Integrated Circuit Cards, Tags and Tokens

New Technology and Applications

Edited by
P. L. Hawkes, D. W. Davies
and W. L. Price

BSP PROFESSIONAL BOOKS
OXFORD LONDON EDINBURGH
BOSTON MELBOURNE

First published 1990

British Library
Cataloguing in Publication Data

Integrated circuit cards, tags and tokens.
1. Smart cards
I. Hawkes, P. (Peter)
II. Davies, D.W. (Donald Watts),
III. Price, W. L.
004.5'6

ISBN 0-632-01935-2

BSP Professional Books
A division of Blackwell Scientific Publications Ltd
Editorial Offices:
Osney Mead, Oxford OX2 0EL
(Orders: Tel. 0865 240201)
8 John Street, London WC1N 2ES
23 Ainslie Place, Edinburgh EH3 6AJ
3 Cambridge Center, Suite 208, Cambridge MA 02142, USA
107 Barry Street, Carlton, Victoria 3053, Australia

Set by Setrite Typesetters Limited
Printed and bound in Great Britain by
MacKays of Chatham PLC, Chatham,Kent

Contents

Preface

The 'smart' card single chip computer in a plastic credit card shape is widely promoted by its numerous suppliers and their agents as the ultimate microcomputer destined to be carried by everyone everywhere sometime soon.

Why, where, when, questions from prospective card holders amongst the public and the key intermediaries like the bankers, retailers, medical profession, public administrators and telephone companies do not always receive straight answers. The benefits of using smart cards are less tangible than the early costs of introducing systems based on these intriguing devices. In this book we attempt to help the reader resolve the many paradoxes associated with the smart card and its close relatives, the radio tag, the integrated circuit digital memory card, the token and electronic coin.

Amongst the many paradoxes bedevilling the whole subject are the following.

Most of the tens of millions of smart cards now produced annually are not 'smart', more usually they are the humbler relative called the integrated circuit digital memory card. Most of these are used for vending applications like public payphones where an equally cost effective result can apparently be achieved with an optical recording card.

The commonest smart cards produced have on one face of the card electrical inter-connections to the read/write authorisation units. This type of card is the subject of international standards work. However, for many applications these contact smart cards are being challenged by the new contactless radio linked cards such as those available from GEC and AT&T.

But even these new contactless radio linked cards are not as new as they seem. They are predated by the well established radio tag used in the access control field to identify animals, people or goods.

Mars Electronics have shown that it is possible to design an electronic coin having the shape and size of a conventional coin but functioning as a

stored value device. There are many other prospective designs of smart 'card' where non-card shapes are preferable for good mechanical and economic reasons. We thus have the paradox that the only real justification for the smart card being card shaped and sized is the transient problem of devising a terminal which will read both magnetic strip and embossed cards as well as smart cards.

Another paradox lies in the claims for smart card security. The card is hailed as the ultimate in security for both access control and as an instrument in financial transactions. In the latter application the smart card is capable of dispensing and recording as data transferred value (equals money). Card stored or emitted files of data, the equivalent of money, obviously require protection from deliberate or accidental misuse both from the authorised card holder breaking the rules and from thieves. To protect card stored data and emitted messages requires data protection measures. These are best based on the applied mathematical techniques of cryptography. The chapter by Dr D. W. Davies describes some of the basics of this most important software area.

Given satisfactory software and economic and durable hardware most application systems based on smart cards remain vulnerable to misuse of a valid card by unauthorised card holders who have stolen or worse still borrowed genuine cards from the authorised holders.

Establishing the cardholder's right to use a given card is currently based on the holder producing the appropriate personal identity number (PIN) or password. Both PINs and passwords can be readily extorted or otherwise obtained from the cardholder's mind or records. Thus although the smart card itself may be secure against many types of misuse limiting use to the authorised holder can be a real problem. Dr J. R. Parks describes the new technology of biometrics which seeks to reduce current dependence on PINs by making measurements on some characteristic of the person such as voice print, fingerprint or handwriting style in order to confirm that he/she is indeed the authorised cardholder.

Some limitations of smart card systems can be overcome by using them in on-line systems where every transaction must be authorised by real-time checks on centrally held lists of stolen and barred cards. The communications infrastructure for a totally on-line system is very expensive. Arlen Lessin's chapter describes one of the new super-smart cards which operate off-line.

For many large scale applications smart cards remain impossibly expensive. To reduce the burden of cost a multifunction smart card has been suggested with a master card issuer franchising space on his card for other card service providers. However, implementing such a system for new payment services such as satellite subscription TV poses substantial administrative and security problems which may delay the commercialisation of such concepts.

In the field of patents smart card ideas have been patented by inventors in a number of countries as well as France. The early use of smart cards will require careful attention to the possible need for licences under some of these patents. Both suppliers and card issuers will need to be meticulous in their study of the published patents and their validity.

Notwithstanding all the above it seems inevitable to the authors that some form of portable personal data carrier will soon come into widespread use in many parts of our society. Whether the smart card as we know it or alternatives such as the optical card, the high density magnetic card or other similar devices will dominate remains to be seen. It is hoped that readers will find answers to some of their questions in this book and that the references given by the authors of the various chapters will lead them to the basic sources of new information on this increasingly important subject area.

P L Hawkes
London
May 1989

Acronyms

AI	Artificial Intelligence
ANSI	American National Standards Institute
API	Automatic Personal Identification
ASCII	American Standard Code for Information Interchange
ATM	Automatic Teller Machine
BTG	British Technology Group
CBC	Cipher Block Chaining
CFB	Cipher Feedback
CMOS	Complementary Metal Oxide Semiconductor
DARPA	Defense Advanced Research Project Agency
DES	Data Encryption Standard
ECG	Electrocardiogram
EDI	Electronic Data Interchange
EFTPOS	Electronic Funds Transfer at the Point of Sale
EPROM	Electrically Programmable Read Only Memory
FAR	False Alarm Rate
FIPS	Federation of Information Processing Societies
FRR	False Rejection Rate
IC	Integrated Circuit
ID	Identity; Identification
INTAMIC	International Association for the Microchip Card
I/O	Input/Output
ISO	International Standards Organisation
IV	Initialisation Variable

KB	Kilobytes
LED	Light Emitting Diode
LF	Low Frequency
LMK	Local Master Key
LPC	Linear Predictor Coefficient
LTS	Long-Term Spectra
MAA	Message Authentication Algorithm
MAC	Message Authentication Code
NPL	National Physical Laboratory
OFB	Output Feedback
OSI	Open Systems Interconnection
PAN	Personal Access Number; Personal Account Number
PC	Personal Computer
PI	Personal Identification
PIN	Personal Identification Number
POS	Point of Sale
PTT	National Public Communications Authority
Q	Q factor of a circuit
QR	Quadratic Residue
RAM	Random Access Memory
RF	Radio Frequency
RSA	Public Key Cryptoalgorithm (Rivest, Shamir and Adleman)
SD	Standard Deviation
SM	Similarity Measure
S/N	Serial Number
SRI	Stanford Research Institute
UV	Ultraviolet
VDU	Visual Display Unit

List of Trademarks

The following trademarks have been used in the text:

CARL
Cotag
Identikit
Identimat
Innovatron
MagnaCard
Qsign
SIGMA/IRIS
SuperCard
SuperSmart
System 7.5
Talisman
UltiCard
UltraSmart Card
UNO
watermark

Chapter 1

Introduction to Integrated Circuit Cards, Tags and Tokens for Automatic Identification

P. L. HAWKES

(British Technology Group)

In which we discover that the smart card is one of a large family of chip-based artefacts for automatic identification.

1.1 INTRODUCTION

Choosing a title for this book was not easy. People want information on the smart card and its applications. Manufacturers' sales literature is a good starting point but is inevitably biased.

A smart card is commonly understood to be a single chip integrated circuit microcomputer built into a plastic credit card. However most of the smart cards in actual use today are not true microcomputers but nearer memory devices. Many are not single chip, chip cards and some of the best and cheapest of these are not even card shaped!

In fact the smart card is but one of many integrated circuit-based data carriers used in a wide variety of computer systems to help identify people, animals, plants, things, messages, events and places. Indeed it is easier to define what is not a chip-based portable data carrier than to produce an overall definition. Concentrating on automatic identification seems to the author as good a basis as any.

Another surprise is that the history of automatic identification via a personal portable data carrier based upon a digital integrated circuit device goes back to 1968 or earlier. The various designs now available reflect the different origins of the data carriers concerned and their prime applications – anti-shoplifting tags, magnetic stripe identity cards, vending cards, pocket calculators etc.

The achievement of M. Moreno and his French licensees and partners has been to focus worldwide commercial attention of one particular class of integrated circuit memory cards. This is the class of miniature artefacts shaped like a standard plastic credit card, having the same dimensions

and containing hardwired or programmed logic as well as digital storage, i.e. the so-called 'smart' or 'intelligent' memory card. In the early 1980s Roy Bright introduced the adjective 'smart' to describe succinctly the essential characteristics of the single chip microcomputer card. His more recent definition distinguishes between the 'active' smart card and 'passive' smart cards. The important features of the former are described in Chapter 2.

In this initial chapter, I will attempt to survey all the silicon chip-based technologies and the perceived needs propelling their creation and uses.

1.2 BASIC FORM AND FUNCTION

Integrated circuit cards, tags and tokens are components in distributed computer and telecommunications systems. Basically they exploit the low cost high density digital storage capacity of integrated circuit memory chips usually, although not invariably, in association with control circuitry known as logic.

As our children are probably now taught in school, integrated electronic circuits are more or less complex arrays of transistors, diodes and other circuit elements and their wiring interconnections formed by printing, diffusion and other processes within a single die or chip of silicon or other semiconducting crystal.

By selective contact printing and etching device, structures down to a few ten millionths of an inch wide are created and enable the resulting chip to record information and process it very rapidly.

With rapid and continuing progress since the early 1970s, integrated circuit making has progressed until today, a single chip IC some half inch square by a few thousandths of an inch thick, can record up to several million bits of digital data as an electronic charge pattern. The microcomputer's logic equivalent can process data at 20 million or more operations a second.

Further increases in information recording density and data processing speed are expected. Made in arrays on six inch diameter wafers, the chip itself sells for a dollar or two.

Like its competitors, magnetic discs and cards and optical discs and cards, the IC chip presents the technologist with a new information recording medium. Using low cost integrated circuit memory as the basic medium, the system designer has a new tool or instrument with which to disseminate and record information *in a system.*

The basic functions enabled by the IC memory chip are the storage of a 100,000 or more bytes (characters) of text or data and their emission or recording in less than a second. Unlike the optical and magnetic media, on-chip logic permits memory access to be controlled autonomously from

within the chip. The implications of this are far reaching as will be described below.

1.3 GENERIC APPLICATIONS

At the present state-of-the-art, the basic form and functions of various IC cards, tags and tokens can conveniently be classified as shown in Table 1.1. The exact form of memory used in these devices varies widely from UV or electrically reprogrammable memory devices to battery backed RAM (random access memory). Particular products and designs categorised in Table 1.1 are best suited to specific applications. These are summarised in Table 1.2.

Table 1.1 Integrated circuit cards, tags and tokens

Type	Typical capacity (bits)	System interface (s)	End-user/card holder interface
Radio tag	64	RF coupling	Via system interface
Memory only card	16K–1M	6–8 electrical contacts	Via system interface
Wired logic 'smart' card	256 up	6–8 electrical contacts	Via system interface
Programmable logic 'smart' card	8K up	6–8 electrical contacts	Via system interface
RF programmable logic 'smart' card	8K up	RF coupling	Via system interface
Active smart card			
(a) Smart Card International 'UltiCard'	8K up	Direct by contacts or indirect by card user	Direct by onboard display and keyboard
(b) Visa 'Supercard'	8K up	Direct by contacts or indirect by card user	Direct by onboard display and keyboard
(c) NPL 'Talisman' token for RSA messages	30K up	Direct by contacts or indirect by card user	Direct by onboard display and keyboard

Table 1.2 Typical applications of integrated circuit cards, tags and tokens

Type	Actual or proposed application
Radio tag	Identification of specific people, animals, places or goods
Memory only card	Distribution medium for computer programs and data
Wired logic 'smart' card	Vending card for making calls from public telephones, etc.
Programmable logic 'smart' card	General purpose including credit and debit card for use in on line and off line payment systems and 'electronic wallet'
RF programmable logic 'smart' card	As above
'Active' smart card	(a) off line payment systems (b) patient data cards in medicine (c) signing and encryption of electronic mail documents (d) metering of the use of gas, water, electricity, TV, public transport etc. (e) logging of events e.g. accesses to premises

1.4 SYSTEMS

The smart card, tag or token is an instrument, usually the 'key' instrument in a complete system designed to provide a service to the end user, i.e. the person carrying the instrument.

The service provider operates and sometimes designs the system. The appropriateness of the particular card, tag or token for a particular service is measured in terms of speed and ease of use, security and cost. Cost reflects both purchase price and cost of use.

Systems are classifiable into two main types – public and private (see Table 1.3). Private systems are intended for use by a closed user group, typically the employees of the organisation operating the system. An access control system for a company's premises is a common example.

Public systems are designed for use by members of the general public, qualified only by a virtue of being customers of a particular bank or users of a particular public service such as the payphone system.

The important public systems are those like credit cards and charge cards which operate internationally as well as nationally. The relevant

Table 1.3 Public and private IC card, tag and token systems

Class	Card population	Card/terminal ratio	Role of standards	Terminal security and price
Private system	tens to thousands	low (10:1 up)	Useful	Both high
Public system	millions	high (50:1 up)	Quintessential	Both generally low

standards are therefore evolving from suppliers' and service providers' standards into international ones via the appropriate national standards bodies, INTAMIC and similar bodies.

Cards, tags and tokens appropriate for public systems tend to be ultra simple to allow customer activation. Low cost is also essential and generally possible because of the large number of standard units involved. This makes them attractive candidates for use in those private systems where the functional limitations can be tolerated.

Operating generally on a single site, over a restricted geographical area or via private networks, private systems can usually afford to have on line real-time telecomunications with each card terminal in constant touch with the system's control centre. This makes the management of card security relatively easy compared with public systems. However, some 'open' sites like hospitals and hotels present particular difficulties associated with the ever changing authorised user population and the risk of attack by criminals and vandals.

Public systems for payment (revenue collection) and the disbursement of money (revenue distribution) are obviously subject to misuse both by legitimate card holders and imposters. This makes on line real-time notification of lost or stolen cards and of account abuse highly desirable. Quick circulation nationally or internationally of 'hot card' lists is however expensive so most systems incorporate a degree of off line operation. This is also of course vital to allow the authorised card holder to obtain some element of usage even if there is a telecommunications failure. Just imagine a bank which told its current account holders they could not use their cheque books because the bank's computer network had problems!

Terminal security and cost are big issues in both types of system. Many of today's terminals are in well protected environments e.g. ATMs on bank premises. Their operation by customer activation can therefore be trusted, This will not be true of many retail shop terminals. Recent scares about computer program 'viruses' demonstrate widespread concern in the industry about the difficulty of trusting personal computer-based terminals.

This may cause a re-evaluation of the security needs and precautions taken when designing, installing and operating PC-based card systems.

A good solution may appear with the new 'active' or super-smart cards (Table 1.1). Having their own keyboard and display this class of device need not rely on a trusted terminal for most of its operations.

1.5 SOFTWARE AND PROTOCOLS

Software includes the programs governing the operation of a programmable electronic device such as the 8-bit single chip microcomputer in a typical 'conventional' smart card. Also included is the operational data which 'personalises' a card, tag or token to the individual authorised end user and the service providing organisation. This data may be programmed into the various types of memory mentioned above, expressed as a wiring pattern (masked programmed) or via fusible electrical links.

Protocols are essentially the rules of conduct by which the card, tag or token communicates with its system or other similar devices. They can be designed in as hardware or software.

Much of the available on-chip memory can be consumed by a stored program for control of the operation of a programmable device. Thus for any very large scale application a bespoke, hardwired solution consumes less chip area and is therefore cheaper. The pay telephone card is a prime example.

1.6 SECURITY THREATS AND THEIR CONTAINMENT

Since the basic purpose of an IC card, tag or token is to identify the bearer to a system, security lies at the heart of all applications. It is therefore not surprising that improved security against misuse by card holders, authorised as well as unauthorised, is often the main selling point for these components. This emphasis has reached the point where the smart card for example is sometimes presented as a panacea for all manner of retail banking and access control systems.

A project sponsored by the author's employers and carried out by the Data Security Team at the National Physical Laboratory, Teddington, has examined the security of smart cards and systems, identified threats from the likely sources and devised appropriate new hardware and software technology to contain the dangers. A prototype version of NPL's 'Talisman' device was developed with the help of Texas Instruments Ltd. Full details are given in Chapter 6. It is described as an integrated circuit 'token' rather than a super-smart card because the recommended size is greater than a credit card and the shape can differ to suit the application.

The main points relating to smart cards used by people are as follows. The card is essentially used to support the card bearer's identity claim. Once read in an authorisation unit (terminal) and accepted as valid the system allows the card bearer to complete a requested transaction. The relevant transactions include:

- Purchase of goods or services
- Access to private premises or computer resources and data
- Sending or receiving telecommunicated messages of value

The threats come from misuse by the authorised card holder, misuse by an unauthorised card holder or where there is collusion between such parties.

Abuse cannot be entirely stopped except at uneconomic cost so a well designed smart card application must contain it. This can be done for example by denying future services to an authorised card holder who has abused his privileges or by catching a thief either in the transaction or later via an audit trail.

The main basic security weakness of the conventional smart card is that it can be stolen and used by an unauthorised card holder.

The established way to guard against this is to only allow card activated transactions where these are supported by the card holder producing a valid PIN (Personal Identity Number). However this PIN must be entered via the keyboard of an authorisation terminal. As already stated this terminal may not always be trustable. If it is bugged a criminal can discover the secret PIN without the card holder's knowledge, copy or steal his smart card and then obtain access to money, goods, services etc. from his account with the card issuing organisation.

NPL's solution to this with its 'Talisman' IC token is to provide a keyboard on the token itself. With a trusted display on the token this keyboard makes the token's use less vulnerable to untrustworthy terminals. Similar solutions are being pursued by Visa and Smart Card International (see Table 1.1. above) under the terminology 'active' smart card.

For many applications of smart cards and tokens, messages need to be sent from the card to a remote mainframe over an insecure network. To prevent eavesdroppers abstracting, delaying, altering or inserting messages the technique of cryptography needs to be employed. Chapter 8 describes these.

The Talisman token incorporates encryption means for generating a cryptographic version of messages sent from the token to remote computers or other tokens such that the message cannot be read by any but the intended recipient and he can authenticate that the message must have come from that token and no other.

PIN details and other confidential data stored in a smart card, passive or active, or in an IC token can be discovered or altered by unauthorised investigation of the IC memory and its data contents. Data alteration is especially likely for smart cards and tokens used as 'electronic wallets', 'cheque books' or meters. Attacks can be logical (via the contacts etc.), electrical (in the same way or by radiation detection) or physical by opening up the unit and reading the data stored therein. Tamper proofing is possible but very costly so most commercial products are best described as 'tamper resistant'. Known means include sensitive 'triggers' which wipe out card stored data when tamper attacks are detected. Easily broken wires buried in a resin potted chip module are one example of triggers. These can be rendered ineffective by deep freezing so they are not a panacea.

Another area of vulnerability is the PIN itself which can be guessed as well as stolen. This has led NPL and others to investigate the uses of so-called 'biometric' techniques whereby some measurement is made of a personal trait of the authorised card holder and compared with an authenticated card stored reference.

The operation of a biometric device is analogous to the 'eyeball' comparison of a handwritten master signature on for example, a conventional credit card with a new specimen produced on demand for a bank cashier or shop assistant. Not surprisingly then automatic signature verification has received a good deal of attention from NPL, SRI/Visa, De La Rue, Thomson and others. It is a well accepted and legally binding commitment to a transaction. All these designs exploit handwriting timing and rhythm as well as signature outline. Such invisible 'dynamic' signature characteristics are very difficult for a forger to reproduce and quite easy for a computer to analyse given an accurate handwriting encoder.

