State AB as exit from State AB is an implied call operation. Both a call and an implied call, that is exit from State AB, will cause TOSCNT 25310 and TOS-1CNT 25312 to be incremented. A return will cause TOSCNT 25310 and TOS-1CNT 25312 to be decremented.

Referring to BOSCNT 25314, Bottom Frame Pointer is initially loaded from JPD Bus 10142 to point to MIS 10368 Frame 1. Again, incrementing or decrementing of BOSCNT 25314 is dependant upon CS 10110 operating state and operation to be performed. BOSCNT 25314 is enabled to count upon exiting from State M1. In addition, DEVCMD subfield of a current microinstruction word must indicate that BOSCNT 25314 is to be incremented or decremented. BOSCNT 25314 will be incremented or decremented upon exit from

EP 0 067 556 B1

State M1 as indicated by microinstruction word DEVCMD subfield.

SEM 25320 monitors relative values of Current and Bottom Pointers residing in TOSCNT 25310 and BOSCNT 25314 and provides outputs to EVENT 20284 for purposes of controlling operation of MI 10368 and MOS 10370. SEM 25320 is comprised of a Read Only Memory, for example 93S427s, receiving Current and Bottom Pointers as inputs. SEM 25320 detects 3 Events occurring in operation of TOSCNT 25310 and BOSCNT 25314, and thus in operation of MIS 10368 and MOS 10370. First, SEM 25320 detects an MIS 10368 Stack Overflow. This Event is indicated if the present value of Current Frame Pointer is greater than 8 larger than the present value of Bottom Frame Pointer. Second, SEM 25320 detects when MIS 10368 contains only one frame of information. This event is indicated if the value of Current Frame Pointer is equal to the value of Bottom Frame Pointer. In this case, the previous frame of MIS 10368 resides in MEM 10112 and must be fetched from MEM 10112 before a reference to the previous stack frame may be made. Third, SEM 25320 detects when MIS 10368 and MOS 10370 are full. This Event is indicated if the present value of Current Frame Pointer is 16 larger than the present value of Bottom Frame Pointer. When this Event occurs, any further attempt to write a frame onto MIS 10368 or MOS 10370 will result in a MOS 10370 Stack Overflow. EVENT 20284 responds to these Events indicated by SEM 25320 by initiating execution of an appropriate Event Handling microinstruction sequence, as previously described. It should be noted that MIS 10368 and MOS 10370 are addressed in the same manner, that is through use of Current, Previous and Bottom Frame Pointers and certain microinstruction word subfields. Primary difference between operation of MIS 10368 and MOS 10370 is in the manner in which stack overflows are handled. In the case of MIS 10368, stack frames are transferred between MIS 10368 and MEM 10112 so that MIS 10368 is effectively a bottomless stack. MOS 10370, however, contains a maximum of 8 stack frames, in a present embodiment of CS 10110, so that no more than eight Events may be pushed onto MOS 10370 at a given time.

GR 10360 is addressed in a manner similar to SR 10362, BIAS 20246, and RCWS 10358, that is through ADRSRC 25316, DSTADR 25318, and SDADRMUX 25322. Again, register select fields of source and destination register addresses are provided by microinstruction word SRC and DST subfields. Frame select field of source and destination register addresses is, however, specified by microinstruction word CONEXT subfield. In this case, microinstruction word RS and RD subfields specify that frame select fields of source and destination register addresses are to be provided by CONEXT subfield. Accordingly, ADRSRC 25316 and DSTADR 25318 provide CONEXT subfield as SRCFADR and DSTFADR inputs to SDADRMUX 25322.

Having described structure and operation of RAG 20288, TIMERS 20296 will be described next below. Referring to Fig. 254, a partial block diagram of TIMERS 20296 is shown. As indicated therein, TIMERS 20296 includes Interval Timer (INTTMR) 25410, Egg Timer (EGGTMR) 25412, and Egg Timer Clock Enable Gate (EGENB) 25416.

d.d.d. Timers 20296 (Fig. 254)

Referring first to INTTMR 25410, a primary function of INTTMR 25410 is to maintain CS 10110 architectural time as previously described with reference to Fig. 106A and previous descriptions of CS 10110 UID addressing. As described therein, a portion of all UID addresses generated by all CS 10110 systems is an Object Serial Number (OSN) field. OSN field uniquely defines each object created by operation of or for use in a particular CS 10110. OSN field of an object's UID is, in a particular CS 10110, generated by determining time of creation of that object relative to an arbitrary historic starting time common to all CS 10110 systems. That time is maintained within a MEM 10112 storage space, or address location, but is measured by operation of INTTMR 25410.

