(19) Japan Patent Office (JP)

(12) **Japanese Unexamined Patent Application Publication** (A)

(11) Japanese Unexamined Patent Application Publication Number

H11-120092

(43) Publication date: April 30, 1999

(51) Int. Cl.[6]	Identification codes	FI	
G06F 13/00	301	G06F 13/00	301P
3/06	304	3/06	304B

Request for examination: Not yet requested Number of claims: 14 OL (Total of 11 pages)

(21) Application number	Japanese Patent Application H10-181019	(71) Applicant	597004720 Sun Microsystems, Inc. 901 San Antonio Rd. MS PAL 01-521, Palo Alto, CA 94303, United States of America
(22) Date of application	June 26, 1998	(72) Inventor	Fay Chong 10019 Oakleaf Place, Cupertino, CA 95014, United States of America
(31) Priority claim number	883923		
(32) Priority date	June 26, 1997		
(33) Priority country	United States (US)	(74) Agent	Patent Attorney Kazuo SHAMOTO (and 5 others)

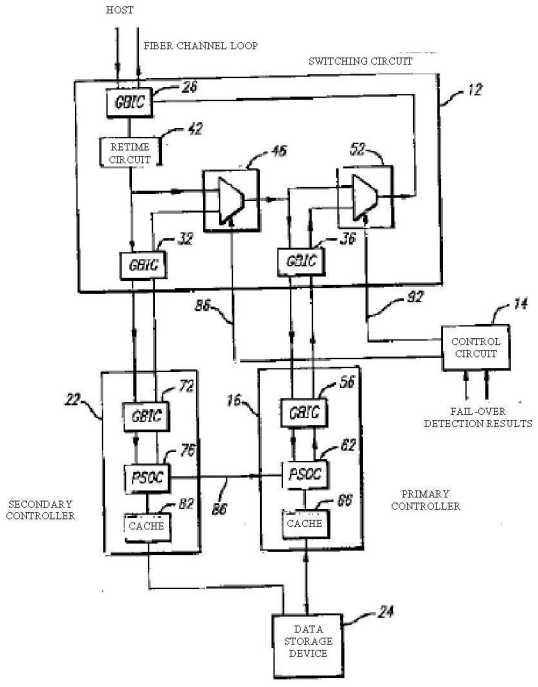
Continued on the last page

(54) (TITLE OF THE INVENTION) METHOD AND APPARATUS FOR HIGH AVAILABILITY AND CACHING OF DATA STORAGE DEVICES

(57) (ABSTRACT)

(PROBLEM) To enable high availability and efficient caching of data storage devices.

(MEANS FOR SOLVING) The apparatus of the present invention comprises a primary controller 16, a secondary controller 22 having the same address as that of the primary controller, a switching circuit 12 coupled to the primary and secondary controllers, and a control circuit 14 coupled to the switching circuit. In a normal operation, the control circuit sets the switching circuit so that the primary controller receives and responds to input data supplied from a host and the secondary controller receives the input data. In a fail-over operation in which the primary controller fails, the control circuit sets the switching circuit so that the primary controller is disabled and the secondary controller receives and responds to the input data supplied from the host. The fail-over is transparent to the host.



DOCKET ALARM

Find authenticated court documents without watermarks at docketalarm.com.

(SCOPE OF THE PATENT CLAIMS)

(CLAIM 1) An apparatus, responsive to a host, for high availability and caching of data storage devices, comprising:

a primary controller;

a secondary controller having the same address as that of the primary controller;

a switching circuit coupled to the primary and secondary controllers; and

a control circuit coupled to the switching circuit;

wherein in a normal operation, the control circuit sets the switching circuit so that the primary controller receives and responds to input data supplied from the host and the secondary controller receives the input data; and

in a fail-over operation in which the primary controller fails, the control circuit sets the switching circuit so that the primary controller is disabled and the secondary controller receives and responds to the input data supplied from the host.

(CLAIM 2) The apparatus of claim 1,

wherein the switching circuit includes first and second multiplexers respectively coupled to the primary and secondary controllers; and

the control circuit switches the first and second multiplexers to respectively control the primary and secondary controllers to the normal and fail-over operations.

(CLAIM 3) The apparatus of claim 2, wherein the primary controller includes a first cache, and the secondary controller includes a second cache.