Chapter 7 describes the current state-of-the-art in biometrics including signature dynamics, hand geometry, fingerprints, retinal and hand blood vessel scanning and speaker verification. To be used effectively with a smart card or token the biometric validity decision must be made by the on board microcomputer using locally stored reference data.

Promising solutions leading perhaps to a biometric smart card are being worked on by a partnership between NPL, the British Technology Group and several equipment suppliers and card issuers. These solutions may soon result in a cost-effective biometric smart card or token. Meanwhile an interesting compromise is to store 'mug shots' in digitised form, in a smart card. Human operators of manual terminals can then compare the card stored 'mug shot' with the claimant's appearance and then authorise or deny the requested transaction. This should prove a useful compromise for some markets like physical access control. Clearly it is inappropriate for markets like self-service banking and shopping.

Table 1.4 Choosing IC card media for access control

Application	Example	Key feature				
		Smart Card	Super-Smart Card (Token)	Memory IC Card	Radio Tag	COMM. NIS
(A) Logical access	(1) Reading, writing or erasing financial & medical data in databases	Program controlled data store gives versatility	As smart card but more secure	EPROM versions give audit trail	Hands free (passive) operation	Tag also usable for physical access (B) below
	(2) Electronic Data Interchange including EFTPOS	Message integrity by private key cryptographic s/w	Message integrity by public key cryptographic s/w	—	—	Latent demand for EDI should generate a substantial token market
	(3) Point to multipoint data distribution	SDI by private key cryptographic s/w	SDI by public key cryptographic s/w	—	—	Teletex & satellite opportunities
(B) Physical access to premises & sites	(1) Employee access	Data store holds personal & biometric details	Secure PIN validation	Store big enough for mugshots	Hands free (passive) operation	
	(2) Employee location by zone	—	—	—	Passive operation	Tag identification by terminals at zone boundaries

Table 1.4 Choosing IC card media for access control

Application	Example	Key feature				
		Smart Card	Super-Smart Card (Token)	Memory IC Card	Radio Tag	COMM. NIS
(C) Physical access to routes (travel)	Road tolls, airline & train season tickets	—	—	EFROM based security	Passive operation	Moving vehicle problem for tags
(D) Telecommunications	Public telephone cards	Rechargeable	Rechargeable	EPROM based security	—	The leading application in the world (20–30 million)
A-Z Multifunction cards or tags	Employee access, work log & discount card	Program controlled datastore gives versatility	Power and security for most purposes	—	—	Super smart card/ RF tag combination would meet all listed needs but so would suitably programmed CT2/ pager device

1.7 OTHER DEVELOPMENTS

Before the ISO standard smart cards are established internationally new designs are appearing with alternative or additional features to open up new applications.

Chapter 3 describes the GEC ic card with its secure low cost RF coupling method for card to terminal interaction.

Two other developments worthy of note come from the opposite ends of the product spectrum of Table 1.1.

The humble radio tag has now fully established itself as a viable solution to the access control problem (Table 1.4). There are over fifty suppliers worldwide. In this country Cotag and its competitors have delivered hundreds of systems to the smaller organisations with a need to restrict site entry to a few hundred employees and some authorised visitors. The systems work well and are cost-effective. John Falk of Cotag describes radio tags and their manifold uses in Chapter 4.

1.8 FUTURE PROSPECTS

As the still fledging industry matures there seem to be two opposing tendencies. The first is to migrate towards very low cost standard devices manufactured on a huge scale.

At the opposite end of the spectrum are the active devices like the NPL's Talisman Token. In the author's view these different approaches will coexist.

There may also be scope for the integration of the identification and metering functions of the active smart cards and tokens to be integrated as software into other products like conventional and portable terminals and telephones.

Chapter 2

Smart Card Technology– A US Pioneer's Viewpoint

ARLEN RICHARD LESSIN

(Chairman & President, Lessin Technology Group, Inc)

The early pioneers were visionary, seeding the not yet existing market for a then unknown technology. Those activities, however, are now making possible diversified and economically feasible applications.

2.1 INTRODUCTION

The smart card entered the US very quietly in 1980. Drama followed quickly, but this story will have to be part of another book. The event occurred at a November international meeting called INTELCOM '80. The place was the Los Angeles Convention Center. Amidst chromium tubes in angular designs and free-flowing vintage champagne, the French government introduced its new telecommunications capabilities and some other technology developments. The main exhibit area was generally quite impressive.

The smart card (Carte à Mémoire) display was not impressive – composed of a few mock-up 4KB capability cards and a non-working prototype terminal sitting casually on a red table cover. This exhibit was less than centrally located. Having talked with some of the attendees later, none of those who passed by or stopped at that table saw the technology displayed there and believing they had seen anything important. This writer was apparently the major exception. Having visited the exhibit and being immediately captured by its possibilities, a light switched on for me. That conference marked the start of my long involvement in the technology, including two years as special consultant to the French government, introducing smart cards to the US. Whether I have been justified in my faith is still in the proving stage, however, in 1989 the future appears bright – but not yet here. Therefore I have decided to meet the issue head-on by establishing the Lessin Technology Group, Inc. (TLC). This New York based consultative and systems organisation was mandated to accelerate the acceptance of smart card and related technologies in the US and internationally.

2.2 EARLY DEVELOPMENT

In 1980 the US government had already granted the five key US patents on smart card technology to Innovatron. That company was founded by French financial/technology journalist, Roland Moreno in 1974 after he invented the key elements of the smart card. At the time, international promotion of the technology was in the hands of Intelmatique, the French government's international technology group, which was responsible for promoting the introduction of the smart card along with several other French products. These included videotex systems, electronic directory systems, low cost consumer terminals, low cost facsimile terminals, telewriting, and audiographics teleconferencing. The key elements, however, proved to be videotex and smart cards; and at the time, videotex was clearly the Government's central concern and smart cards a quite secondary one. In 1980, anyone predicting that nine years later the smart card would be more prominent than videotex was considered either a visionary or, more likely, less than observant of apparent realities.

In France, the banks gave the production version 8 Kbit EPROM smart card early heavy support. To this day, as a result, its potential financial applications are the primary emphasis in much of the literature. But from the start, those involved in the technology saw that it had potential uses that went well beyond the financial industry. For example, it could carry personal medical history, give personal access to confidential information, or serve as a key carrying one's retinal scan or digitised fingerprint or photograph to provide access to areas of great physical security. Philips Data Systems, one of the original three French card and terminal manufacturers licensed by Innovatron, published a descriptive graph in 1982 suggesting that smart cards could be used for payment cards, prepaid cards, subscriber cards for telephones and transport cable TV, authorisation and access, guarantees and service for automobiles, applications and social security services. Some of those applications – and a few not on the Philips list – have been implemented recently. So the ideas behind them aren't all new; it is rather more a signal example of use beginning to catch up with vision. Vision is always parent to the reality.

It took nearly two years for the smart card to formally surface again in the US, although there had been substantial media attention to its potentialities and much work behind the scenes on its behalf had been done. The year 1982 is significant in smart card history for several reasons. In France it was the year that the first home banking system using smart cards went into operation, its first terminal installed in an apartment in Velizy, a Paris suburb, and also the year of the first major Point of Sale (POS) trials. Two trials also started that year in the US.

- First Bank Systems of Minneapolis ran a small smart card trial involving 10 North Dakota farmers, in conjunction with a home banking system videotex trial.

- The US Department of Defense issued a few hundred cards to soldiers at Fort Lee, Virginia. The cards were to be used for identification at post entry and exit points and at the post exchange (PX) where US soldiers and their families can buy goods at greatly reduced prices. At the time, the US Army was looking for an alternative to the cardboard photo ID cards it had been using for generations. The smart card was one of several technologies tried out at different military bases in the US. The smart cards were used in special point-of-sale terminals. These first US experiments did not lead to substantial distribution of smart cards. The Department of Defense in effect made no decision on the smart card – the program started at Fort Lee was shelved. However, currently the Pentagon has been considering issuing a purchase order for several hundred thousand to millions of cards. However, this interest has not yet been implemented.

These were the only visible US trials of smart card technology. In general, US banks were resistant to the technology, essentially because of their major commitment to magnetic stripe implementation. However, several other important activities were actually going on behind the scenes. Chase Manhattan Bank, Security Pacific Bank, American Express, Bank of America, and the US Department of Agriculture, among others, launched major research projects into the potential uses of smart card technology. Some of those experiments have led to new applications of the cards, while other organisations are still waiting and watching the technology develop (including experimental laboratory development and limited test implementations of a non-contact version of the technology by AT&T among others). Constraints to acceptance in the US included limited memory, requirement for readers/terminals at all transaction locations, high cost, no reusability, and the prejudice against foreign technology – the NIH or 'not invented here' syndrome.

Throughout this time, most of the technological development of smart cards was going on in France and, to a lesser extent, Japan. US companies were basically playing with the French technology and often trying out applications that were already in commercial development in France. All of these French and Japanese cards were 'passive' cards, based on a new terminal infrastructure being implemented. Some initial reasons for US slowness to implement smart card technology have been non-updateability of most cards, commitment to magnetic stripe card systems, the need to supply power to the card via an originally costly smart card terminal/reader infrastructure, and the fact that the applications software was *not* in the card, but in the terminal or a PC or computer beyond the terminal. This made changes costly and difficult. That requirement changed in October 1985 when Visa announced its program to test the Super Card – a value added, multiple function, large memory, charge card of the future. It was to be an advanced smart card with its own built-in terminal – including a keyboard, display screen and battery power all in a credit card package. The applications software resides inside the card in

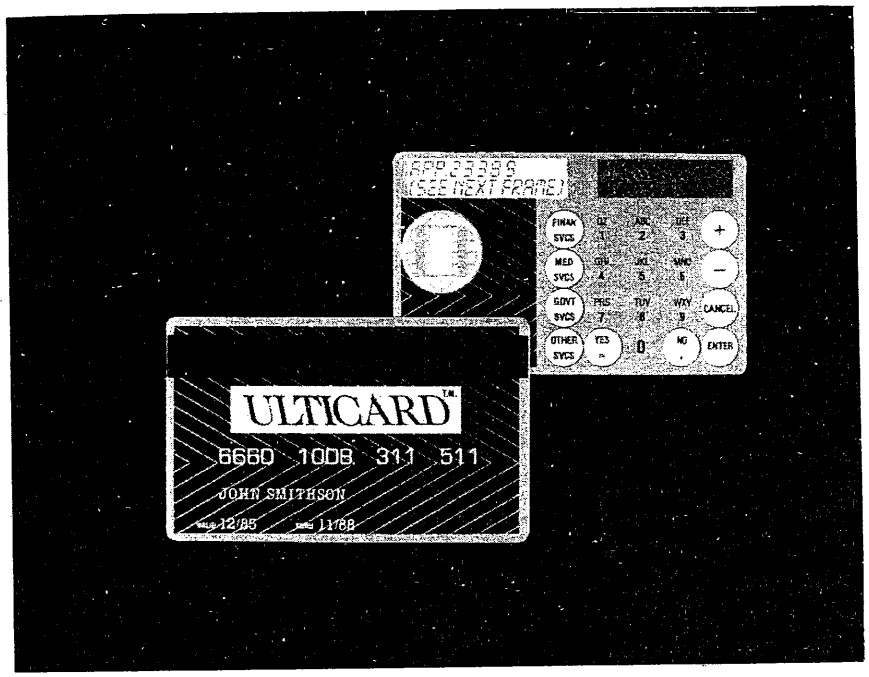

Fig. 2.1 The UltraSmart Card (formely Ulticard).

the SmartCard International (SCI) UltraSmart type of Super Card and in its SCI MagnaCard (conventional type). It was July 1985 when SCI proposed the A. R. Lessin concept for such a card to Visa. In November this new smart card had 64 Kbits of reprogrammable memory, a two-line readable display, alpha-numeric keypad, and battery power. SCI and Visa signed a development contract in January 1986 and SCI delivered the first working prototype Super Cards at the Visa International board meeting in San Juan, Puerto Rico in May 1986, only 16 weeks after the contract was signed. SCI calls this card, UltraSmart Card, (shown in Figure 2.1), Super Card being one version and an increasingly generic name for this version of smart card technology (although 'SuperSmart' has been registered by Visa). Since then, SCI has worked with Arthur D. Little, Texas Instruments, and other American companies in continuing to develop the UltraSmart Card technology and to find new applications for it.

2.3 NEW GENERATION SMART CARDS

This new generation of self-contained smart cards is a major advance in the technology. It radically reduces the need for terminals -- although it can be compatible with most imprint, magnetic card swipe, smart card, and other point-of-sale terminals. (The present generation of this kind of cards are in the process of becoming compatible with ATMs.) To make

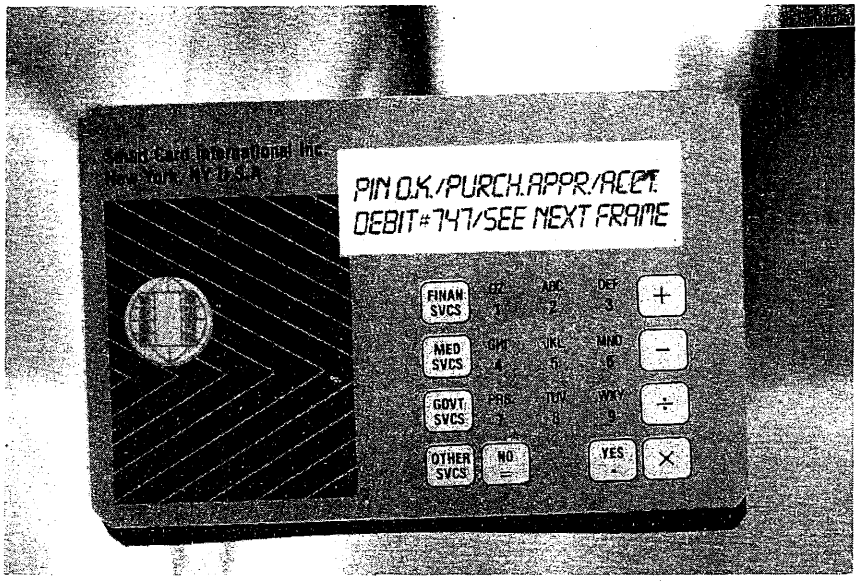

Fig. 2.2 A PIN unlocks the UltraSmart Card for purchase use or access to other functions, services and multiple applications.

a purchase, the user unlocks the card with the proper personal identification number (PIN Figure 2.2). The clerk enters the relevant information on the card's keypad regarding the product being charged. The processor checks the credit balance and displays an authorisation code which the clerk copies onto a standard credit slip, or enters into an electronic terminal, for transmission to a data bank.

UltraSmart Card is much more than a simple smart card with a built-in screen. It is a complete portable computer – informally named 'The Pocket PC' – essentially credit card size. Previous versions of the smart card have suffered because they had no way to address memory logically in the way it is done in standard computers. Instead all data has to be stored and retrieved by physical address. Thus users could not simply tell the card that they wanted to see such and such a data file, as they can on a computer. Instead, the card issuer – or whoever designs the retrieval software – must have intimate knowledge of the arrangement of the card's memory and must tell the card the precise physical addresses to go to get the data. This has made creating application software much more cumbersome and expensive.

UltraSmart Card – and SCI's conventional smart card, MagnaCard – were the first smart cards to implement a full operating system, although recently there have been others. However, the proprietary SCI system, CARL (CARd Language), handles all memory management in the same way the operating system of a normal computer does. Thus data can be accessed by logical memory management, in the same way the operating

system of a normal computer does. Data can be accessed by a logical file name, regardless of where it is stored in the computer. This greatly facilitates development of application software. It allows several different applications – such as in the Visa version to act as a credit card account, an electronic checking (debit) account, and an appointment calendar or address file – to share a single card since the data for each application can be protected from the others. The UltraSmart Card version extends this multifunctionality to medical, government, insurance, educational, professional and virtually an unlimited variety of other applications. In fact its developers anticipate that this type of technology will normally be used in this manner. Also if desired, two or more applications can share a single data file even further expanding capabilities.

2.4 FINANCIAL USES

MasterCard in 1984 had already announced its own tests of the traditional smart card technology. In late 1985 it mounted tests of both French and Japanese versions of the 'passive' technology involving thousands of cards in point of sale (POS) applications in Columbia, Maryland, and Palm Beach, Florida. One of MasterCard's main reasons for testing the smart card was its perception of existing and projected credit card fraud. An example of card fraud would occur when someone steals your card along with your wallet in New York. He then flies it to California where the stolen card is used illegally for three days to run up hundreds of dollars in bills for items which are then sold on the black market. This has been a growing problem and is potentially a big business, possibly involving organised crime in the US. MasterCard blames it in part for $900 million in Visa and MasterCard losses in 1984 and has estimated that those losses will exceed $2 billion in 1990 – more than the national debt of some countries. MasterCard was attracted to the smart card because of its superior security, which it believed would stop most of the fraud.

Unfortunately, the MasterCard tests – which were intended to determine whether consumers would accept the technology – proved inconclusive. This was not because the cards or terminals failed or because consumers didn't like them. It was because those running the program did not adequately educate either card holders or retailers in using the technology properly and also the cards and terminals were not adequately integrated into the POS system. This is, I believe, an absolute requirement in regard to utilising smart card hardware and software. They must be integral parts of the overall user system. This caused MasterCard to pull back. In fact both MasterCard and Visa reached the conclusion that they must co-operate in the introduction of any new charge card technology. Installing millions of point-of-sale terminals and training millions of store clerks to use the new technology is too expensive a proposition for either to do alone.

Unfortunately the two companies disagreed – and continue to disagree – on the basic reasons for being interested in the technology. MasterCard remains convinced that the elimination of credit card fraud and reduction in credit losses will be enough to justify the expense of converting from magnetic stripe to smart card technology. Visa sees combating these problems as a secondary issue. It is mainly interested in providing cardholders with new services, and it is not willing to go ahead in a major implementation of smart card technology unless it is convinced that these extra services will generate the revenues to pay for conversion. Therefore, Visa's Super Card project concentrates on adding new services such as electronic checking, conversion of currency into international denominations, electronic note pad, calculator and clock. Also, elimination of the smart card terminal infrastructure is very important for Visa, because it changes the worldwide economic justification for the payment function drastically. The objective is generating profits, **not** new bank technology.

The real answer, of course, is that both organisations are right. US banks are facing a business crisis caused by the deregulation of their industry and the entry of dynamic new competitors who have already taken away a large amount of their traditional business. They need to offer new services in order to hold onto the business they have left. At the same time credit card fraud is reaching material proportions. If it isn't stopped, it will become a major problem for the US banking industry in the 1990s. The UltraSmart 'active' card could provide solutions to both.

A 1987–88 smart card study performed by Booz-Allen-Hamilton, the large US consulting firm, was funded jointly by Visa and MasterCard. Focusing on the constraints and costs of implementing conventional smart card systems, it put a damper on short-term ultilisation of this form of the technology. However, it did not address the facts of UltiCard technology and the opportunities it offers in requiring only limited reader infrastructure (being a reader as well as a processor and memory device itself) and therefore radically reducing widespread implementation costs. A current Frost and Sullivan study predicts a $20–25 billion business in the mid 1990s.

Unfortunately, in late 1989 the Visa/MasterCard debate shows no signs of resolution, and blocks full-scale inplementation of either standard smart card or unified (UltraSmart Card) technology in the US financial industry. The large banks and other major credit card companies such as American Express and Carte Blanche adopted a wait-and-see attitude, also probably put off in part by the initial cost of converting from magnetic stripe technology. What will change their positions will probably be recognition of such factors as the economies of reuseability and longevity of smart cards, off line Super Card (UltraSmart Card) capabilities, the cost of credit card fraud, and the cost of the phone lines needed for on line credit authorisation that is the main present day (1989) answer to fraud

growth, – led by identified opportunities for new revenue sources. Thereby, the pressure to adopt the alternative, vastly more secure, totally off line system provided by smart card technology will likely increase until it portends to become irresistible even to the most conservative and change-resistant bankers and banks.