INTTMR 25410 is a 28 bit counter clocked by a 110 Nano-Second Clock (110NSCLK) input and is enabled to count by a one MHZ Clock Enable input (CLK1MHZENB). INTTMR 25410 may thereby be clocked at a one MHZ rate to measure one microsecond intervals. Maximum time interval which may be measured by INTTMR 25410 is thereby 268.435 seconds.

As indicated in Fig. 254, INTTMR 25410 may be loaded from and read to JPD Bus 10142. In normal operation, the MEM 10112 location containing architectural time for a particular CS 10110 will be loaded with current architectural time at time of start up of that particular CS 10110. INTTMR 25410 will concurrently be loaded with all zeros. Thereafter, INTTMR 25410 will be clocked at one microsecond intervals. Periodically, when INTTMR 25410 overflows, architectural time stored in MEM 10112 will be accordingly updated. At any time, therefore, current architectural time may be determined, down to a one microsecond increment, by reading architectural time from the previous updated architectural time stored in MEM 10112 and elapsed interval since last update of architectural time from INTTMR 25410. In the event of a failure of CS 10110, architectural time in MEM 10112 and INTTMR 25410 may be saved in MEM 10112 by reading elapsed intervals since last architectural time update. When normal CS 10110 operation resumes, INTTMR 25410 may be reloaded with a count reflecting current architectural time. As indicated in Fig. 254, INTTMR 25410 is loaded from JPD Bus 10142 when INTTMR 25410 is enabled by a Load Enable input (LDE) provided from DP 10118.

Referring to EGGTMR 25412, certain CS 10110 Events, in particular Asynchronous Events previously described with reference to EVENT 20284, are received or acknowledged by EVENT 20284 only at conclusion of State M1 of first microinstruction of an SOP. As certain CS 10110 microinstructions have long execution times, these Asynchronous Events may be subjected to an extended latency, or waiting, interval

before being serviced. EGGTMR 25412, in effect, measures latency time of pending Asynchronous Events and provides an output to EVENT 20284 if a predetermined maximum latency time is exceeded.

As indicated in Fig. 254, EGGTMR 25412 is clocked by a 110 Nano-Second Clock input (110NSCLK). EGGTMR 25412 is initially set to zero by load input (LDZRO) at end of State MI of the first microinstruction of each SOP executed by CS 10110, or when specifically instructed so by DEVCMD subfield of a microinstruction word. EGGTMR 25412 is incremented when enabled by Clock Enable (CLKENB) input from EGGENB 25416. There are two conditions necessary for EGGTMR 25412 to be incremented. First condition is occurrence of an Asynchronous Event, which is indicated by input ASYEVNT to EGGENB 25416 from EVENT 20284. Second condition is that 16 or more microseconds have elapsed since last increment of EGGTMR 25412. This interval is measured by an output from fourth bit of INTTMR 25410 which, as shown in Fig. 254, is connected to an input of EGGENB 25416. EGGTMR 25412 is a four bit counter and will thereby overflow and generate output OVRFLW to EVENT 20284 256 microseconds after beginning of an SOP if an Asynchronous Event has occurred and if at least 16 microseconds have elapsed since start of that SOP. EGGTMR 25412 thereby insures a maximum service latency of 256 microseconds for Asynchronous Events.

e.e.e. Fetch Unit 10120 Interface to Execute Unit 10122

Finally, as previously described FU 10120's interface to EU 10122 is primarily comprised of EUDIS Bus 20206, for providing EUDPs to EU 10122's EUSITT, and FUINT 20298. Operation of EUSDT 20266 and EUDIS Bus 20206 has been previously described and will be described further in a following description of EU 10122. FUINT 20298 is primarily concerned with generating Event Requests for conditions signalled from EU 10122 so that these Events may be serviced. In this regard, FUINT 20298 is primarily comprised of gates receiving Event Requests from EU 10122 and providing corresponding outputs to EVENT 20284. Another interface function performed by FUINT 20298 is generation of a "transfer complete" signal generated by FU 10122 and provided to EU 10122 to assert that a EU 10122 result read from EU 10122 to FU 10120 has been received. This transfer complete signal indicates to EU 10122 that EU 10122's result register, described in a following description of EU 10122, is available for further use by EU 10122. This transfer complete signal is generated by an output of FUSITT 11012 as part of microinstruction sequences for transferring data from EU 10122 to FU 10120 or MEM 10112.