(CLAIM 4) The apparatus of claim 2, further comprising a data storage device coupled to both the primary and secondary controllers.

(CLAIM 5) A method for providing high availability and caching for a data storage device, comprising:

(a) a step of preparing a primary controller;

(b) a step of preparing a secondary controller having the same address as that of the primary controller;

(c) a step of coupling a switching circuit to the primary and secondary controllers;

(d) a step of coupling a control circuit to the switching circuit;

(e) a step of controlling, in a normal operation, the control circuit to set the switching circuit so that the primary controller receives and responds to input data supplied from a host and the secondary controller receives the input data; and

(f) a step of controlling, in a fail-over operation in which the primary controller fails, the control circuit to set the switching circuit so that the primary controller is disabled and the secondary controller receives and responds to the input data supplied from the host.

(CLAIM 6) The method of claim 5,

wherein the switching circuit includes first and second multiplexers respectively coupled to the primary and secondary controllers; and

each of steps (e) and (f) comprises a step of switching the first and second multiplexers to respectively control the primary and secondary controllers.

(CLAIM 7) The method of claim 6, wherein the primary controller includes a first cache, and the secondary controller includes a second cache.

(CLAIM 8) The method of claim 6, further comprising a step of coupling a data storage device to both the primary and secondary controllers.

(CLAIM 9) An apparatus, responsive to first and second hosts, for high availability and caching for data storage devices, comprising:

a first controller including a first primary controller and a second secondary controller;

a second controller including a second primary controller and a first secondary controller, wherein the first secondary controller is a backup of the first primary controller and the second secondary controller is a backup of the second primary controller;

a switching circuit set coupled to the first and second controllers; and

a control circuit coupled to the switching circuit set;

wherein in a normal operation, the control circuit sets the switching circuit set so that the first primary controller receives and responds to input data supplied from the first host and the first secondary controller receives the input data from the first host, and the second primary controller receives and responds to input data supplied from the second host and the second secondary controller receives the input data from the second host; and

in a fail-over operation in which one of the first and second controllers fails and becomes a failing controller and the other of the first and second controllers becomes a surviving controller, the control circuit sets the switching circuit set so that the primary controller of the failing controller is disabled and the secondary controller of the surviving controller receives and responds to the input data directed to the primary controller of the failing controller, and the primary controller of the surviving controller receives and responds to the input data directed to the

DOCKET ALARM
Find authenticated court documents without watermarks at docketalarm.com.