2.5 AGRICULTURAL USES

While development of strictly financial uses has slowed, other US applications have actively developed. For instance, in the three years (1986–1988) the US Department of Agriculture (DoA) has instituted a program using smart cards to track production quotas for peanut farmers. Under US law, farmers are given yearly quotas on the amount of peanuts they can sell. If they exceed those quotas, they are heavily fined. Because farmers typically do not bring all their crop to market at once, or sell it all to one buyer, it is difficult for farmers to make sure of what they are permitted without exceeding their allowances. Copies of the records of each sale used to be mailed to the DoA where they were reviewed and an updated statement of the remaining quota mailed back – a process that often took two weeks – during which time the farmer may well have made further sales. Thus farmers could easily make costly mistakes.

In 1985–86, the Department of Agriculture began issuing conventional smart cards to farmers in a pilot program that has since grown into the first major full-blown implementation of the technology in the US with 58,000 cards issued to farmers. Each card contains the quota for the farmer to which it is issued. Each time farmers sell peanuts they deduct the amount they sell from the card using standard point-of-sale terminals. This allows everyone involved to know whether the sale falls within the allowable quota and how much the farmer has left on his quota immediately and accurately. The quota program, which is designed to benefit the farmers, has been made much more efficient and most quota overruns eliminated. This program is now being supplemented by a similar one that will apply to tobacco farmers.

2.6 SECURITY USES

Conventional smart cards are also finding a place as electronic 'keys' to high security areas such as mainframe computer installations of such companies as Bank of America and Royal Bank of Canada, among others. Smart cards are excellent for this because they can carry multiple security levels. For instance, the SCI MagnaCard can contain up to 22 different keys or levels of security, each of up to 254 characters. The National Security Agency (NSA) will utilise these capabilities as part of a security access program scheduled to involve one + million cards from three US suppliers in the 1989–92 time frame. These can be nested so that they would have to be entered in a specific order. Also, the card can

be programmed to refuse to operate after a predetermined number of incorrect consecutive tries are made at entering a key (usually three). Furthermore, smart cards can carry biometric information such as digitalised photographs, fingerprint or retinal scan data for use with very high security devices. The smart card is capable of holding this information in a very secure manner, making it extremely difficult to alter or forge.

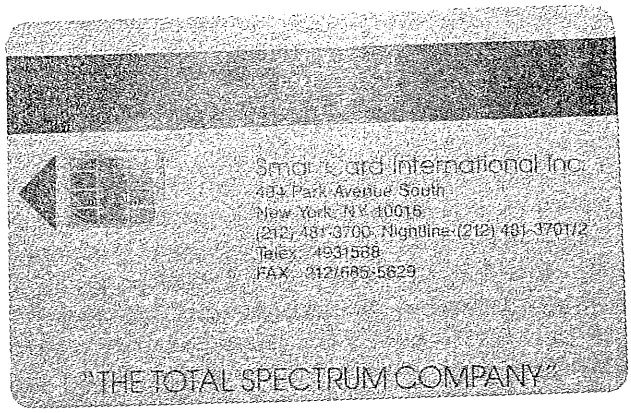

Fig. 2.3 The SCI MagnaCard. This example is of a traditional type of smart card, issued in other forms by various suppliers. It can contain different sizes of fixed or eraseable memories. It requires a reader/terminal for access.

2.7 MEDICAL USES

UltraSmart Card type technology is also finding new applications, particularly in medicine. Modern health care may be the most specialised, decentralised, and information-intensive of all human activities. Better information at the right time and place means better health care. More efficient information technology means lower costs.

Medical and health care professionals have been quick to see the possibilities of smart cards, to improve care and contain costs. UltraSmart Card, for example, has the memory capacity to store and revise extensive, dynamic, individual medical records. UltraSmart Card provides complete multilevel security, allowing open access to emergency information while preserving individual privacy and protecting doctor-patient confidentiality. UltraSmart Card's stand-alone capability is ideal for self-monitoring regimens, home care, regular medication or treatment regimens and other outpatient programs and services.

The Methodist Hospital's Institute for Preventive Medicine and the Baylor University College of Medicine, both in Houston, perhaps the world's largest medical complex, developed and tested a medical systems UltraSmart Card in partnership with SCI (Figure 2.4). This multifunction card was designed to be used by patients with heart disease, diabetes and hypertension as part of a medical regimen. Selected over-weight patients could also use it as part of a self-monitoring regimen designed to modify

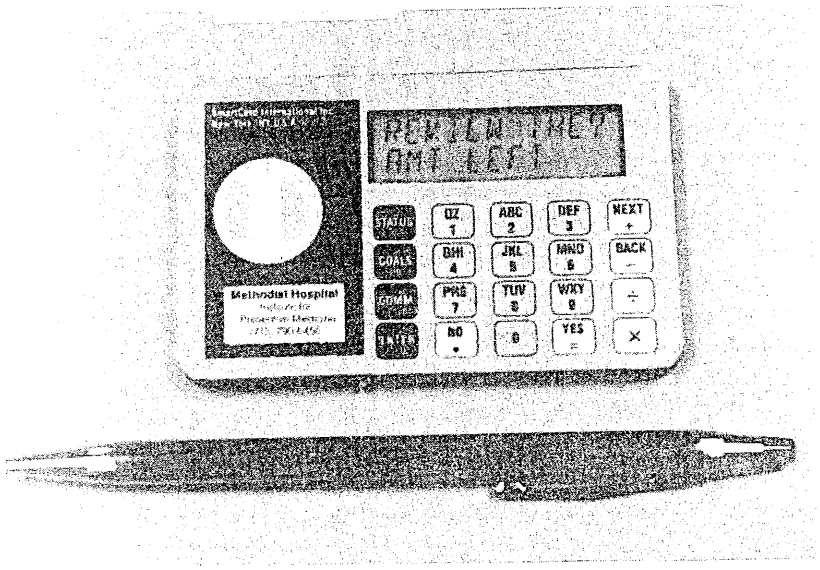

Fig. 2.4 Medical systems UltraSmart Card.

behaviour patterns associated with overeating. This program could be further enhanced by a US government National Institutes of Health Research study grant to SCI. Patients could use the card to keep track of food intake, exercise, weight, medication, use of alcohol or tobacco and behavioural correlates (stress, boredom) between clinical visits.

Dr J. Alan Herd, formerly Medical Director of the Institute for Preventive Medicine, described the potentially significant impact of 'computer-in-a-pocket' technology on the success of health-related behavioural modification. During the past 15 years, health professionals have developed programs to help control many chronic diseases. In addition to scheduled medication, most treatment programs require patients to keep written records of health-related behaviours. Referring to these programs, Dr Herd praised the increased reliability and accuracy of UltraSmart Card type records and the capability to offload those through a PC interface for graphic presentation. 'The immediacy of feedback and the precision of reports greatly enhance the opportunity for educating and improving performances of subjects when changing dietary patterns in an important part of the medical regimen,' he said.

InfoMed, the largest US company committed to providing custom information management systems to home care agencies nationwide, seeks to reduce health care costs and free maximum staff for patient care by increasing administrative efficiency. Current InfoMed services include distributed data processing systems, advanced home care information systems, stand-alone business application software, personal computer systems, and a handheld nursing computer – potentially one of the first

important UltraSmart Card type of applications. There are 10,000 home nursing agencies in the US.

The UltraSmart Card type of Nursing Computer could perform two major tasks for each nurse – daily scheduling and patient care programming. Updated daily via a read-write interface with a local office PC, the Nursing Computer would carry a complete plan of the prescribed medical regimen for each patient to be visited that day including medications, dosages, frequencies, possible side effects, and handling instructions. It would also contain patients' allergies, patients' names, addresses, telephone numbers, and care hours scheduled for each patient, and the names and phone numbers of their pharmacies.

The nurses could record each action they take with each patient, as they do it via the keyboard. At the end of each day this information would be transmitted electronically into the PC at the nursing station for billing, report generation, and such other administrative needs as complying with government regulations. Preliminary indications are that visiting nurses will be able to increase their patient load from the current five per day to as many as seven. Clearly this example manifests a potentially major application of the benefits to all involved parties, from provider of technology through to patient, of this type of system.

InfoMed is also evaluating a patient computer that could prompt and record self-monitoring regimens, self-administered medications, outpatient treatment appointments, and home care nursing visits. Each patient's emergency medical information file could be immediately accessible to anyone. A personal medical history, accessible only to the patient and doctor, could also be included.

2.8 INSURANCE SALES AID

Connecticut Mutual Insurance Co., other insurance companies and SCI have jointly tested UltraSmart Card as a sales aid for insurance agents. UltraSmart Card type of technology is capable of instantaneously calculating insurance rates utilising the most current insurance tables and can also simultaneously store the actual rates of several customers. The customer data can be quoted and can be brought back to the field office and transferred to a PC or mainframe for further processing or record keeping purposes. Insurance tables may be updated and downloaded at any time by inserting the UltraSmart Card into a card reader which has been connected to a PC at the home office.

To use this type of card, agents enter their PIN which protects sensitive data from unauthorised access. They specify the type of insurance required for quotation by the client, existing or prospective, from a menu on the card – one card can handle calculations for several different kinds of insurance. They choose an appropriate insurance plan from a menu. They are prompted to enter the sex and age of the client and whether they

smoke. Then they enter the face amount of the policy. The card calculates and displays the standard annual premium and the preferred annual premium. It displays the guaranteed cash value of the policy, total death benefit at age 65, and such options as dividends, surrender values and available load amount for whatever year is specified. This process enables information to be delivered immediately in either home or office, creates better client service, and creates more revenues for both the agent and his company. Other insurance carriers have indicated interest in following these types of applications.

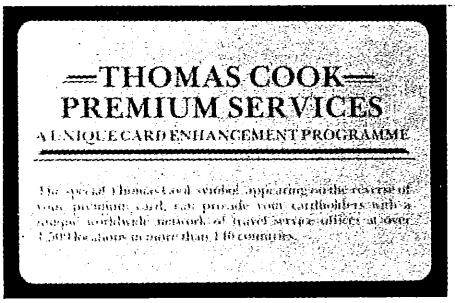

Fig. 2.5 One side of the UltraSmart Card prototype designed to deliver travel and related financial services.

2.9 TRAVEL AND RELATED FINANCIAL SERVICES

Via both its US and UK offices and marketing/engineering facilities, SCI and Thomas Cook (Peterborough, UK) in 1987, signed a very significant development contract for prototypes of the first advanced (UltraSmart Card) smart card for the delivery of travel and related financial services (Fig. 2.5). The card provides its built-in keypad and display screen to record and monitor for travellers their spending on such items as expenses, and electronic travellers cheques or electronic cash, plus itinerary planning, airline and hotel preferences and corporate/organisation travel policies.

Peter Middleton, Chief Executive of Thomas Cook described the application: 'Our commitment to this major project reflects our determination to deliver travel and financial services with greater efficiency at reduced costs. While these services will enhance our core business, they will also simplify travel complexities before, during and after journeys. 'For example, once back in the office the traveller can insert his or her card into a smart card reader attached to a PC and immediately receive a hard copy print-out of incurred expenses collated by category.'

The Cook's prototypes were delivered in 1989 and are expected to be the forerunner of production numbers for future major use by Cook and other related organisations worldwide.

2.10 FUTURE DEVELOPMENT

As exciting as these developments are, they are only the beginning for smart card technology in the world. Back in 1982, this writer developed a probable timetable for smart card implementation in North America that so far has proven to be valid and which I believe will continue to be accurate. It projected that in 1983–1985 the first early tests would take place in the US, and that happened. It also predicted customisation of the technology and development of the early market, both of which happened. In the short-range future – 1986–1988 – it predicted the appearance of the first substantial applications and the beginnings of proliferation of uses, and that occurred on schedule. In the mid-range future – 1989–1992 – it projected the development of more extensive applications, initiation of mass manufacturing, and the creation of a major industry. With mass production, smart cards can be expected to cost anywhere from 50 cents to $50 each, depending on what they have built into them and what software is required in them. And in the longer-range future – 1992–2000 – my timetable projected worldwide, applications – essentially smart cards in many aspects of human and organisational life.

This does not mean that every human being will have a smart card; however, for a very large number of the world's populace smart cards are likely to manifest themselves in all kinds, shapes, and sizes. This is because they are the logical extension of distributed processing – from paper to desk top computer to pocket or purse PC. We will probably never totally eliminate cash or paper, but we will truncate the use of both. The reason all this will happen is that the smart card technology – and in future applications the UltraSmart Card/SuperCard form of it – will meet the immediate and daily needs of a large number of present and potential users, whether in a city or tribal wilderness, to access government entitlements.

Potential smart card users today are looking for solutions to the increasingly vital problems of managing information effectively and conveniently. Bank cheque customers want a cheque book that automatically balances, that guarantees that no one can forge a cheque and is always available when wanted. Medical patients want an easy way to keep track of their diet without having to carry around a notebook and pencil and constantly scribble down information and add up lists of calories. Government wants to provide human services to its citizens more effectively and less expensively.

However, vendors/manufacturers of smart cards must learn to conform to customer requirements which go beyond selling products. Virtually all these actual and potential users want most or all of five criteria to be met in a system context:

(1) applicability to their needs and conformity with existing operations environments and/or systems;

(2) easy delivery and accessibility of information;
(3) security, which has been a hallmark of smart card from its inception;
(4) flexibility, which the technology provides inclusive of such improvements as logical data file storage; and
(5) economic practicality and viability.

As new generations of the technology appear with more improvements in processor speed and memory capacity, these devices will inevitably find many more applications in the government, medical, industrial, petroleum, retail, auto rental, insurance and many other industries. They will be adopted as handy devices for accessing information on networks – a skill that is already becoming vital to many professionals. They will also likely replace floppy discs because of security as well as increased memory capabilities. Transportation companies will use them instead of tokens and public phones and vending machines will use them instead of coins, but most applications probably have yet to be identified and developed.

Much of all this is happening now in France and Western Europe and in Japan. And yes, eventually even MasterCard and Visa will use them in one form or another, pressured by their acceptance by such a broad cross-section of the marketplace. In the US, as in other countries, the entire thrust of technology today is towards increasing computerisation, towards more effective control and dissemination of information in cars and appliances, home security and energy control, as well as throughout offices. This makes the replacement of the present, 40-year-old magnetic stripe technology with the new 'smart' digital technology and its vastly superior information/service delivery capabilities inevitable both in everyday applications, a host of new ones, and eventually in the 1990s in traditional credit card applications as well.

There is a growing perception in the US and elsewhere among independent research groups, technologists, smart card users and potential users that the smart card will change the way we do things – that smart card technology will change the world. It is likely, despite a slower start, to take hold most broadly and profoundly in the US. It is, I believe, certain to be the major disseminated information technology in the world as we enter the 21st century.

SUMMARY –
AN AMERICAN PERSPECTIVE CHRONOLOGY
OF THE HISTORY OF SMART CARD TECHNOLOGY

1970	Dr Kunitaka Arimura files first basic smart card patent in Tokyo (Japan only).
1974	Roland Moreno (A French financial/technology journalist) invents and files first broad based smart card patents in

	France and major industrial countries worldwide. (There were some 75 licences worldwide by 1989.)
1975	First of five Moreno patents granted to Innovatron (French patent holding company founded by Moreno).
1976–78	First licences to develop the technology sold to three French companies (Honeywell Bull, Schlumberger, Philips by lnnovatron).
1975–79	All five patent filings in US granted to Innovatron.
1979	French government (PTT) founds Intelmatique as an international organisation to introduce and promote smart card (Carte à Mémoire) and other new French technologies (including Videotex).
1980	French government (Intelmatique) retains Arlen R. Lessin and his consulting firm to represent it and its developing smart card industry in North America.
1981–83	Smart card technology is introduced to the world. Japanese and others start research. French launch first pilots in France and the US.
1981	First US pilot using French terminals and Bull cards as home banking test in North Dakota.
1982	US Department of Defense tests military ID Philips Data Systems smart cards at Fort Lee, Virginia.
1983	Lessin founds SmartCard International, Inc. (SCI) with Moreno's endorsement to work seriously on developing product and defining the US market. SCI negotiates the first non-French card patent licence with Moreno and Innovatron. SCI is the first dedicated smart card company in the world.
1983	Bull (Microcard) starts US operation.
1983–85	SCI performs research (market and product technology) and is retained as smart card consultant by such companies as AT&T, Visa, and Texas Instruments.
1985	SCI proposes to Visa the Lessin concept of a 'super smart card' which would contain large reprogrammeable and updateable memory, keyboard, display and self power – potentially also utilising biometrics and high security algorithms and crytography. It would vastly enlarge capabilities for services rendered and revenue potentials.
1985–89	Innovatron grants licences to multiple companies (mostly US, French and Japanese) to enhance commercial capabilities of technology. Tests begin in Japan, Norway, Italy, Britain, Finland, Germany, New Zealand and Australia.
1986	SCI signs agreement with VISA to develop and produce 100 'super smart card' prototypes and delivers first working card 16 weeks later.
1986	AT&T introduces first non-contact smart card for use in

public phones and other potential applications.

1986 Bull Microcard awarded conventional smart card contract by US Department of Agriculture for peanut farmer application.

1986 SCI completes public securities offering and becomes first full spectrum publicly funded and traded smart card company in the world.

1987 SCI files US and international patents on its UltraSmart Card and also proprietary software and hardware technology and operating system software.

1987 PC3 issued smart card patent.

1987 Multiple tests/trials begin internationally.

- France commits to reaching 32 million cards by the end of 1988 – including bank cards.

1987 Smart Card Applications and Technologies (SCAT) organisation founded and holds first conference in US.

- US interest increases significantly, especially from Federal and State governments.
- SCI completes SuperCard prototype deliveries to Visa.
- Bull (Microcard) signs contract for US Marines security card. Some small US companies market smart card test contracts.
- Schlumberger contracts to supply smart cards to State of Michigan for job placement ID and other services.

1988 Significant market events:

- SCAT breeds ESCAT (European Smart Card Applications and Technologies) meetings (1988 and 1989) in Helsinki. Also, ASCAT (Asian Smart Card Applications and Technologies) is being organised and its first conference is set for Tokyo in 1990.
- Market develops substantively in other than banking sector.
- First US UltraSmart Card technology patent is issued to Lessin and other SCI associates (patent assigned to SCI).
- SCI signs contract with Hawaii state government to develop smart card model program to render human services to residents (SCI UltraSmart Card type).
- Kodak-Pathé contracts for smart card transportation application (SCI UltraSmart Card type).
- AVIS contracts for smart card auto leasing service tracking application (SCI UltraSmart Card type).
- Schlumberger, PC3, SCI approved by US National Security Agency (NSA) for security application smart cards.

1989

- 'SmartStart' for secure access to PCs developed (SCI).
- Trials of first supermarket POS smart card applications in the US (PC3).
- US Veteran's Administration issues request for proposal for smart cards containing medical profiles of infirm veterans.
- US Department of Agriculture issues request for proposal for smart cards for 1990 food stamp delivery trial.
- State of Michigan broadens its smart 'opportunity card' program for state residents.
- AT&T and Olivetti reach agreements for Olivetti to purchase and issue AT&T card for mass transportation application.
- Lessin leaves SCI to found Lessin Technology Group, Inc. (LTG) with mandate to provide smart card and related technology consultative and systems services world-wide. LTG is first company to address the bringing technology 'from concept to practice'.
- LTG establishes European office in UK. Roy D. Bright, former head of Intelmatique (France), is Managing Director.
- LTG and SCAT organisation agree to hold executive briefings in US and overseas under the aegis of the International Smart Card Institute (ISCI). Its purpose is to disseminate unbiased information about the technology and its realistic applications. The President of ISCI is Lessin.
- Trials and plans for applications profliferate in wider ranges, including newly identified areas.

1990–93 Watershed years for the acceptance internationally of smart card and related technologies.

Chapter 3

A Contactless Smart Card and its Applications

JOHN McCRINDLE

(Marconi Research Centre)

We discover a very practical alternative to the contacts of the original smart card.