Having described structure and operation of FU 10120, including DESP 20210, MEMINT 20212, and FUCTL 20214, the structure and operation of EU 10122 will be described next below.

C. Execute Unit 10122 (Figs. 203, 255—268)

As previously described, EU 10122 is an arithmetic processor capable of executing integer, packed and unpacked decimal, and single and double precision floating point arithmetic operations. A primary function of EU 10122 is to relieve FU 10120 of certain arithmetic operations, thus enhancing efficiency of CS 10110.

Transfer of operands from MEM 10112 to EU 10122 is controlled by FU 10120, as is transfer of results of arithmetic operations from EU 10122 to FU 10120 or MEM 10112. In addition, EU 10122 operations are initiated by FU 10120 by EU 10122 Dispatch Pointers invited to EU 10122 by EUSDT 20266. EU 10122 Dispatch Pointers may initiate both arithmetic operations required for execution of SINs and certain EU 10122 operations assisting in handling of CS 10110 events. As previously described, EU 10122 Dispatch Pointers are translated into sequences of microinstructions for controlling EU 10122 by EU 10122's EUSITT which is similar in structure and operation to FUSITT 11012. As will be described further below, EU 10122 includes a command queue for receiving and storing sequences of EU 10122 Dispatch Pointers from FU 10120. In addition, EU 10122 includes a general register file, or scratch pad memory, similar to GRF 10354. EU 10122's general register file is utilized, in part, in EU 10122 Stack Mechanisms similar to FU 10120's SR's 10362.

Referring to Fig. 203, a partial block diagram of EU 10122 is shown. EU 10122's general structure and operation will be described first with reference to Fig. 203. Then EU 10122's structure and operation will be described in further detail with aid of subsequent figures which will be presented as required.

As indicated in Fig. 203, major elements of EU 10122 include Execute Unit Control Logic (EUCL) 20310, Execute Unit IO Buffer (EUIO) 20312, Multiplier Logic (MULT) 20314, Exponent Logic (EXP) 20316, Multiplier Control Logic (MULTCNTL) 20318, and Test and Interface Logic (TSTINT) 20320. EUCL 20310 receives Execute Unit Dispatch Pointers (EUDPs) from EUSDT 20266 and provides corresponding sequences of microinstructions to control operation of EU 10122.

EUIO 20312 receives operands, or data, from MEM 10112, translates those operands into certain formats most efficiently used by EU 10122. EUIO 20312 receives results of EU 10122's operations and translates those results into formats to be returned to MEM 10112 or FU 10120, and presents those results to MEM 10112 and FU 10120.

MULT 20314 and EXP 20316 are arithmetic units for performing arithmetic manipulations of EU 10122 operations. In particular, EXP 20316 performs operations with respect to exponent fields of single and double precision floating point operations. MULT 20314 performs arithmetic manipulations with respect to mantissa fields of single and double precision floating point operations, and arithmetic operations with regard to integer and packed decimal operations. MULTCNTL 20318 controls and coordinates operation of

EP 0 067 556 B1

MULT 20314 and EXP 20316 and prealignment and normalization of mantissa and exponent fields in floating point operations. Finally, TSTINT 20320 performs certain test operations with regard to EU 10122's operations, and is the interface between EU 10122 and FU 10120.

5
a. General Structure of EU 10122
1. Execute Unit I/O 20312

10 Referring first to EUIO 20312, EUIO 20312 includes Operand Buffer (OPB) 20322, Final Result Output Multiplexer (FROM) 20324, and Exponent Output Multiplexer (EXOM) 20326. OPB 20322 has first and second inputs connected, respectively, from MOD Bus 10144 and JPD Bus 10142. OPB 20322 has a first output connected to a first input of Multiplier Input Multiplexer (MULTIM) 20328 and MULT 20314. A second output of OPB 20322 is connected to first inputs of Inputs Selector A (INSELA) 20330 and Exponent Execute Unit General Register File Input Multiplexer (EXRM) 20332 in EXP 20316.