3.1 INTRODUCTION

The late 1940s and 1950s saw the introduction and growth in numbers of the plastic card. By the early 1970s their use was widespread in areas such as finance, travel and entertainment. At the same time, computers developed rapidly, culminating with the introduction of the microprocessor in 1971. This device was to appear in all forms of equipment and consumer products from spacecraft to washing machines.

The idea of combining these two products by embedding a microprocessor in a plastic card was conceived in the mid 1970s. Today, this product, the smart card, looks like having a greater impact than the plastic card and microprocessor combined. During the next decade it will pervade all areas of our lives as it is used in finance, medicine, the armed services and telecommunications, and in more far-reaching applications such as passport replacements. It is also set to change fundamental ways of working that have been carried out for centuries as it becomes a direct electronic replacement for money.

The first applications of smart cards took place in France using cards which had surface contacts for powering up the electronics within the card, and for communications. This type of card requires a read/write unit with a slot into which the card has to be presented correctly and forced home. Operational reliability is very much dependent on the careful treatment of both the card and the read/write unit which are prone to wear, contamination and damage (deliberate or otherwise). In many cases, before a new product is widely accepted it is superseded by a new generation which becomes the internationally recognised product. The GEC ic Card is one of a new type of smart card which is now emerging. As a result of research which began in 1983, GEC Card Technology has overcome the deficiencies of the earlier contact type smart card with the introduction of a revolutionary new concept – a

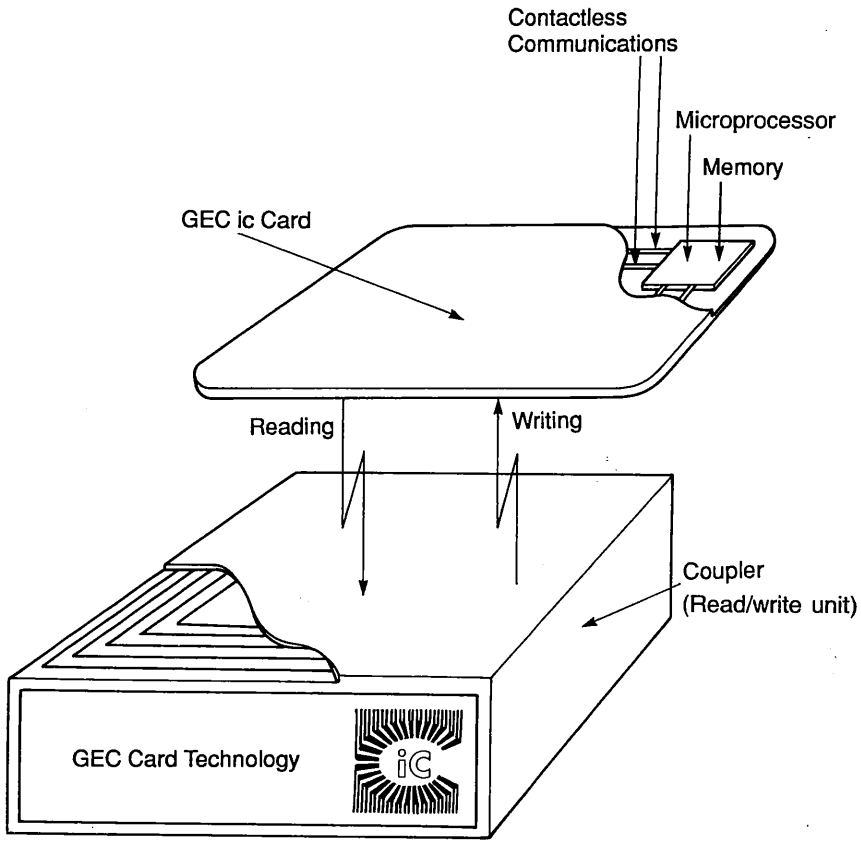

Fig. 3.1 The GEC intelligent contactless Card and Coupler (read/write) unit.

smart card which can communicate via a unique contactless interface (Figure 3.1).

3.2 THE GEC INTELLIGENT CONTACTLESS (ic) CARD

The GEC intelligent contactless (ic) Card has been designed to be a conveniently sized token that is both computationally powerful and very secure. Figure 3.2 shows the main elements of the card, and read/write unit, or coupler as it is known in its simplest form.

Sealed into the card is a microprocessor, memory and contactless interface. In practice the microprocessor and memory are on one chip.

The microprocessor is the intelligence within the card. When activated it executes instructions programmed into the memory. It can perform

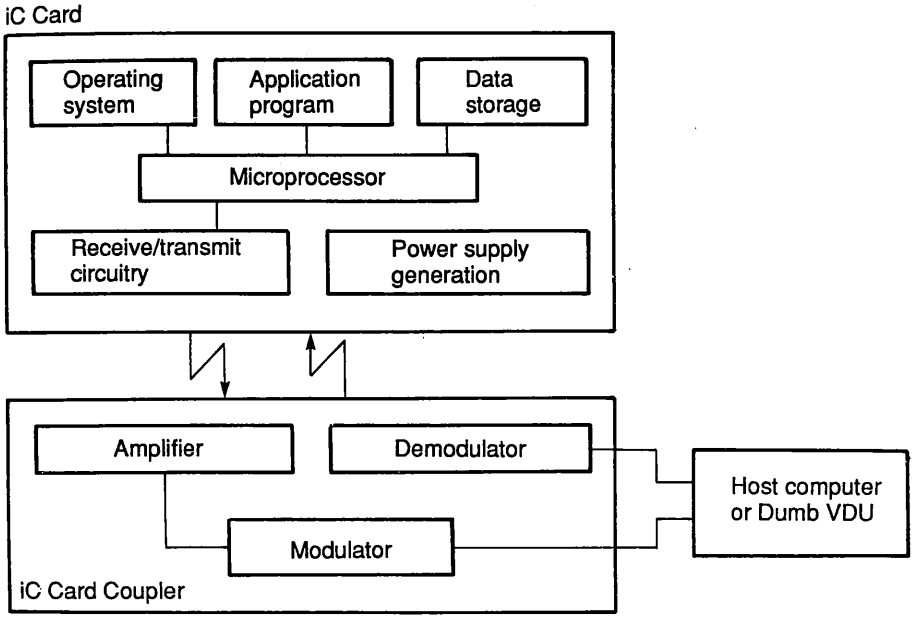

Fig. 3.2 Block diagram of the elements of the GEC ic Card and read/write unit.

calculations, manipulate data, carry out security checks to ensure that data stored in the card is not revealed to unauthorised persons, and it can scramble and descramble data sent to and from it. It can also compress data that is to be stored in the card and expand data that is to be transmitted to the outside world.

The memory in the ic Card is functionally partitioned into three areas. One area is taken up by the operating system which controls those functions such as data transmission and reading and writing of memory, which are fundamental to the card's operation. The operating system is embedded in the chip at the manufacturing stage and cannot be altered. In a second area resides the application program, that is the segment of executable instructions which defines in full the behaviour of the card within the environment of the particular application to which the card is to be put. This program is loaded by an operating system function which may be irreversibly disabled. The remainder of the memory is completely under control of the application program, and is generally used as read/write data storage.

The power for the electronics and communications link is provided by the contactless interface. Built into the ic Card is a small coil of wire which develops a voltage across it when the card is in the presence of an inductive radio frequency field. The voltage is rectified and regulated on

the card to provide a steady power supply for the electronic elements. Also built into the card is a reset circuit, designed to activate the card when it is in the correct field. The RF field is also used as a bidirectional data path. Data is sent to the card in serial fashion by shifting the frequency of the RF carrier and this is decoded in the card.

The coupler, which is linked to the host system by a standard (RS232) serial data line, acts as an interface for the data flowing between the card and host system. It produces the RF field, the frequency of which is determined by the state of data from the host, which is decoded in the card. The coupler also contains amplitude modulation (am) detection circuitry for converting information received from the card into a serial data stream. The data rate can be anything from 300 to 9600 baud with even or odd parity, if desired, and any number of stop bits. These characteristics are fully under control of the card software. The field strength from the coupler is sufficient to allow card operation up to 20 mm from the coupler but falls off rapidly after this distance. As its name suggests, the coupler merely acts as the means by which the card communicates to and from the other system components.

These technical features provide the GEC ic Card with unique benefits. The contactless interface means that the life of the card is much longer than conventional smart cards which are susceptible to contamination, damage and wear of surface contacts. In use the card can be placed in any orientation on the surface of the coupler, or the coupler can be mounted underneath any non-metallic working surface. This is very useful in banks and at retailing checkouts, keeping counters free of equipment. Unlike conventional card reading mechanisms, the couplers are sealed electronic units with no moving mechanical parts. Thus they are able to withstand harsh environments and severe treatment yet remain very reliable, require minimal maintenance and can be easily incorporated into existing equipment. As both card and coupler electronics and connections are totally sealed (Figure 3.3) they are capable of working in wet and dirty environments.

3.3 SECURITY FEATURES

Many of the applications of the smart card require security somewhere in the system whether it is for the protection of sensitive data, protection against eavesdropping or protection against illegal use. The GEC ic Card has a range of security features impossible in magnetic stripe cards and unrivalled by many computing based products. It offers:

(1) Protection against the manufacture of fake cards because of the highly complex technical nature of the card and high cost of manufac-

Fig. 3.3 The GEC ic Card and Coupler electronics are totally sealed making them ideal for use in harsh environments.

turing equipment which acts as a deterrent to all but the largest criminal organisations.

(2) Protection against easy access to the electronics, and hence data storage area, by complete encapsulation of the electronics.

(3) Protection against the probing of data lines between microprocessor and memory by incorporating both the elements on a single micro-electronics chip.

(4) Protection of the application program through the ability to 'blow' a software fuse thereby destroying the means by which the card can reload a new program.

(5) Sumcheck protection against the alteration of memory contents.

(6) Protection against altering and adding to the dialogue between the card and a terminal by authentication software specifically designed for the card.

(7) Protection against using recorded dialogue to establish authentic communication and against rerouting messages by verification software specifically designed for the card.

(8) Protection, through encryption, against deciphering dialogue between the card and terminal.

(9) Positive personal identification of the card holder by comparison of a personal characteristic (e.g. signature, fingerprint, facial features)

of the legitimate card holder stored on the card with the same feature of the person presenting the card at an access point. The comparison can be carried out within the card, thus maintaining complete secrecy of reference data.

(10) Protection, by the card invalidating itself when repeated attempts are made to gain access by continued keying in of possible personal identification numbers or forging of signatures.

In totality, the security offered by the ic Card is virtually unrivalled by any other low cost computing based product.

3.4 APPLICATIONS

Applications for the smart card can be divided broadly into three categories: data carrier, where the card is used as a convenient portable and secure means for storing data; conditional access, where the card is used as a secure means of identifying the holders entitlement to gain access to a site, a computer, a software package or a service; and financial, where the card is used to replace credit cards, cheque books or money. Each card is by no means restricted to one application only. A card can accommodate several different functions spanning all three categories. For instance, one card could be used to hold medical data, provide access to a computer system and act in a financial capacity.

As a data carrier the card has many applications in the medical field. Used as a general medical card, the ic Card could contain such information as the holder's address, date of birth, name and address of his/her doctor, allergies, recent medical history, serious complaints, drugs being taken and donor wishes. The card could be carried by the individual and in the case of an emergency, for example the holder collapsing in a street or being involved in a road accident, would provide immediate medical information to the ambulance crew (Figure 3.4) or the doctor in a hospital casualty department. The speed with which vital information would be available could well save lives. The card is also particularly suited to patients requiring regular treatment or regular monitoring e.g. diabetics, dialysis patients. In these applications the card allows key information to be provided easily and quickly to the doctor at each appointment and data can be easily added to the card.

Military applications include electronic identity tags for servicemen and women. The card can contain details of the holder, service records, medical history, entitlements etc. The card is particularly suitable for data logging. At remote or unattended sites it could be used to record temperature, events etc. Periodically it could be collected and returned to a central point for the logged information to be read off the card.

Fig. 3.4 The GEC ic Card could contain medical details about the holder. In the case of an emergency it could provide vital medical information to an ambulance crew or doctor in a hospital casualty department.

As a maintenance record, the card could be conveniently attached to equipment. The paperwork that goes with military and high value industrial equipment is often considerable. The smart card provides an easily updatable compact way of storing such data.

There are many industrial applications for the card. For example, it could be used to program computer-numerically-controlled (CNC) machines replacing punched cards or magnetic tapes. Alternatively a card could be used to store a record, for monitoring purposes, of the progress of manufactured components throughout stages of their manufacture. In the automobile industry, such a card might subsequently form the basis of a vehicle's servicing record.

In the airline field the card could be used as an electronic ticket with a complete analysis of the passenger's preference for 'smoking' or 'non-smoking' seat, as well as dietary needs. For regular travellers it could log the number of trips flown with a particular airline to give a free or reduced fare flight after a number of trips have been made.

In the area of secure access, the card can act as an electronic key to control access of personnel to facilities where sensitive work is carried out

or data is held. The most common type of security access devices are keys, badges and magnetic cards. These all suffer from the same basic drawbacks that they can be easily duplicated and when stolen or passed on to someone else, either wilfully or through coercion, they can allow entry because there is no link with the person to whom the device was issued. The ic Card overcomes these weaknesses because it is very difficult to reproduce and has the capability of storing a digitised personal characteristic of the owner (e.g. fingerprint). With suitable verification equipment, this data can be used at the point of entry to identify whether the cardholder is the legitimate owner of the card. The card also has the benefit that it can easily be individually personalised to allow access to only certain facilities depending on the security clearance of the card holder. Additionally, as the cardholder progresses through a security system, a log of the person's movements can be stored on his card as a security audit trail.

Computers often hold sensitive information and access to this information has to be controlled. The smart card offers a solution. It can hold a cryptographic key to allow access to various areas of a database depending on the card holder's level of authority.

The smart card also offers a solution to the problem of unauthorised copying of software. By storing a key part of a software program in the card, the complete program will only be able to run with the smart card present.

Direct Broadcasting by Satellite (DBS) and Cable Television are going to become more widespread in future years. The smart card offers a means for payment and the key for reception. Customers will be able to purchase an ic Card that will provide the necessary key to unscramble the picture. Cards and decoding equipment could be supplied through TV rental companies. After, for instance, an interval of one month the key required to decode the signal can be changed so that the user has to return to the rental shop to have, upon payment, the card updated with the new key. Viewing time statistics could be simultaneously collected.

Banks' major clients can use the ic Card as the key to secure access of the bank's mainframe computers for corporate cash management. The card is a secure token for individual companies to access their bank accounts and financial services from remote personal computers on their own premises. This service could later be extended into home banking.

In the general financial area the card can be used in a number of ways. It can be used to replace the cheque book. At a point of sale the smart card has the capability to compare the card holder's personal identity entered by means of a four digit number, or characteristics of a digitised signature, with a secretly held reference in the card. A correct comparison will then allow the automatic transfer of funds from the purchaser's bank account to the retailer's bank account.

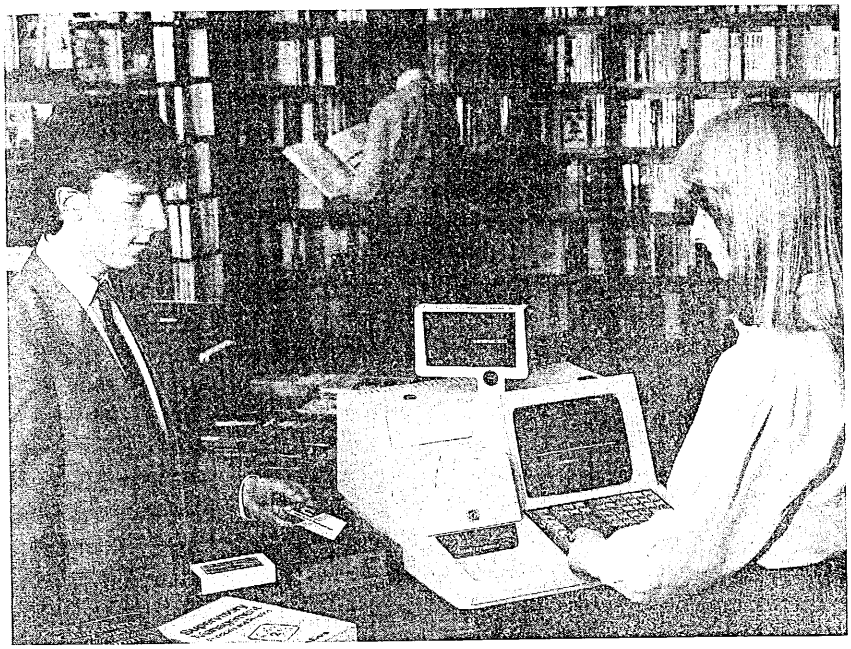

Fig. 3.5 The GEC ic Card can be used as the means for paying for goods at a retail outlet.

The card can also be used as an electronic wallet replacing cash. Here the card will have prepaid amounts which can be used for payment of low value purchases in shops (Fig. 3.5), at vending machines and car park entry points by the automatic deduction of the appropriate amount. The payment made will be held securely within the vending machine, probably on another smart card, for subsequent reconciliation.

As an electronic token, the card is equivalent to the electronic wallet but instead of cash, holds units of consumption such as electric and gas units and telephone charge units. In applications such as these the card could also provide additional facilities. In the case of the electricity/gas card it could monitor and store when units are being used; information which could be extracted from the card when next the token value is replenished. In the case of the telephone card it could also hold telephone numbers for speed dialling.

In the longer term the card could be used as a social services card carrying individuals' child allowance, pension entitlement or social security entitlement. It could be used as a driving licence, tax disc and log book, readable electronically through the car windscreen. One day it is envisaged there could even be an 'electronic' passport where the card is simply laid

he counter of immigration control to securely validate the holder pedite the immigration and visa checking process.

HE FUTURE

seems little doubt that the smart card will start to have a major in the early 1990s. It is already being extensively adopted in France. ın most of the major electronics companies are rapidly developing ards and a number of trials are underway. In the USA, potentially est world market for smart cards with a reported 825 million plastic ic stripe credit and debit cards already in existence, major trials plementations are already beginning or are expected soon. In the GEC ic Card is being used in a number of areas including the first smart cards by a UK bank for financial applications. Since its intro-ı, the GEC ic Card has attracted worldwide interest and orders een received from the USA, Europe and Australasia. It is set to major share of the emerging smart card market.

Chapter 6

Secure Transactions with an Intelligent Token

W. L. PRICE AND BERNARD J. CHORLEY

(Head, Data Security Group, National Physical Laboratory)

The intelligent token is a type of supersmart card providing security for electronic banking and commerce.

6.1 INTRODUCTION

The NPL intelligent token was conceived and developed as part of a research programme, beginning in 1982 and continuing for eight years, sponsored by the Tokens and Transactions Control Consortium (TTCC); the latter was set up by the British Technology Group and the UK Department of Trade and Industry. The research programme was carried out at the UK National Physical Laboratory and the members of the consortium included organisations from amongst the suppliers and users of access control and data security technology. Descriptions of the NPL intelligent token can be found in conference papers by Price and Chorley [1,2]. Studies of a similar concept have been carried out by the OSIS organisation (now TeleTrusT) [3], though, as far we know, this system has been demonstrated only by simulation

We give here an account of the general design philosophy of the NPL token, together with some details of the design itself and a summary account of the potential applications.

In the context of access control (to computers and, indeed, many other systems) there is usually a need for a reliable means of personal identity verification. In many computer-based transaction processing systems there is almost always a further need to ensure the integrity of transaction messages initiated by or on behalf of the system user. The design philosophy of the NPL token aims to meet these two needs in the one device.

It is becoming increasingly common for access to computer systems to depend on knowledge of some password, often coupled with possession of a token; a very common experience is use of a token in the form of a bank plastic card, associated with a password in the form of a personal identification number or PIN. Unfortunately the security of the plastic

card, with magnetic stripe, leaves a great deal to be desired; it is far too easily copied, though anti-forgery features are added by the manufacturers. The known insecurity of the magnetic stripe card is one of the reasons behind the development of the 'smart card', which is the subject of much of the present volume. The smart card has the advantages of far greater storage capacity, coupled, in some cases, with internal processing power. Smart cards are used in many applications including payment for services, payment for goods and operation of personal or corporate bank accounts. A consensus view holds that smart cards are far less easily copied or forged than magnetic stripe cards.

Of course, if a smart card designed to be used without password is lost, then anybody finding it can use it; this is the case with 'smart' cards designed for telephone prepayment applications. Therefore, in applications requiring higher levels of security, such as in accessing bank accounts, a PIN is usually associated with the card. Correct response from the card depends on correct presentation of the PIN to the smart card terminal.

It is well known that the security of systems which depend on PINs leaves a lot to be desired. Habits of people with PINs can be very bad, such as writing them on the cards or sharing them with other people; it is not uncommon to wait at a bank ATM whilst the person in front presents a number of cards on behalf of colleagues, with the relevant PIN for each card. The alternative to the PIN is personal identity verification by one of several possible biometric techniques. Biometric methods measure the way in which the card holder carries out a specified task, such as signing one's name, or else measure a physical characteristic of the card holder, such as fingerprint pattern. The aim of the biometric system is to make impersonation extremely difficult. Several problems hinder the introduction of biometric methods; these include variability of results, relatively slow response in some cases, unpopularity with the users, etc. Until more acceptable biometric methods are available, we fall back on the PIN, which is at least better than using a card with no password.

Theft of PINs by thieves must be prevented at all costs. Such theft can take place in many ways, including bugging a transaction terminal to collect PINs. Whilst the risk of a bank automatic teller machine being bugged for this or any other purpose is very small, this cannot be said for the kind of terminal which we are beginning to find on retail counters in shops. Cost considerations dictate that the degree of tamper resistance that can be built into retail terminals is low. PINs presented on low security retail terminals may therefore be subject to disclosure.

If a PIN associated with a magnetic stripe card is stolen, then the thief may be able to create a false card and thus deceive the access control system. If a PIN associated with a smart card is stolen, then it is likely that the thief can only profit by his action if the card is also stolen. This

must be regarded as a real possibility and therefore there may be an advantage in a system which gives better protection to the PIN.

6.2 DESIGN PRINCIPLES OF THE TOKEN

One way in which the PIN can be given better protection is to provide a keyboard for its entry which is under direct user control. Thus the PIN is not entered on the keyboard of the retail terminal, but upon a keyboard specially mounted on the token itself. The first design principle of the NPL intelligent token is therefore that the PIN should be presented to the system on a keyboard integral with the token. We shall later describe how the token is able to check the PIN validity and then satisfy the terminal that the PIN was valid without actually disclosing the PIN to the terminal. Clearly it is not sufficient for the token to receive the PIN on its keyboard, check it and then send a message saying 'the PIN was correct' to the terminal, since this message could be sent by a false token with an incorrect PIN. The message must be such that the terminal can rely upon it.

In a transaction processing system the user is usually dependent on displayed information on the terminal; the question is whether the user can always trust this information. It is a common experience to motorists to buy petrol from a pump with digital light emitting diode display; elements of these displays frequently fail and an amount (petrol or value) is displayed which is different from the correct value. It is not inconceivable that a retail terminal display might be deliberately altered by someone seeking to defraud the system owners or users. Again this brings us back to the customer's personal token. If it were possible to display the transaction details, particularly the amount of money to be committed, on a display under customer control, then this problem would be much reduced, at least from the customer's point of view; comparison of token and terminal displays may give added confidence. For this reason the NPL intelligent token carries its own display for communication with the customer, which is the second important design principle of the token.

We have indicated earlier the importance that must attach to the integrity of transaction messages which control the movement of funds between user and retailer accounts. It is frequent practice to protect transaction messages by encipherment techniques, often by authentication based on symmetric encipherment algorithms. An alternative and attractive method of ensuring integrity is based on the digital signature derived from an application of public key cryptography. As we shall see, the NPL token is able to satisfy a terminal that a correct PIN has been offered without disclosure of that PIN; this property depends upon the application

of a digital signature. The basic token design therefore includes the ability to calculate digital signatures using a stored secret key. Since the ability to calculate signatures is a fundamental requirement in the device, it is convenient to apply this ability to calculating signatures on transaction messages authorised by the token holder. Transaction messages signed in this way can be checked anywhere in the transaction processing system where a reliable copy of the corresponding public key is available.

6.3 REALISATION OF THE TOKEN DESIGN PRINCIPLES

We have now identified the three fundamental design principles of the NPL intelligent token – integral keyboard, integral display and ability to calculate digital signatures. We proceed to discuss the ways in which these design principles have come to be implemented in physical and logical terms. The central part of the design is an implementation of the RSA public key cryptosystem [4]; this software implementation runs on a fast signal processing chip, the Texas Instruments TMS32010.

Personal identity verification (more strictly PIN verification) begins with presentation of the token to a terminal by the user; the terminal senses the presence of the token and generates a random number which it sends as a challenge to the token; at the same time the token signals to the user (using its own display) that the PIN must be input on the token keyboard. The token is designed to check the PIN and, if the PIN is correct, to sign the random number just received from the terminal using, for this purpose, the secret RSA key contained within the token. (Should the PIN be incorrect, the signature process does not take place and the user is given a limited number of attempts to get the PIN correct, failing which the token is disabled.) Having produced a transformation of the random number by the signature process, the token returns the transformed number to the terminal. The terminal, after having generated the initial random number challenge and while the token is preparing its signed reply, will have sought the public key corresponding to the token; this can come either from a reference source of public keys or can be supplied by the token itself in the first exchange of data with the terminal, in which case the version of the public key is supplied already signed by the secret key of a superior authority (the public key of the authority must then be known to the terminal). Given the public key corresponding to the token, the terminal can check the validity of the returned signature of the random number challenge; correct signature implies correct PIN presentation on the token keyboard. Figure 6.1 illustrates the sequence of events in identity verification.

Because the token is capable of generating RSA signatures, it is a simple extension of its functionality to permit the signature of transaction messages. In a retail point-of-sale system these messages would be pre-

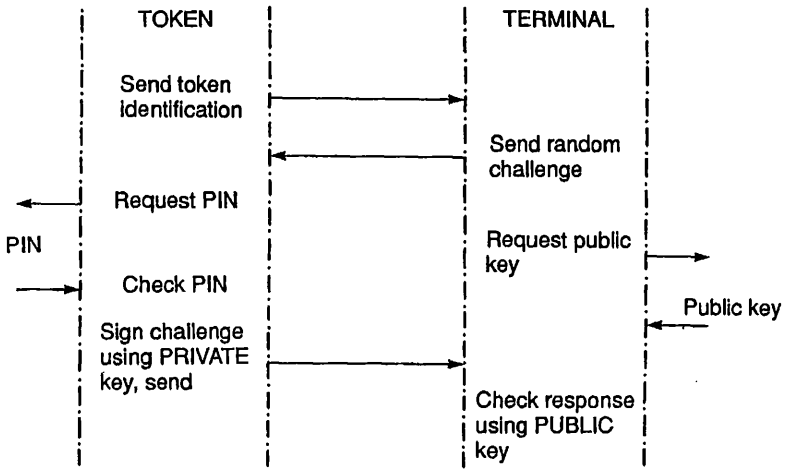

Fig. 6.1 PIN checking, token challenge and response.

pared on the retailer terminal and sent to the token for approval by the token holder (inspection in the token window by the user) and, if approved, signed by the token and returned to the terminal. The terminal can check the validity of the signature and then send the signed message to a transaction processing centre. The signature validity can be checked by any entity in the system having access to the public key corresponding to the signature token. To avoid replays of transactions, it is necessary to include a time and date field in the message. Transaction numbering does not lend itself conveniently to prevention of replay for token originated transactions; tokens accessing multiple services would require a serial number for each and hosts offering services would need a number for each token in valid issue. Figure 6.2 illustrates the sequence of events in transaction signature.

It is an interesting extension of the design that the initial random number challenge may be omitted and replaced by the transaction message. In this case the protocol is shortened by arranging that the identity verification is checked by the signature on the transaction message.

The ability of the token to sign messages can be extended to cover messages in general; the application of the token is not restricted to value transactions.

6.4 THE PROTOTYPE TOKEN

The NPL intelligent token was created in prototype form in a unit measuring 36 cm × 15 cm × 2 cm; this device contained 21 discrete

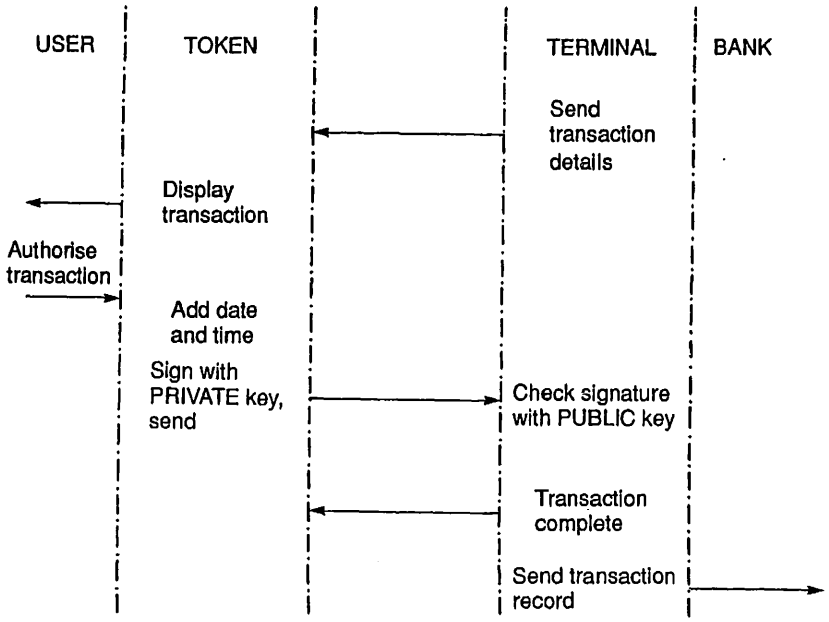

Fig. 6.2 Transaction authorisation and signature.

integrated circuits, including the TMS32010 and an Intel 8085 to act as controller. Battery maintained RAM was provided for storage of parameters such as keys and a record of transactions.

Consideration was given to a semi-custom designed RSA processor chip in the early days of the project. It was considered at that time that the technology was not readily available for such a device and so an alternative method of implementing the algorithm was sought. Texas Instruments had just released the first in a series of fast processor chips designed mainly for signal processing applications. This device, the TMS32010, is a 16-bit microprocessor with an instruction execution time of 200 ns. The instruction set includes a signed 16-bit multiply executed in one cycle. Although not ideal (overflow and the sign bit caused problems) this processor was programmed to perform the RSA calculations. The new design continues to use this processor.

The TMS32010 has limited program and data memory spaces of 4K words and 144 words respectively. Further, there is no high-level language compiler and the speed of the processor makes it difficult to interface to slow peripheral chips. For these reasons, all the remaining functions of the token were placed under the control of a second processor. This split has several advantages; the two processors can work in parallel, thus

reducing or even hiding the RSA calculation time, a high-level language can be used for the application software, none of the TMS32010 memory resources are wasted, and peripheral interfacing is easier and uses fewer components.

Creation of the prototype enabled the development team to demonstrate the correct functioning of the device in applications such as access control, point of sale transactions and signature of alphanumeric messages. Since the prototype was comparatively large, the next stage was to engineer a smaller version. In order to reduce the size, the functions of a number of the separate integrated circuits were absorbed into one application specific integrated circuit (ASIC). The chip count was thereby reduced to 11; further space was saved by using surface mounting technology. The result, produced in collaboration with Texas Instruments, was a device similar in size to a medium sized pocket calculator (about 14 cm × 9 cm × 1 cm).

In the new version, the Intel 8085 is replaced by a member of the TMS7000 series of 8-bit microprocessors. As before, this processor also controls the RSA processor and all peripherals, maintains the secret data and runs the application program. 32 K bytes of program memory are available, most applications to date have used only half of this amount. 8K bytes of battery backed RAM are built in for the storage of keys and other data needed for applications.

All of the decoding logic, address latching and bus de-multiplexing for both processors have been reduced to a single semicustom chip designed at the NPL and fabricated in 1.8 micron CMOS gate array technology by Texas Instruments. Figure 6.3 is a block diagram showing the important physical features of the token. The display is a 16 character by 2 line liquid crystal display and the keypad consists of 4 rows of 3 buttons. The reasons for including these on the token are given elsewhere in this chapter. The clock maintains information about the date and time of day, this is used in some applications to date stamp messages. Communication between the token and the outside world is by way of a three-wire serial interface.

Communication between the two processors takes place over an 8-bit bidirectional bus buffer constructed in the semicustom chip. A data block containing the message, exponent and modulus required for an RSA calculation is sent to the RSA processor via this interface, the result is returned in a similar way. Block transfers are completed in a few microseconds, a time not considered significant when set against the calculation time.

The token contains secret information unique only to itself. No other token contains the same secret so the compromise of one does not put the security of the whole system at risk. Nevertheless, measures must be taken to detect tampering and destroy the secret information upon detection. Possession of the secret key would enable an intruder to falsify

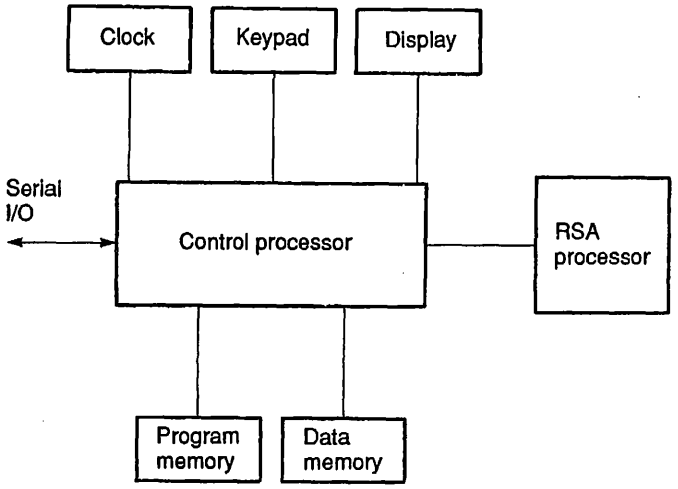

Fig. 6.3 NPL token block diagram.

transactions in which the token was engaged. The tamper resistance built into the prototype token simply detects when the case is opened and destroys the secret data by discharging the RAM chips. Although adequate for our prototypes, this form of tamper detection is nothing more than a gesture and new circuitry has been designed to perform these functions properly. We have indicated elsewhere the important functions performed by the display and keyboard; clearly these devices must also be included in the tamper protected area. It is necessary to note that tamper resistance will add significantly to the cost and size of the token.

The speed required of the RSA calculation implies a device that consumes a lot of power. It is unlikely that this can be supplied conveniently to the token without direct electrical contact so the token takes its power from the terminal into which it is plugged. RSA calculations are only required when the token is communicating with a terminal, so it could be supplied separately; other functions such as a calculator, clock and a database would then be run by the TMS7000 under battery operation.

The implementation of the RSA algorithm on the signal processor is capable of carrying out one full operation of the algorithm in about 3 seconds. This implementation has been described by Clayden [5]. An increase of speed has been achieved by using the method [6] based on knowledge of the two primes that make up the public modulus. Knowledge of these primes is permissible in the device that holds the secret key. In other words the two primes method can be used to generate signatures, but not for checking them. By this means it has been possible to reduce the algorithm time for signature generation to about 1.4 seconds. For

signature checking it is possible to gain a substantial increase in speed by limiting the public exponent to, say, 17 bits. This places no restriction on the size of the secret exponent and does not reduce the security of the signature operation. The speed of signature checking is not exactly inversely proportional to the size of exponent because of overheads in the computation, but a very substantial reduction in checking time is possible. In our implementation, signature checking using this method takes 60 ms.

6.5 MINIATURISATION

It is possible to argue that a token carrying display and keyboard should not be reduced in size to that of the standard plastic card as has happened with the smart cards. Very many people now carry small pocket calculators as a matter of course and the intelligent token could easily be reduced to a compatible size. Smart cards have not yet been in wide enough use to allow a judgement to be formed as to their durability. If the ISO draft standards for physical properties of smart cards are to be regarded as firm guidance, then cards are going to need the ability to withstand very severe handling. A small calculator-like object can be made much stronger. We understand, of course, the desire to make the smart card compatible in dimensions with the existing magnetic stripe card standards because of the volume of installed equipment accepting the latter.

Of the 11 chips in the current token, 6 are memory devices and would be the first target for chip count reduction. Integrating these devices into the processor would also increase the security of the token. A low power processor could be permanently powered by battery; under these circumstances it could also act as the time of day clock. A two chip token could be built today using semicustom microcomputers now available. A single chip design using the latest digital signal processors is a distinct possibility in the near future.

6.6 BIOMETRICS

It should be plain from the introduction to this chapter that we are not satisfied with the level of security achievable in systems designed around personal identification numbers. We have mentioned the possibility of replacing PINs by biometric methods. Because the NPL token is the size of a small calculator, it is a realistic aim to mount a suitable interface for biometric identity verification on the case of the token; this could be based on written signature, fingerprint, vein scanning, etc. It is convenient that the type of processor required for realising the biometric measurement is just the same type that has been used for the RSA implementation in

the NPL token, a signal processor, such as the TMS32010. If a biometric identity verification method is to be implemented in the intelligent token, further work is required and this depends on achievement of acceptable performance in the chosen biometric identification method. It is unrealistic to expect that any biometric method will be totally error free and the overall transaction system design will need to take this into account. When a PIN is presented to an identity verification system there is no margin for error – the PIN is either correct or it is not; in a biometric system a degree of tolerance must be built in to allow for variability in measurement. Because of this degree of tolerance, a higher level of security may be achieved by retaining the PIN as a supplementary check on identity.

One method currently of interest relies upon the unique features in the blood vessel patterns on the back of the hand. This development, called Veincheck, is under investigation on behalf of the British Technology Group. If built into the token it might consist of a row of four infrared emitter/detectors built into the base. Identification would be performed by wiping the base of the token over the back of the owner's hand before confirming a transaction.

6.7 FUTURE DEVELOPMENTS

Clearly, given demand for the device, it will be feasible to miniaturise the intelligent token still further, possibly even down to the compass of an ISO standard plastic card, yet carrying keyboard and display. Products of this type are already being designed by various manufacturers, but not making use of the signature capability of the NPL device. A smaller device would undoubtedly be more convenient and, therefore, possibly be more acceptable to users. Adoption of ISO card dimensions would ensure compatibility with magnetic stripe card readers adapted to take intelligent tokens. However, a very small size is not necessarily an advantage. The durability of a very small design of this nature is not yet tested. User habits with plastic cards are notoriously bad and a very small token would have to be resistant to considerable mishandling.

Even more critical than the size of the device is its cost. For mass penetration of a market low cost is essential, especially if the competition is provided by cheap magnetic stripe plastic cards. If the cost of intelligent tokens remains high, then their use may be limited to applications where cost is not so important.

If it becomes possible to achieve a sufficiently low unit cost, then mass markets may become available and the token used for transaction control and payment for a wide range of services in many environments, such as in the home, in public utilities (transport, telephone, etc.), the office and

in shopping. If the amount of internal storage can be increased significantly, the token applications can extend to serving as a personal memo, communicating with a database held on a personal computer. It is also possible to envisage applications in the medical field, with personal medical records held within the token, and prescriptions issued by the doctor's terminal directly to the token and the pharmacist's terminal reading them securely. Because many people are quite happy to carry small calculators on their person, the addition of calculator functions to the token might be an attractive option. Addition of such a capability might overcome the problem of token cost. People willing to buy a pocket calculator might be only too willing to pay a little extra for the substantial addition in capability.

In this chapter we have attempted to point out the need for an identity token with greater security than the magnetic stripe card. We are also concerned that terminal security in smart card systems may be a weakness. The intelligent token has the advantage of greater safeguard for PINs, whilst offering the further capability of ensuring the integrity of transaction messages. The cost of providing terminal security can be reduced by introduction of the intelligent token, because the terminal need not contain protected secret parameters.