15 FROM 20324 has an output connected to JPD Bus 10142. A first input of FROM 20324 is connected from output of Multiplier Execute in General Register File Input Multiplexer (MULTRM) 20334 and MULT 20314. A second input of FROM 20324 is connected from output of Final Result Register (RFR) 20336 of MULT 20314. EXOM 20326 has an output connected to JPD Bus 10142. EXOM 20326 is a first input connected from output of Scale Register (SCALER) 20338 of EXP 20316. EXOM 20326 has second and third inputs connected from outputs of, respectively, Next Address Generator (NAG) 20340 and Command Queue (COMQ) 20342 of EUCL 20310.
20

2. Execute Unit Control Logic 20310

25 Referring to EUCL 20310, EUCL 20310 includes NAG 20340, COMQ 20342, Execute Unit S Interpreter Table (EUSITT) 20344, and Microinstruction Control Register and Decode Logic (mCRD) 20346. COMQ 20342 has an input connected from EUDIS Bus 20206 for receiving SDPs from EUSDT 20266. COMQ 20342 has, as described above, a first output connected to a third input of EXOM 20326, and has a second output connected to an input of NAG 20340. NAG 20340 has, as described above, a first output connected to second input of EXOM 20326. NAG 20340 has a second output connected to a first input of EUSITT 20344.
30 As previously described, EUSITT 20344 corresponds to FUSITT 11012 and stores sequences of microinstructions for controlling operation of EU 10122 in response to EU 10122 Dispatch Pointers from FU 10120. EUSITT 20344 has a second input connected from JPD Bus 10142 and has an output connected to input of mCRD 20346. mCRD 20346 includes a register and logic for receiving and decoding microinstructions provided by EUSITT 20344. In addition to an input from EUSITT 20344, mCRD 20346 has first outputs providing decoded microinstruction control signals to all parts of EU 10122. mCRD 20346 also
35 has a second output connected to a first input of Input Selector B (INSELB) 20348 and EXP 20316.

3. Multiplexer Logic 20314

40 Referring to MULT 20314, MULT 20314 includes two parallel arithmetic operation paths for performing addition, subtraction, multiplication, and division operations on packed decimal numbers, integer numbers, and mantissa portions of single and double precision floating point numbers. MULT 20314 also includes a related portion of EU 10122's general register file, a memory for storing constants used in arithmetic operations, and certain input data selection circuits. That portion of EU 10122's GRF residing in
45 MULT 20314 is comprised of Multiplier Register File (MULTRF) 20350. Output of MULTRF 20350 is connected to a second input of MULTIM 20328. A first input of MULTRF 20350 is connected from output of RFR 20336 and a second input of MULTRF 20350 is connected from output of MULTRM 20334. First and second inputs of MULTRM 20334 are in turn connected, respectively, from output of RFR 20336 and from output of Container Size Logic (CONSIZE) 20352 of TSTINT 20320.

50 MULTIM 20328 selects the data inputs to MULT 20314's arithmetic circuits and has, as previously described, first and second inputs connected respectively from first output of OPB 20322 and from output of MULTRF 20350. Output of MULTIM 20328 is connected through Multiplier (MULT) Bus 20354 to input of Multiplier Quotient Register (MQR) 20356 and to input of Nibble Shifter (NIBSHF) 20358. Another input to MQR 20356 and NIBSHF 20358 is provided by Constant Store (CONST) 20360. CONST 20360 is a memory
55 for storing constant values used in MULT 20314 operations. Output of CONST 20360 is connected to MULT Bus 20354. MULT 20314's arithmetic circuits may thereby be provided with inputs from OPB 20322, MULTRF 20350, and CONST 20360.

MULT 20314's arithmetic circuitry is comprised of two, parallel arithmetic operation paths having, as common inputs, outputs of MULTIM 20328 and CONST 20360. Common termination of these parallel
60 arithmetic operation paths is Final Register Shifter (FRS) 20362. A first arithmetic operation path is provided through NIBSHF 20358, whose input is connected from MULT Bus 20354. NIBSHF 20358's output is connected to a first input of FRS 20362 and a control input of NRBSHF 20358 is connected from an output of Multiplier Control Logic (MULTCNT) 20364 and MULTCNTL 20318.

MULT 20314's second arithmetic operation path is provided through MQR 20356. As described above,
65 MQR 20356's input is connected from MULT Bus 20354. MQR 20356's output is connected to first and