REFERENCES

[1] Chorley, B. J. and Price, W. L. (1986) 'An intelligent token for secure transactions' *Proc. IFIP/Sec'86*, Monte Carlo, December 1986, 442–450.

[2] Price, W. L. and Chorley, B. J. 'The intelligent token or 'super-smart' card' *Proc. SmartCard 2000*, Vienna, October 1987 (proceedings to appear).

[3] Commission of the European Communities (1986) Open Shops for Information Services (OSIS), Final Report, Cost Project 11 ter, CEC, June 1986.

[4] Rivest, R. L., Shamir, A. and Adleman, L. (1978) 'A method of obtaining digital signatures and public key cryptosystems'. *Comm. ACM.*, **21** (2) 120–126.

[5] Clayden, D. O. (1985) 'Some methods of calculating the RSA exponential' *Proc. Int. Conf. on System Security*, Online, London, October 1985, 173–183.

[6] Quisquater, J.-J. and Couvreur, C. (1982) 'Fast decipherment algorithm for RSA public-key cryptosystem' *Electronic Letters*, **18** (21) 905–907.

Chapter 8

Cryptography and the Smart Card

D. W. DAVIES

(Data Security Consultant)

Cryptology is the key technology for secure systems.

8.1 Introduction

The close relationship between smart card and cryptographic techniques can be looked at from two directions. The smart card can be used as a component of a cryptographic system to improve its convenience or level of security. From the other viewpoint, the smart card itself is the main component of the system and cryptography is called upon to help it with its task. In this chapter we shall mainly adopt the second viewpoint, which is centred on smart card applications but first let us look at the smart card as an adjunct to cryptography.

The confidentiality of data on a communication line can be protected by enciphering it. Encipherment is a transformation which makes the transformed data seem meaningless to an outsider, yet which allows an inverse transformation, for those authorised to receive the information, which turns it back into its clear text. To separate the authorised readers from others, the authorised readers hold a secret value called a *cryptographic key* without which decipherment is impossible. In the usual form of cryptography, this secret key is used as a parameter for both the encipherment and the decipherment functions.

When cryptography is used to protect data travelling some distance, before it can go into operation a secret key must be established at both the sending and the receiving end. Conveying the key from one place to the other entails a risk of losing it to an opponent. A smart card can be used to store a key for secure transport. The use of this key can be authorised by means of a password, known only to authorised users, and the smart card itself can take part in the complex process of key management. Some of the techniques are described later in the chapter.

Sometimes, cryptography is used to encipher information not for communications purposes but to protect it while it is stored locally. It might be difficult to protect the local store from illicit access or information

stored on a removable medium might be stolen or copied. When cryptography is used for stored data, the keys are not transported but their security is very important because they can unlock all the protection provided. Most computers are physically insecure, so a smart card can be used to hold the keys and the card taken away by its owner and stored in a safe place. Here also, a password can be used to unlock the secret key from the card.

A related problem of cryptography is the protection of the cryptographic mechanism itself. Not only must the key be protected but also the place where the cryptographic transformations take place. Smart cards can help in this problem by becoming, themselves, the 'cryptographic engine' of the system. If they have enough processing power for the purpose, they can hold all the protective mechanism of a secure system, particularly at the terminal end where the processing demands are less severe. The computer itself, which might be an intelligent terminal, is physically insecure and any part of its store or process is open to tapping or 'bugging'. To counter this, cryptographic methods are used and the keys, together with the cryptographic transformations, are contained entirely in smart cards. When these are removed, the system is locked up and the information it contains is safe against illegal access.

These are examples of the close relationship between smart cards and cryptography, seen from the side of the cryptographer who regards the smart card as an additional tool. Our viewpoint in the rest of this chapter is to think of the smart card as a main component of the system and see how cryptographic techniques are used for its purpose.

8.2 PROTECTION FROM PASSIVE AND ACTIVE ATTACKS

Cryptographic techniques can be used in a large number of ways and for many different purposes. The basic purpose is to protect a system against misuse by impostors or unauthorised people. The first stage in protecting a system is to analyse the threats to the system and the risks they entail. We shall consider only those threats that are amenable to cryptographic protection and, as a first step we divide these into *passive* and *active* attacks.

A passive attack attempts to read information without changing it. Examples are the tapping of a telephone line, stealing or copying a diskette, observing a password by looking at the keyboard while it is entered or picking up stray electromagnetic radiation from which a meaningful signal can be reconstructed. Generally speaking, these attacks are not difficult to carry out and in a widespread communication network it is impossible to prevent them. The tapping or bugging of voice conversations is a highly developed art which can be applied (with a few

changes) to the collecting of digital data from communication lines, I/O channels, processors, stores, keyboards or any other part of the information system. In these circumstances, to preserve the confidentiality of information it must be transformed by encipherment and then transformed back into clear form for processing, printing or display. Wherever the information is in clear form there must be other ways to protect it, such as not allowing unauthorised persons into offices where information is displayed or printed.

On the other hand, an active attack is one which seeks to alter the information, perhaps to falsify a transaction, prevent a debit to a bank account from reaching it or even destroy an entire file. Generally speaking, these active attacks are much more difficult to carry out and they require more skill and sometimes more luck to achieve their purpose. When they do succeed, the consequences are often more serious than those of a passive attack.

It might seem that encipherment was all that is needed to prevent an active attack on data but this is not so. Some of the more extreme active attacks such as placing a bomb in the computer room are not amenable to cryptographic protection but, these aside, there is a much wider range of possibilities against which protection is needed. Suppose, for example, that an important file is stored in enciphered form and updated from time to time, while still being enciphered. If we are not careful, an attacker simply replaces today's file by an older one which will pass as genuine because it is properly enciphered. Whole transactions might be knocked out on a communications line without detection. Some encipherment methods allow an attacker to change individual bits of a message or file without knowing the cryptographic key. Thus there is more to protection against an active attack than encipherment alone and the methods used have in the past been known as 'authentication' because they ensure that the information remains authentic. This term is discouraged by the recent usage of the International Standards Organisation and we speak therefore of *data integrity* meaning that the data takes the values it was intended, not those altered or substituted by an impostor.

In an extended communication system it may not be possible to prevent the changing of information but we can at least hope to detect when the information loses its integrity. Thus we shall normally be speaking about integrity verification rather than the prevention of an active attack.

One form of active attack is the *masquerade* which means an unauthorised user pretending to the system that he is an authorised one. It goes further than this because any part of the system can be subject to a masquerade. A well known example is the program which asks the user to present his password. If the program is an impostor, the passwords can be given away to an enemy. Protection against masquerading should therefore be included in every interaction within the system where it is physically

possible for an intruder to slip in. In the work of the International Standards Organisation this is known as *peer entity authentication* since the two communicating parts of the system are entities at the same level, i.e. *peer entities*. The particular case which we meet most frequently is the authentication of a user to the system. Masquerading as an authorised user is the commonest type of active attack on teleprocessing systems, typified by the 'hacker' who mainly has to contend with rather simple password systems. Authentication of users can take much more secure forms, using smart cards, which will be described later in section 8.5. As in the present use of 'automatic teller machines' (cash dispensers), a token which the user holds together with a password which he remembers adds a lot to the security. A smart card, which is very difficult to forge, makes an ideal token for this purpose.

Masquerading forms a bridge between passive and active attacks. By tapping the line to find the password or guessing the password or trying a lot of passwords the attacker obtains the means to enter the system in the guise of a genuine user. This gives him the possibility of both a passive and an active attack, the active attack requiring no more skill than normal use of the terminal, if a public network is used for access.

In the next three sections of this chapter we shall describe the use of cryptography against passive attacks, preservation of data integrity and user authentication.

8.3 CRYPTOGRAPHY

Cryptography is an ancient art which, like many others, is changing rapidly with the development of information processing technology. We are concerned only with ciphers, which use a cryptographic algorithm and not with codes which are based on large, arbitrary code books. Figure 8.1 shows the nature of a cipher which employs two algorithms E and D to encipher and decipher information respectively. Encipherment operates on the plaintext using a cryptographic key (k) as parameter. Decipherment

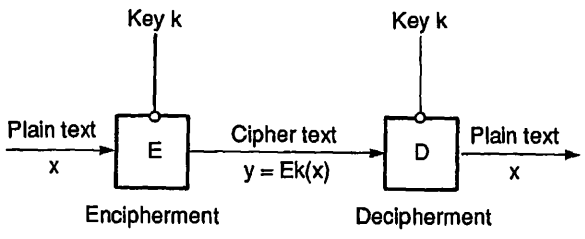

Fig. 8.1 Notation for encipherment and decipherment.

uses the same key to operate on the ciphertext y and restore the plaintext x. The symbols x and y can either represent single blocks of data on which the transformation takes place or streams of data, for example, a stream of ASCII characters passing over a communication link. This distinction corresponds to two different types of cipher algorithm, a *block* cipher and a *stream* cipher. Later we shall give examples of both.

8.3.1 Attacks on a cipher

The cipher derives its strength, in part, from the key. If the secrecy depended on the algorithm alone, when this algorithm became known to an enemy, either by chance or by hard cryptographic work, it would be necessary to go back to the drawing board and design a new algorithm. All modern ciphers make use of a key so that, even when the algorithm is known, decipherment is not possible unless the particular key employed is known.

The possible number of different keys is important, like the number of 'differs' of the keys you use in a lock. If this is too small, a processor can be set to try all the keys and see which one generates a plausible plaintext. The total number of different key values is known as the *size of the key space*.

If nothing were known about the plaintext and it could therefore be any random set of digits, the most elaborate and fastest machine for searching through the keys would still tell us nothing. We must either have some knowledge of redundancy in the plaintext or, better still, some examples of actual plaintext with their ciphertext equivalents. In many cases, probable words or phrases from the plaintext can be guessed when their place in the text is unknown. With any of these clues about the plaintext an attack can be mounted, by searching through all the keys, in principle. A very large keyspace prevents the breaking of a cipher by simple key searching but it does not ensure that other, cleverer methods will not succeed. Only long experience with the practical breaking of ciphers can enable the strength of a cipher to be evaluated. Ultimately, there can be no guarantee because the whole range of possible methods of attack can never be enumerated.

8.3.2 The Data Encryption Standard

The US Government employs 'Federal Information Processing Standards' for its own use. In 1973, the US National Bureau of Standards announced that it was contemplating a standard for data encryption and asked for proposals. The response was disappointing but after a second call in 1974

a proposal from IBM was seen to have promise and, after some changes, it was published in 1975 as a draft. Eventually this was adopted in 1976 as the *Data Encryption Standard* defined by FIPS Publication 46.

The Data Encryption Standard or DES had a much wider influence than was originally intended. The US standards body ANSI made it a US standard and it was widely adopted, particularly in banking and financial services. Because of restrictions placed on the export of chips for the DES algorithms from the USA, its use outside the USA has largely been confined to banking and financial services. It became a *de facto* standard within this community.

The DES is a block cipher with plaintext and ciphertext blocks of 64 bits and it employs a key of 64 bits, but has 8 parity bits so that its keyspace depends only on 56 bits.

The structure of the algorithm is described in detail in reference [1]. It uses a combination of bit permutations and substitutions. By *substitution* we mean that an input field is used as an address to look up a table and produce an output field. The DES employs eight substitutions, each with a 6 bit input and 4 bit output. These are known as the 'S boxes'.

From the outset, the strength of the DES has been controversial. The size of its keyspace was criticised on the basis that a complete search through its keyspace could be carried out in less than one day by a machine containing one million devices, each able to test one million keys per second. The cost of such a machine was argued about but believed at the time to be in the region of $10M. The structure of the S boxes was clearly not random but the criteria for choosing these tables has never been published and the work done to develop the DES is classified. Many felt that this was unsatisfactory for a standard.

The whole idea of a published algorithm was a new one. Though the secrecy of the algorithm is not assumed to be essential to the security of a cipher, no cipher had been published, studied and discussed so widely and in such detail.

Studies in many places revealed interesting properties of the DES, including the existence of four *weak keys* and twelve *semi-weak keys* which give the cipher special properties. Interesting facets of the design of the DES continue to be discovered. However, none of these has shown any significant weakness of the algorithm beyond that which is implied by the size of its keyspace.

As information technology improved, the cost of a key searching machine decreased until, by about 1985 it became obvious that a searching machine would be within the bounds of expenditure by a large corporation or a large criminal organisation, assuming that the searching time was allowed to extend to about one month. This is a real threat when the same key is used for longer than this time. Multiple encryption using the DES with at least two different keys has been adopted by the financial

community to overcome this problem. This is particularly relevant in key management for a large system where a master key remains in place for some time.

By 1988 the US had decided not to renew their endorsement of the DES for federal purposes. When this was announced in 1986 it caused consternation among the financial users but the US Government stated that it would not discourage use of the DES for financial transaction purposes.

At the present time, adoption of the DES or any other cipher, as an international standard has been stopped by a resolution by the ISO Council so there is no role for the international standards body in the development of a replacement. This leaves the financial community in an unsatisfactory position, particularly where the use of the DES for safeguarding large payments is concerned. For small payments, with a good key management system, the DES is usually considered adequate.

The algorithm was designed for implementation solely in hardware on a special chip, according to the US Federal standard. The ANSI standard removed this condition and many systems now employ the DES implemented in software on a microprocessor. Because of its origins and the original intention to use hardware, the DES is not a convenient algorithm for software implementation and was outside the ability of the early smart cards. Recent developments in smart cards have made the DES more feasible so that it is now possible to implement the DES in software within the same smart cards. Where this is the algorithm required by financial transactions, incorporation of the cipher in a smart card is a useful contribution to security, enabling the card to store its own secret key and use it internally.

A replacement for the DES will probably be a block cipher with the same size of plaintext and ciphertext blocks in order to ease the change from DES to the new cipher. The key size would have to be larger, at least 64 bits. A replacement for DES should be designed to be easy to implement in hardware and software.

8.3.3. Methods of using a block cipher

There is a limited number of applications for a cipher which can only handle 64 bits. A longer message could be broken into blocks of 64 bits, for example blocks of eight ASCII characters, and each block in turn could be enciphered. This simple method of enciphering a longer stream is not recommended for two reasons. Firstly, an enemy could extract pieces of the message made of a number of blocks and reuse them or rearrange them in messages of his own construction. Secondly, though eight characters may have a large number of potential values, in some contexts there

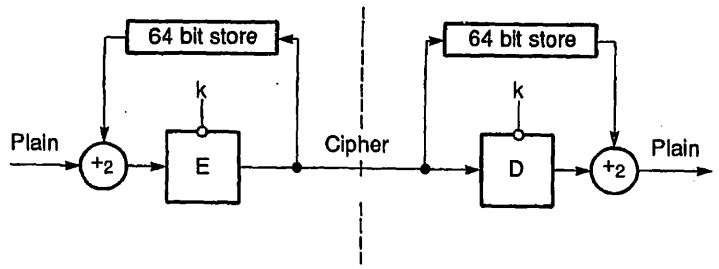

Fig. 8.2 Cipher block chaining.

would be blocks which repeated quite frequently and could be recognised in ciphertext, so that methods used for breaking codebooks could be applied. For example, in English text the words 'of the', with the three spaces, when it occurs on the eight character boundary will probably be the most common block and can thus be recognised. In computer output, sequences such as all zeros or all spaces will easily be detected.

To overcome these problems a method of using the DES was introduced known as cipher block chaining (CBC) and this can be applied to any block cipher. It is illustrated in Figure 8.2 which shows the ciphertext produced by the sender added, modulo 2, to the next plaintext block. This chains each block to its neighbour, preventing the codebook analysis method and the separation and reassembly of blocks of text. When applying this method, the block which is modulo 2 added to the first plaintext block is called the *initialising variable* or IV. In many systems, a new IV is sent when each new key is distributed and the IV is kept as secret as the key.

Decipherment presents no problems, since the ciphertext can be stored and modulo 2 added to the plaintext of the next block. There is no synchronising problem in the sense that, whatever happens to modify or momentarily interrupt the ciphertext, provided the boundaries of the 64 bit block can be recognised at the receiver, the system will recover. For example, a single error in ciphertext results in completely random output for the plaintext block into which it is transformed. The same error goes into the 64 bit store and emerges in the following block, so two blocks of text are affected. After this, the error does not propagate any more. Even this amount of error extension can be troublesome. It can interact badly with error correction schemes and it greatly increases the average error rate on the line.

The CBC mode of operation is most useful where the information already has some imposed block structure, as in formatted messages. When these messages have been assembled in a store, the CBC encipherment can be applied to the whole stored message and then the result transmitted or stored, as required.

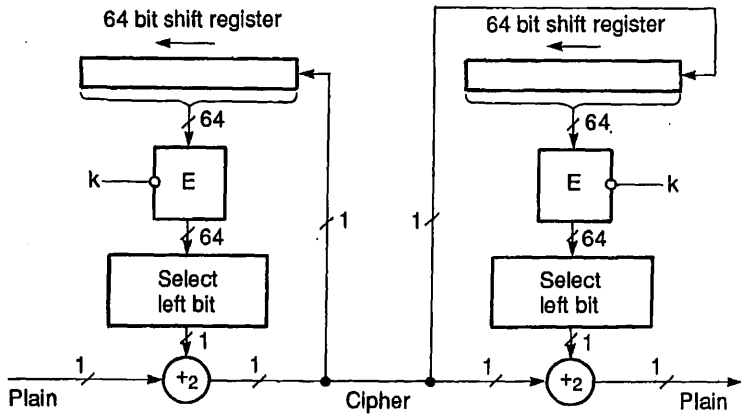

Fig. 8.3 One-bit cipher feedback.

This mode of operation is not convenient when the natural unit of communication is smaller than a block. For example, much communication takes place in the form of 8 bit units or less, for example 8 bit ASCII characters or 5 bit telegraph codes. For these, the use of a 64 bit block cipher would mean that characters were held up, waiting for transmission, until the block was filled. Interactive communication between two operators would require that, after a short pause, the residual characters were padded out and transmitted. To satisfy these requirements, another method of using a block cipher was introduced which is *cipher feedback* (CFB). This is illustrated in Figure 8.3. At the top of the figure is a 64 bit shift register into which the ciphertext is introduced, bit by bit, from the right hand end. Thus it contains, at each end of the communication line, a record of the last 64 bits of the ciphertext. At each end of the line, this block is enciphered and from the result only the left hand bit is extracted. This bit stream is added into the plaintext to produce the ciphertext. Effectively, it is a random bit stream generated by feeding back the ciphertext itself. Note that the cipher algorithm is in this case used only in the encryption mode; in fact a reversible algorithm is not really needed.

Because this works on each bit as it arrives, it is completely transparent to the procedure used by the communicating parties, whether they are people or machines. It is normally used within the OSI architecture at the physical layer because it treats only the bit stream, without reference to its structure, and passes on each bit as it arrives.

The entire DES algorithm calculation must be made again for each bit, so if there is a limited speed of processing, this method will be slow.

As with the CBC mode, the shift register must be loaded before the process begins. In this case the initialising variable (IV) can be sent in clear over the line in a preamble. It can be in clear because the block is

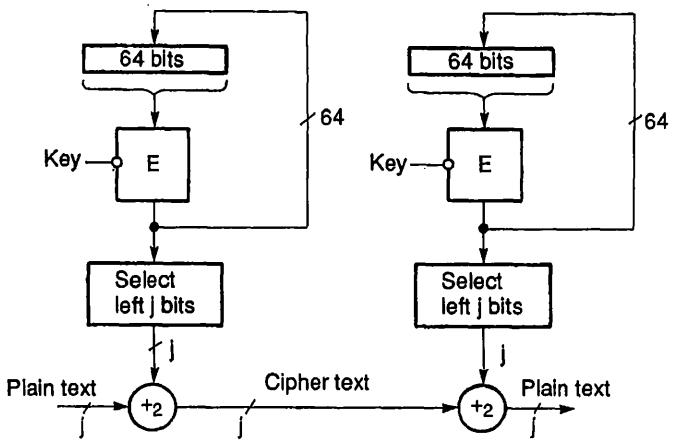

Fig. 8.4 Output feedback.

enciphered before it is ever used. Another possibility is to send no IV but to transmit random data for the first 64 bits, which will ensure that the two registers are synchronised.

For this method, there is no synchronising because there is no 64 bit or other boundary to be observed. If an error occurs on the line or, for example, a bit is missed or an extra bit inserted, this causes an immediate error in the delivered plaintext and also affects the shift register during the next 64 bits. Consequently the error extension property of this algorithm produces 65 bits of error for a single error on the line. Error extension is a penalty paid for the advantage of any self-synchronising cipher method.

When the error extension property of the CBC or CFC is unacceptable, another mode called *output block feedback* or OFB should be used. This also enciphers by means of adding, modulo 2, a random bit stream to the plaintext and adding the same bit stream to restore the ciphertext. In this case, the bit stream is generated *independently* of the message content and is therefore unaffected by errors on the line. The method is illustrated in Figure 8.4. Here, the contents of the register at the top of the diagram are enciphered and put back in the register for each cycle of operation, at both ends. Provided the key employed is the same and the registers start with the same values, they should remain in step and generate, for each operation, the same 64 bit pseudo-random output. This can be added into the message stream as any size of unit that is convenient, for example as 64 bit blocks or as 8 bit blocks to match the character size or as 1 bit at a time. For the highest speed of operation, the whole 64 bit block is used.

With this mode of operation there is no error extension but synchronisation is a real problem. Each end of the line has a pseudo random

number generator based on repeated DES encipherment of the register. The cycling of these two generators must be controlled by the clock rate of the line if they are to remain in step. A noise burst or short break in the transmission on the line can cause them to get out of step, after which the output at the receiving end is a random stream and communication is lost. In a practical system, there must be a means to *detect* loss of synchronism and *restore* it. For detecting loss of synchronism, there must be some redundancy in the plaintext which can be measured easily by the receiver. There must also be a return channel to tell the sender that synchronism has been lost, a protocol for exchanging a new starting value for the register and a method of restarting the two ends simultaneously. The starting value can be transmitted in clear across the line because it is enciphered before it is used. Some systems of this kind interleave information with the enciphered data which is accumulated to form a new starting value. It is possible to resynchronise at intervals, whether synchronism is lost or not and in this way be sure to restore synchronism, at least after an interval. Some systems resynchronise whenever a certain pattern is recognised in the ciphertext. However this can cause troubles when errors on the line make one end believe the pattern has occurred and the other does not. Maintaining synchronism in these stream ciphers is a difficult problem, particularly if they must be transparent to any pattern in the plaintext and therefore not dependent on plaintext redundancy to recognise loss of synchronism.

The three modes of operation that we have described, CBC, CFB and OFB, between them cover most of the applications of a block cipher for data confidentiality. We shall return later to its application for data integrity. These modes of operation are applicable to any block cipher and therefore form a rather general and useful aspect of cryptographic technique.

Cipher feedback can be applied with the feedback path carrying more than one bit on each cycle of the device. Apart from the one bit cipher feedback we have described, the next most common form is 8 bit cipher feedback in which 8 bits are selected from the left hand end of the cipher output for addition into the message stream. At each operation, the 8 bits of cipher text produced are shifted into the shift register, so that after 8 operations the register contents have been completely renewed. This is a convenient way to encipher an asynchronous transmission with 8 bit units. Other feedback widths are possible but are rarely used.

8.3.4 Key management

Under the heading of key management we include the entire life cycle of a key from its creation, through its distribution and use to its eventual

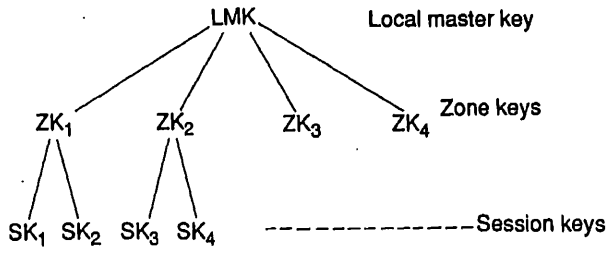

Fig. 8.5 A key hierarchy.

destruction. Keys should be chosen randomly within their total range (key space) and the ideal is to employ a true random number source such as a circuit which amplifies and samples random noise. A second best alternative is to generate a pseudo-random sequence using a cryptographic algorithm and some starting values called 'seeds'. Where random numbers must be generated within a smart card this is normally the method used.

For the distribution of keys, storage in a smart card for carrying to the place where they are to be used is effective, assuming that the smart cards are physically able to protect them from reading.

The principle employed in many key management schemes is that the same key should not be used too many times or over a very long period. For this reason, the keys that are used for enciphering or providing integrity checks on data are *session keys* which are changed frequently. These session keys can be transmitted over a communication line from a convenient central source provided that a master key is available to encipher them in transit. This is the first stage of a *key hierarchy* in which one key enciphers another. In practice, hierarchies of several layers are employed as illustrated in the example of Figure 8.5. This shows a *local master key* (LMK) at the top of the hierarchy which is used to encipher a number of *zone keys*. Each zone key is employed between a pair of units (or perhaps more) which have direct links with one another. The zone keys have a long life so they are used to encipher the session keys which actually carry out the encipherment of data or the integrity checking.

The top level key or LMK is used locally to encipher all the zone keys which are employed at that location. The purpose of this encipherment is to allow the zone keys to be stored outside the protected environment. This is necessary where there is a complex key hierarchy and the storage capacity of the cryptographic module is limited. These days, storage limitations are less frequent but external storage allows more than one security module to share zone keys if they all have the LMK stored in them. Zone keys are the master keys which operate between distant units and must be exchanged between these units. These can be carried in a smart card or similar protected unit. The encipherment of sessions keys

using zone keys is part of the automatic key distribution carried out over the communication lines.

The full details of a practical key management scheme are complex and should use sequence numbers and identifying information for all the session keys that take part. Several methods of linking these data together have been employed; and the terms *notarisation* and *offset* will be found in the literature and the standards on key management. Among the many aspects of the design of secure systems, key management is the one that is more often specialised to particular applications. In recent years, smart cards have taken an increasing part in these customised key management schemes.

There is one key management scheme published as ANSI standard X 9.17 and (modified) as an international standard ISO 8732. It is intended for the management of keys in wholesale financial transaction systems. Taking into account all its details it may not be suitable for use outside this range of application but the terminology of its messages and the schemes of encipherment it uses are found in key management schemes for other applications. Because of its complexity it is best referred to in the original standards documents.

The key hierarchy is a useful concept because it enables the working keys to change while reducing the frequency at which the top level keys must be transported. It can never eliminate the need for physically transporting some keys before a system can be started up. Because these top level keys control all other cryptographic keys, their security against illicit reading is vital and there are risks in physically transporting any secret data. A recent development in cryptography is the *public key cipher* and this can make a radical change in the top level key management.

8.3.5 Public key cryptography

The principle of cryptography shown in Figure 8.1 employs a secret key k as a parameter in both the encipherment and decipherment algorithms. This is the way in which individual pairs of users can ensure that their communications are produced from others who do not know the secret key. The principle of a secret key has been accepted as fundamental by cryptographers for more than a century and the majority of today's ciphers are of this kind. We call it a *symmetric* cipher because the knowledge of the sender and receiver is the same and because, with this secret key, encryption can be applied to information passing in either direction.

A remarkable new idea was proposed by Whitfield Diffie and Martin Hellman in 1976, who noted that only the receiver of the information, who needs to decipher it, requires a secret key. They proposed a type of

cryptography in which the key used for *encipherment* was not secret. Clearly, if there are two different keys they must be closely related in order that encipherment and decipherment shall be relatively inverse. If the encipherment key is not secret it is essential that knowledge of this key does not betray the secret decipherment key, so the relationship between the two keys is an unusual one. The tricky nature of the functions used in public key cryptography is even clearer if we consider that the whole process of encipherment, its algorithm and its key, can be made public without betraying the processing of decipherment. Anyone can perform the translation from plaintext to ciphertext but it requires secret information to perform the inverse function.

Functions of a kind which are easy to perform in one direction but for which the inverse is extremely hard to calculate have been known for some time under the name *one way functions*. Encipherment is a special type of one way function because it can be inverted if a secret key is known. This is sometimes called a 'one way function with a trap door'. After the first description of this principle, it was two years before a practical scheme was published by Rivest, Shamir and Adleman and this crypto system, the *RSA* cipher, is still the most practical public key system. Figure 8.6 shows the principle of a public key cipher.

As a starting point for generating the two keys we assume a seed value *ks* which has been chosen at random. From this value, using published algorithms F and G, the two keys are generated. *ke* is the encipherment key and *kd* the key used for decipherment. The key *ke* is described as a public key though there is no necessity to make it public. If an enemy does learn its value, this still does not reveal the secret key *kd* which must be retained and protected by the receiver of the information. The figure shows that the receiver should also control and protect the device which finds *ks*, generates *ke* and *kd* and uses the latter for decipherment. This is an *asymmetric* cipher and if two-way communication is needed the other party will have to generate his pair of keys and send out the public key for the encipherment of information on its way to him.

If the public key is really made public it allows anyone to encipher data and send it to the owner of the corresponding secret key. For a very large population of users of which anyone may need to send information to any other, this is a valuable feature because secure communication requires only that all parties shall know the public keys and then secret communication can take place between any pair. For n such users only n key pairs have to be generated and their public values placed on a list which is made known to all. With asymmetric or secret key cryptography, the number of keys needed for this situation would be $\frac{1}{2}n\ (n - 1)$.

Someone who receives a message enciphered with his public key can decipher it and, by its redundancy, know that it was enciphered with the correct key. This tells him that the sender knew his public key but if this

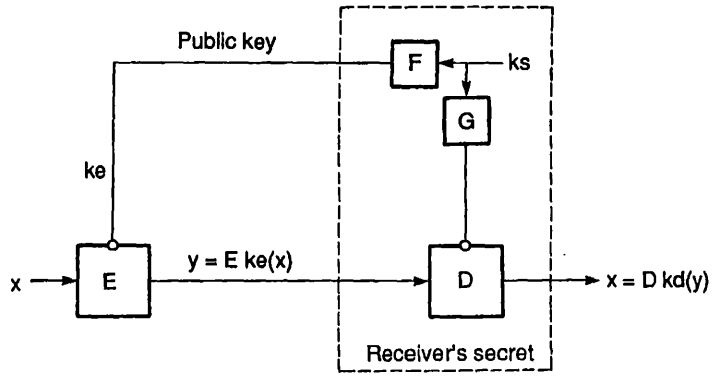

Fig. 8.6 Principle of the public key cipher.

key is widely known it does not identify the sender. In other words, there is no authentication of the received message. By contrast a symmetric scheme in which each communicating pair shares a secret key authenticates the sender of the message. There are many ways to provide this *source authentication* using public key cryptography. Suppose, for example, that each time a message is sent a random number field is included in the message. The reply from the receiver of the message also quotes this random number (as well as giving a new random number so as to continue the process). After the first message has passed, this procedure provides authentication of the origin of each message. The sender of each message knows the public key of the intended receiver and only that receiver can decipher the message. Therefore if his random number comes back in reply, it must have been deciphered at the destination, so the source of that reply must be the owner of the corresponding secret key. This authenticates the source of the reply. This simple procedure can maintain source authentication with public key cryptography. There are alternative ways to provide authentication using *digital signature*, described in 8.4.

The distribution of secret keys for symmetric ciphers is a great problem. With public key cryptography, the numbers to be exchanged do not require secrecy; they do require to be authentic. If an intruder could plant his own public key in place of the true one, he could intercept the messages and decipher them. But the intended receiver would find them undecipherable so the intruder has achieved little unless he can re-encipher the messages with the true public key and pass them on. This kind of active attack, which must intercept every message and re-encipher it without detection, is extraordinarily difficult and for many purposes could be regarded as impossible, nevertheless authenticity of the public key is an important security factor.

In a large community with a common set of public keys, this can be handled by installing a public key registry. The participants' public keys must be correctly registered and the users' access to the registry must be protected against falsification. Good solutions have been found to these problems by using the digital signature principle and by developing a good protocol for the key registration.

Public key operations involve heavy processing, in particular the operation of decipherment. It seems a fact of life that public key cryptography entails a heavy load of processing. For the first decade of this new technology the processing load imposed severe limitations on what could be done with it. Typical microprocessors took a minute or more for the decipherment operation and this could not be used directly for practical cryptography. Much faster special chips could have been made but none became available. The first application was therefore to ease the problem of distributing secret master keys for a symmetric key hierarchy. Compared with the physical movement of a secret key from one place to another, a minute or more of processing was convenient and much safer. The advance of technology has changed this so that now cheap microprocessors can perform the decipherment function in about a second. This still does not produce a good data encipherment rate but there are many transaction systems where it is adequate. Furthermore, the nature of the RSA cipher enables encipherment to be performed about 50 times as fast as decipherment by making a suitable choice of keys.

The first chip design for RSA operations came on the market in 1987 and gave an order of magnitude speed improvement over the fast microprocessors. This enabled bulk handling of RSA decipherment and made the use of public key cryptography without a symmetric cipher practical for the first time. The incorporation of RSA functions into a smart card is the next step. This can make a qualitative change in security and is the subject of Chapter 6 of this book.

Chip designs for much faster RSA processing have been proposed but commercial activity has been slow. The investment made in the development of DES chips was not very profitable for most suppliers. This has discouraged the design and development effort needed for chips for RSA operation. The gradual development of technology will overcome these limitations in a few years.

8.4 DATA INTEGRITY

By verifying data integrity we ensure that no unauthorised change has been made to information, that is to say we verify that no active attack has taken place. At one point the information is cryptographically 'sealed' in such a way that any later change can be detected. Two processes are

involved, the first which creates extra information at the time that the information is known to be correct and the second which allows the integrity of that information to be checked at a later stage.

Data integrity checking can be applied to communication by using the first process before sending and the second process when information is received. Alternatively it can be applied to stored data so that the reader of the data can verify its integrity, meaning that it has not been changed by an unauthorised person while it was in store.

8.4.1 Principle of integrity checking

The principle of the cryptographic method is shown in Figure 8.7 in the context of data communication. A message M is transmitted from source to destination. Together with the message is sent a number A which is called an *authenticator*. More strictly, it should be called a *data integrity check field*. The value of A is derived from the whole of the message using a calculation which employs a secret key k as parameter. At the receiving end, the same calculation is performed on the received data using the same secret key and the result is compared with the incoming value of A. If these are equal, the integrity of the message M has been verified. If an error appears at this point, it may be due to deliberate unauthorised changes on the line, or it could be due to an error or using the wrong key.

The algorithm used is cryptographic, in the sense that if a number of examples of messages together with their authenticator values are known, it must not be possible to deduce the value of k. The value of A must also depend on every bit of the message so that no part of the message can be altered without detection. In general, authenticators have been developed from cryptographic algorithms but this is not essential because, unlike a cipher algorithm, no corresponding inverse is needed.

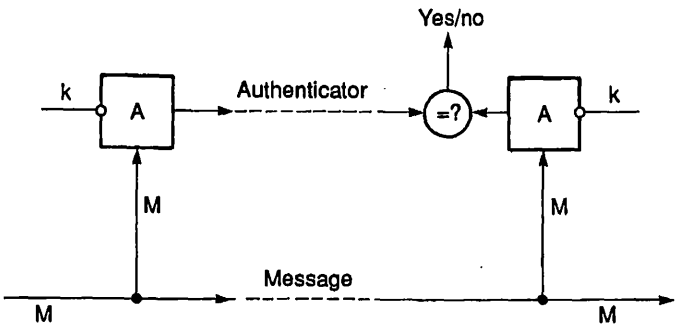

Fig. 8.7 Principle of message authentication.

The size of the authenticator field is not dependent on the message size so the authenticator, although it depends on the whole message, is usually of a fixed size, such as 32 bits. The main criterion is that the equality on which the verification of identity depends should be very unlikely to happen by accident. In fact, a 16 bit authenticator with a probability of accidental agreement of 1/65536 would be sufficient for most purposes, but 32 bit values have become well established.

8.4.2 Prevention of misuse and replay of messages

If the same key is used for some time, entire messages might recur, and then their authenticator values would be identical. Making use of this, an intruder could repeat an old message and have it accepted as genuine. This would be serious if the message is a payment. Another possibility is that a message intended for one purpose could be misused for another and the authenticator would still match; although data integrity in a narrow sense has been preserved, this would not be a safe system.

Both these problems can be solved by careful choice of the format of the message which is authenticated. To avoid misuse of the message, it should contain full identifying information such as its source, persons authorising it and the identity of any other persons or accounts involved. Provided this information is comprehensive, misuse of the message becomes impossible. For example, all key management messages should contain the identity of the source, the purpose for which the key is to be used and any other information which limits its range of use.

To prevent the reuse of a message for its original purpose, all messages should carry a sequence number of some form. If the messages are passing between two entities only, a sequence number can be maintained for this one source and destination. The receiver checks the sequence number. If one source sends messages to a number of destinations, the destinations can check the increasing sequence but not that the sequence is complete. This at least ensures that no message can be reused but does not guard against the deletion of a message in the sequence.

Where deletion of messages is not a significant risk, the sequence number can be replaced by a time and date stamp. This is very convenient when a large population can be either senders or receivers of messages, as in electronic mail.

8.4.3 Algorithms for data integrity checking

Data integrity checks for financial messages are, in most cases, based on the DES algorithm. A method of using this algorithm to calculate an authenticator field is shown in Figure 8.8. This corresponds with the

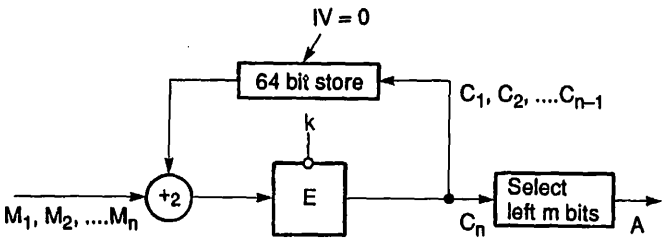

Fig. 8.8 Authentication using the cipher block chaining mode.

mode of operation we described in section 8.3.3 as cipher block chaining. When CBC mode is used for calculating an authenticator, the IV is zero. The message is divided into sections of 64 bits, padding out the final block to 64 bits if necessary. (Padding out with zeros is acceptable.) The first block is enciphered using the secret key, then added modulo 2 to the second block and enciphered again. This process is continued until all the blocks have been used. From this procedure, a 64 bit number is produced and, by convention, we use the left hand 32 bits of this result.

The usual name for this type of authenticator is a *message authentication code* (MAC). The terminology is so well established that the term MAC normally denotes an authenticator calculated in this way with the DES algorithm. The current ISO terminology reserves the term *authentication* for the authentication of peer entities, notably user authentication. It does not introduce a satisfactory word for the concept of *data integrity check function*. Until a useful single word is adopted it seems acceptable to use the word *authenticator* for this quantity. The term MAC is less satisfactory because the 32 bit field is not strictly a 'code'. It has been so widely adopted that use of the acronym MAC is unavoidable.

There is no published integrity algorithm that has been used as widely as the MAC. When it was considered as an ISO standard for financial messages, the inconvenience of a software implementation of DES was recognised and an alternative algorithm developed specifically for use with mainframe computers was adopted as an alternative. This so called *message authenticator algorithm* (MAA) employs a 32 bit integer multiplication and is therefore not only suitable for mainframes but also for microcomputers of more recent design. In this algorithm, messages are divided into blocks of 32 bits with padding of the final block if necessary. A 64 bit key is used from which an initial calculation produces six 32 bit numbers. Two of these numbers form initial values for the calculation, two of them are used during the cycles which successively employ the 32 bit blocks of the message and two of them take part in a final two cycles of the calculation called the *coda*. The running calculation operates on and produces two 32 bit numbers and, at the final step, these are added

modulo 2 to form the authenticator value. The full definition of the MAA algorithm is contained in ISO 8731 part 2. If the message is longer than 1024 bytes (256 blocks of 32 bits) an authenticator is calculated from the first 1024 bytes then prefixed to the next 1024 bytes and the calculation of the authenticator is started again. The authenticator (derived from the final block) is the authenticator for the whole message.

There are many other data integrity algorithms in use but some are proprietary algorithms which have not been made public. Among these is the algorithm used by SWIFT. In smart cards, a proprietary algorithm known as *Telepass* has been used. As the capability of the microprocessor in smart cards improves, it may become possible to adopt the ISO algorithm MAA.

8.4.4 Digital signature

The DES based algorithm and MAA employ a secret key which must be known to the entity which calculates the authenticator value and also to the entity which checks it. If this key is to be kept secret it must not be made widely known; therefore this type of integrity check is primarily used bilaterally. It needs the prior distribution of a secret key. The technique protects the information against falsification by any third party who does not know the secret key. It does not protect either of the two entities against dishonesty of the other. The receiver of the information could alter it and since, with a knowledge of the secret key, he could generate a valid authenticator value, this does not protect the information against changes. This technique would be worthless as a method for protecting the information in a bank cheque, because the receiver has a motive for falsifying it.

Because the receiver could alter or forge a document the sender can falsely accuse him of doing so. After having sent a message, such as payment, the sender may wish to deny having done so and the evidence of an authenticator value calculated with a secret key gives no protection for the receiver against this charge. If this kind of dispute between the sender and receiver of a message is to be resolved, they must not share exactly the same information, in other words the *asymmetric* nature of public key cryptography is required in this situation.

A simple change in the operation of a public key cipher can provide a type of authentication which meets the requirement exactly. This is illustrated in Figure 8.9 where the algorithms and notation are those of Figure 8.6 and the operations have been changed around so that the message x is first subject to 'decipherment' then to 'encipherment'. For the RSA algorithm the two operations D and E are, in fact, identical and their different effect is due only to the keys they use.

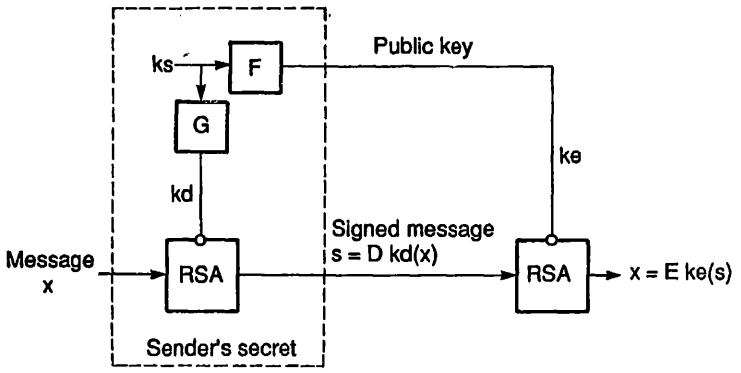

Fig. 8.9 Principle of public key signature.

The message produced in this way is transformed but is not secret because anyone, with the aid of the public key, can restore it to its plaintext form. The existence of the restored message (recognisable as such by its redundancy) is proof that it came from the sender who owns the public key. Therefore, if the public key is authentically known, the transformed message is evidence of the message's origin. In effect, the value *s* and the public key *ke* could be taken to a third party for arbitration. The existence of the restored message text *x* as a function of these quantities, together with the known identity of the owner of the public key, is proof of the origin of the message. Figure 8.9 shows the principle of the method but in this form it would be inconvenient. The receiver of the message would need to store the 'signed message' *s* but for all practical purposes only the plaintext x is useful. For very long messages this would be particularly inconvenient. Also, it does not correspond with the way in which we understand signatures, where the signature is a relatively short piece of information added to the message. A more convenient method is shown in Figure 8.10. Here, the message is sent in plain; a signature consisting of one RSA block is formed by applying the signature process to the quantity H(M) which is a one-way function of the whole message. At the receiving end, the same one-way function is formed and is compared with the quantity generated by applying the RSA process to the received signature.

There are many ways to calculate the one way function H, some using the DES algorithm and others using RSA operations. No standards have yet been established in this area. With the signature scheme of Figure 8.10, messages of indefinite length can be signed with one RSA block, which is typically of length 64 bytes.

Encipherment with a public key cipher does not, by itself, provide authentication of the message. The signature process, in either of the

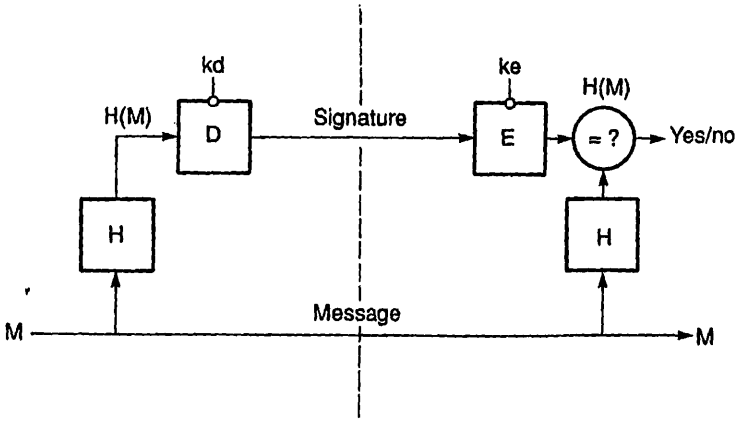

Fig. 8.10 Signature separated from the message.

forms we have described, does not provide message confidentiality. By comparison, an asymmetric cipher provides encipherment and, with certain precautions in the format of the message, it also provides a measure of authentication. Public key cryptography separates these two properties into two different processes. If both encipherment and authentication are needed, the secret keys of the sender and the receiver are both involved. The message shown in Figure 8.10 can be enciphered and the encipherment can include the signature block or not, as required. Alternatively, the simple protocol described in section 8.3.5 can provide authentication for an exchange of messages.

Both digital signature and public key cryptography depend on the authenticity of public keys. In a large group who use it for message interchange all the public keys are held in a central registry. For distribution of these keys the messages can be signed using the key of the registry so that each user need only have an authentic value for the registry public key to make the whole key distribution scheme secure.

One of the main virtues of digital signature as an authentication method is that the public key can be made known as widely as required. Therefore messages authenticated in this way can be checked by any party which requires to verify their authenticity. This is important when a message passes through a number of hands in the course of its processing or when a single message is to be broadcast widely.

A practical use of digital signature was held back by the computational complexity of public key cryptography but the introduction of fast processors now makes it practical for nearly all purposes, soon including its use in smart cards. By using a short exponent in the RSA scheme the process of checking a digital signature can be made very fast so that only

the generation of the signature requires the full length RSA operation, this taking about one second in a fast microprocessor.

8.5 USER AUTHENTICATION

In a communication system with a layered architecture (such as OSI) two entities which communicate with each other are in the same layer and are called 'peer entities'. Authentication is the means by which two communicating peer entities each establish each other's identity in a secure manner. This is the general case of *peer entity authentication*.

As an example, consider two cryptographic modules which are used to verify the integrity of data passing between them. They do this by means of a secret key which they share in common. If information is checked with the secret key and proves to be genuine, it must have arisen from a place where a secret key is known so that the appropriate check function (authenticator) could have been produced. Therefore, if the key management is safe, this also provides peer entity authentication. The receiver knows not only that the information received is genuine but also that it arose from the corresponding module (peer entity) since it is the only one holding that secret key. Thus peer entity authentication in this case is a natural consequence of secure key management. Indeed, data integrity is of very little value if it is not combined with authentication of the source of the data. Data which has not been falsified during the communication process is of no value if it was falsified by a masquerading peer entity.

The type of peer entity authentication which is of most importance in practice is authentication between a *user* and a computer system with which the user is interacting through a communication network. If the users are mobile and can appear at different terminals or at different ports on the communication network, authentication of the terminal is not enough and there must be authentication of the user himself. Generally this must be 2-way authentication so that the user is not fooled into interacting with a masquerading central system which could cheat him into believing that a transaction had been made or obtain information from him by this trick.

8.5.1 Existing methods

By comparison with the security which can now be obtained, user authentication is often no more than the presentation of a password, which can be very insecure. The best password systems can have a reasonable level of security. These use passwords which change at every access attempt and these passwords are presented both by the user to the system and by the system to the user.

Passwords came into use with the first interactive systems because they could be implemented with nothing more than a simple program at the host. Far too much trust was placed in this insecure method with the results that we now see in the activities of 'hackers'.

In an attempt to provide better user authentication or to supplement the security of passwords, a number of schemes for 'biometric' user authentication have been developed. These attempt to measure a characteristic of the user which he cannot easily change such as a fingerprint, shape of hand or blood vessels on the retina. Alternatively they use an action by the user over which he has little conscious control such as the dynamics of his manual signature or his voice. Biometric methods have been improved steadily but they all have perceptible error rates either by rejecting genuine users or accepting false ones. There is a trade-off between these two types of error but some level of error seems unavoidable.

8.5.2 User authentication employing a smart card

A cryptographic exchange of data can provide effective peer entity authentication but if one of the entities is a person, he must hold a cryptographic device on which the processing can take place and which he will ensure is not lost or stolen. This is an example of a token (something which a person holds) being used to authenticate the user. Tokens such as magnetic striped cards, which have no processing power, can easily be read, copied or forged and are insecure. To be effective, a smart card is needed, which can protect its key physically and also contains the cryptographic function for performing the authentication protocol.

There are several protocols that can be used. The most commonly used protocol is shown in Figure 8.11. The host sends a random number R (the challenge) which is processed by the card using the secret key k and returned (S) to the host. The host also has the secret key and can perform the same calculation, comparing its result with the one received from the smart card. If these agree the smart card is taken to be authenticated and, by implication, the user also. A card can be protected against unauthorised use by someone who finds or steals it by requiring a password or personal identification number to be given before it will perform its cryptographic exchange. The password can be entered either by a keyboard on the token itself or from a separate keyboard through its electrical interface. The latter raises the problem of a false terminal which can capture the password, though this is of little use without having the token as well.

The challenge and response procedure can use a cipher algorithm in which the secret key held in the card is used to encipher the challenge and produce the response. The inversion property of a cipher algorithm is not needed so any cryptographic function, such as a one-way function could be substituted.

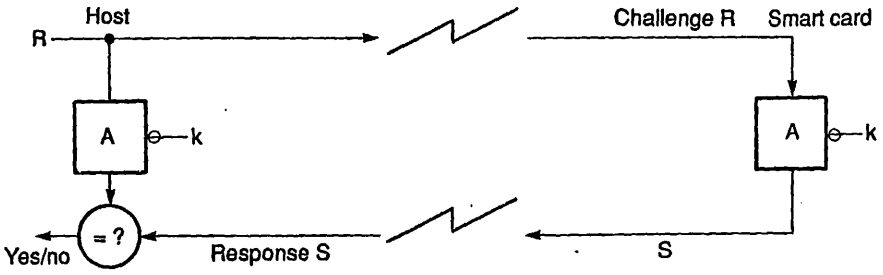

Fig. 8.11 Challenge/response method of authentication.

The token need not have an electrical interface with the terminal if it has a keyboard and display. The user can enter the challenge into the token and read the response from it. There are on the market a number of devices resembling hand-held calculators which work in this way. It requires extra effort from the user, who must read the challenge from the screen of a terminal and enter it into the token, then read the response from the token and enter it into the keyboard of the terminal. The advantage of a token without an electrical interface is that it can be applied to the hardware of the terminal, in fact it can be used by telephone where the numbers are spoken from one person to another.

In order to reduce the effort needed with the non-interface token the transfer of the challenge number can be dispensed with if the response is made to depend on a changing number stored in the token for which a corresponding number is maintained at the host. Figure 8.12 shows one such principle. Registers at the host or in the card hold a value x. This is processed using the secret key k to generate the response S. Then this value becomes the x value for use on the next occasion. This simplifies the user's procedure but ties the token to just one host which is, of course, the most common requirement. Other devices on the market employ a quartz crystal clock in the token to keep it in synchronism with

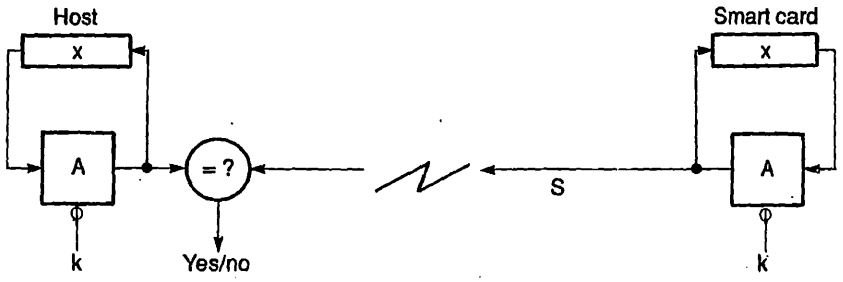

Fig. 8.12 Authentication with a stored value.

the host, enabling one token to be used with a number of hosts. The clock's time and date form the 'challenge' from which the cryptographic response is sent via the user, who transcribes it to the terminal, to the host. Because the clocks will slowly lose synchronism, each active host must make small corrections and be prepared to try a few challenge-response pairs to determine whether the token is genuine.

Hand-held devices for cryptographic authentication of users are valuable chiefly because they can be fitted easily to existing systems. Where a new system is being designed, the electrically coupled smart card is preferable.

8.5.3 Authentication by digital signature

Cryptographic exchanges using secret keys can provide excellent security but a single token's use is necessarily confined to the corresponding hosts which possess these secret values. A much more general method of user authentication can be developed using digital signature. This is the basis of the signature token described in Chapter 6.

The principle is that the token generates a public key pair and holds its secret key protected inside it. Any host wishing to use this token for authentication is given the public key. The cryptographic exchange consists, as before, of a challenge and response. The response consists of the 'digital signature' of the challenge and can easily be verified with the use of the public key. It is, however, important that the challenge cannot be used to create a message which could falsely be said to have been generated by the owner of the token. The protocol and an agreed message format prevent this.

Because the information used by the host consists only of a public key, the number of hosts able to collaborate with a given token is unlimited. In practice, the useful transactions which could follow this authentication stage must be agreed between the host and the token. Therefore it is likely that these tokens will be specialised to certain types of application, such as banking. Within this general area, provided there are agreements on protocols and message formats, the range of possible transactions could be large. For these transactions, a number of banks and financial institutions could collaborate with a single token held by the user. The digital signature principle overcomes the security problems of this arrangement.

8.5.4 The Fiat Shamir authentication principle

Development of a practical signature token is well advanced but there would be an advantage in a comparable authentication method that could

be performed with less computation than the RSA calculations require. In particular, calculating the digital signature is computationally complex, while verifying the signature can be simplified by using a small public exponent. Fiat and Shamir have described such a scheme which is based on the properties of square roots in modular arithmetic.

Like the RSA method, it employs a modulus which is the product of two primes. For the public key, only the product is made known and the numbers are large enough that factorisation is practically impossible. Given such a composite modulus, the properties of square roots that are employed are as follows. About one quarter of all numbers have square roots and these are called *quadratic residues*, QRs. Nearly all the QRs have four square roots which are arranged as two pairs. If m is the modulus a pair is x and $m - x$, rather like positive and negative roots of ordinary arithmetic. It is convenient to make the factors of the modulus of the form $4n + 3$ since it makes it easy to discover if a number is a quadratic residue and, if it is, to find a square root.

The property used in this authentication method is that, if the factors are known, it is easy to calculate the square root of a QR but if the factors are not known, finding the square root is as difficult as factorising the modulus.

Fiat and Shamir developed a number of schemes but, for simplicity, we shall describe the simplest of these methods, though it is not the most efficient.

When the card is first established, it finds a suitable modulus m as if it were an RSA token. The user then generates a message containing the identifying information which the card will authenticate to any enquirer. If necessary, a one way function of this message can be used to reduce its bulk. The resultant number must be made into a QR, which can be done by having a small field the value of which can be adjusted to make it a QR. The card is then able to calculate a square root of this identifying information. That is, it finds a quantity u such that $u^2 = v$, modulo m, where v is the identifying information.

The problem is for the card to prove to an enquirer that it contains v without revealing that value. This is handled by a repeated dialogue of the following kind.

(1) The card finds a random value r and sends $x = r^2$ and $y = v/x$ to the terminal which checks that $x.y = v$.
(2) The terminal sends a one bit decision e to the card to decide the next move.
(3) If $e = 0$ the card sends r to the terminal which checks that $r^2 = x$. If $e = 1$ the card sends $s = u/r$ to the terminal which checks that $s^2 = y$.
(4) The terminal checks that $r^2 = x$ or $s^2 = y$ according to its choice of e.

If both r and s had been sent, the secret u would have been lost, since u = r.s. By offering to send either, but not both, at the choice of the terminal, the card shows its confidence that it knows both, hence knows u.

From the terminal's viewpoint, either x or y might have been constructed as r^2 where r is a chosen value so the outcome could have been a piece of luck with probability ½. For this reason, the above dialogue must be repeated with different values r many times, so that the probability of the card succeeding by luck is very small. For example, 20 iterations would be sufficient.

These 20 operations are still much less onerous than a single RSA exponential calculation since the multiplications and divisions involved are very small tasks compared with exponentials. The communications load for the 60 or so numbers must also be considered. This is unimportant when the authentication is a purely local function.

In an improved scheme of the same kind, the terminal makes five binary choices at each stage instead of just one and, in this case, four cycles of dialogue gives the same low probability of the card succeeding by luck. This also reduces the communications load.

There will probably be useful applications for the Fiat and Shamir authentication scheme but, unlike the RSA procedure, it cannot function as a signature method. A record or 'transcript' of the dialogue is of no value in proving the authenticity of the card after the event because it could be forged by deliberately checking always the easy option, since x and y cannot be distinguished. Fiat and Shamir have produced a true signature scheme using the same principle which offers a saving of computation compared with RSA signature by a factor of about 20. Unfortunately, the amount of data stored in the card and the size of the signature are both increased. Since the original publication, many new versions of this 'zero knowledge' protocol for authentication and corresponding signature schemes have appeared, so that various trade-offs between computational complexity and sizes of keys and communicated data are now available.

8.6 THE FUTURE OF CRYPTOGRAPHY IN THE SMART CARD

Improvements in semiconductor technology will allow closer packing of logic and storage devices on chips, higher processing speeds and lower power consumption. This will remove some of the constraints on complex cryptographic algorithms.

Current technology allows the use of the DES algorithm in a smart card of standard bank card dimensions, in spite of the rather poor match between the nature of the DES algorithm and the simple processors employed in smart cards.

The replacement of the DES algorithm by a more secure cipher and one which is better adapted to be implemented in software or firmware is an urgent requirement and is technically possible but is unlikely to be worked out in the public arena and made the subject of a national or international standard. Where the requirements of banking and other financial services are concerned it is possible that a suitable algorithm for limited use within this community could be developed but there is no organisation established internationally which can easily take up this challenge.

In the next few years, the implementation of the RSA cipher in compact *signature tokens* will be feasible. There is a real possibility for the use of the RSA cipher to overtake the DES and become a *de facto* international standard, regardless of the pressures on ISO. Digital signatures employed within user-held tokens, as described in Chapter 6, can provide a very effective mechanism for personal transactions of all kinds. Though the Fiat Shamir authentication scheme is attractive, the more general purpose nature of the RSA algorithm as cipher, signature and one way function, makes this in the longer term, more significant.

The security of a user token depends not only on its cryptography but also on its physical security and the way in which information enters and leaves it. In this respect, the thin card of the present banking standard may not remain the best choice. A keyboard on the card and a display can improve the security by verifying to the user the exact nature of the transaction being performed, providing the user with information authenticated in the card and receiving instructions and the password (PIN) directly instead of routing it through a less secure terminal. Also, the coupling of the token to the terminal or communications interface must eventually avoid electrical contacts. For all these reasons, a slightly thicker and rigid format for the user-held token seems likely to emerge.

The technology will evolve and, whatever standards eventually emerge, the interplay between cryptography and other aspects of security in the smart card of the future is an interesting prospect.

REFERENCES

[1] Davies, D. W. and Price, W. L. (1984) *Security for Communication Networks* John Wiley.

Index

Page numbers in italics refer to illustrations.