surviving controller.
(CLAIM 10) The apparatus of claim 9, wherein the switching circuit set includes:
a first switching circuit coupled to the first primary controller and the secondary controller in the first controller; and
a second switching circuit coupled to the second primary controller and the first secondary controller in the second controller.
(CLAIM 11) The apparatus of claim 10,
wherein the first switching circuit comprises first and second multiplexers; the second switching circuit comprises third and fourth multiplexers;
and the control circuit includes first and second control circuits;
wherein the first control circuit switches the first and second multiplexers to respectively control the first primary controller and the first secondary controller, and the second control circuit switches the third and fourth multiplexers to respectively control the second primary controller and the second secondary controller.
(CLAIM 12) The apparatus of claim 11,
wherein the first controller includes a first cache coupled to the first primary controller and the second secondary controller; and
the second controller includes a second cache coupled to the second primary controller and the first secondary controller.
(CLAIM 13) The apparatus of claim 11, further comprising first and second data storage devices, each coupled to both first and second controllers.
(CLAIM 14) The apparatus of claim 9, wherein the circuit is a fiber channel circuit.
(DETAILED DESCRIPTION OF THE INVENTION)
(0001)
(TECHNICAL FIELD OF THE INVENTION) The present invention generally relates to a method and apparatus for high availability and caching of data storage devices, and more particularly to a method and apparatus for performing efficient caching and allowing fail-over (that is, switching (switch-over)) in controllers and/or data storage devices to be transparent to a server or a host computer.
(0002)
(PRIOR ART) In a typical client-server system, a plurality of clients are coupled to one or more servers, and the servers are, in turn, coupled to one or more data storage devices. The clients can access files in the data storage devices through associated servers. High availability and efficient caching are important to data storage devices in order to maintain high data communication performance. To achieve these goals, fail-over has been executed in data storage devices and has been commercially available for a number of years. However, in conventional implementations, it is necessary to run software on the server (or host computer) to redirect the I/O from the server over an alternate path or the same path but to a different data storage device address – that is, a secondary address. Different vendors typically use different types of software to control fail-over operations in data storage devices. In a typical corporate environment, equipment from multiple vendors may be used to construct a client-server system. The server of one vendor may include fail-over software that is incompatible with the data storage devices of another vendor. In order to properly set up the system in such a situation, it is necessary to perform extensive testing to resolve the problem of incompatibility. As a result, it is extremely inefficient and time-consuming to set up such a system. One way to solve the incompatibility problem is to always purchase equipment from the same vendor. However, this leads to a loss of flexibility in equipment selection and future upgrades to the system. Restricting hardware buyers to a single vendor also leads to increases in cost.
(0003)
(PROBLEM TO BE SOLVED BY THE INVENTION) Therefore, there is a need for a method and apparatus for high availability and caching of data storage devices allowing fail-over in the controllers and/or data storage devices to be transparent to a server so that extensive testing to resolve incompatibility between equipment of different vendors can be substantially minimized.
(0004)
(MEANS FOR SOLVING THE PROBLEM)
The present invention provides a method and apparatus for achieving high availability and caching of data storage devices. According to a preferred embodiment of the present invention, there is provided an apparatus comprising a primary controller, a secondary (sub-) controller having the same address as that of the primary controller, a switching circuit coupled to the primary and secondary controllers, and a control circuit coupled to the switching circuit. In this preferred embodiment of the present invention, in a normal operation, the control circuit sets the switching circuit so that the primary controller receives and responds to input data supplied from a host and the secondary controller receives the input data. In a fail-over operation in which the primary controller fails (failure occurs), the control circuit sets the switching circuit

DOCKET ALARM
Find authenticated court documents without watermarks at docketalarm.com.

so that the primary controller is disabled (deactivated) and the secondary controller receives and responds to the input data supplied from the host. In addition, the apparatus may further comprise a data storage device coupled to both the primary and secondary controllers.

(0005) According to a second preferred embodiment of the present invention, there is provided an apparatus, responsive to first and second hosts, for high availability and caching of data storage devices. This apparatus comprises first and second controllers, a switching circuit set, and a control circuit. The first controller includes a first primary controller and a second secondary (sub-) controller, and the second controller includes a second primary controller and a first secondary (sub-) controller. The first secondary controller is a backup of the first primary controller, and the second secondary controller is a backup of the second primary controller. The first and second controllers are coupled to the switching circuit set, and the switching circuit set is coupled to the control circuit. According to this alternative preferred embodiment of the invention, in a normal operation, the control circuit sets the switching circuit set so that the first primary controller receives and responds to input data supplied from the first host and the first secondary controller receives the input data from the first host. Further, the second primary controller receives and responds to input data supplied from the second host and the second secondary controller receives the input data from the second host.

(0006) According to this alternative preferred embodiment, in a fail-over operation in which one of the first and second controllers fails, the control circuit sets the switching circuit set so that the primary controller of the failing controller is disabled and the secondary controller in the surviving controller receives and responds to the input data directed to the primary controller of the failing controller. In this fail-over operation, the primary controller of the surviving controller receives and responds to the input data directed to the surviving controller. Further, the apparatus may also comprise first and second data storage devices, each coupled to both first and second controllers.

(0007) Other attainments and a fuller understanding of the present invention will become apparent and appreciated by referring to the following explanations and claims with reference to the accompanying drawings.

(0008)

(EMBODIMENTS OF THE INVENTION) FIG. 1 is a functional block diagram of a fiber channel circuit for high availability and caching of data storage devices according to a preferred embodiment of the present invention. The present invention may be implemented in electronic circuitry. As illustrated, a switching circuit 12 is coupled to a control circuit 14, a primary controller 16 and a secondary controller 22. On the other hand, the primary and secondary controllers are coupled to a data storage device 24. The switching circuit 12 includes GBICs (Gigabit Interface Converters) 26, 32, and 36, a retime circuit 42, and multiplexers 46 and 52. The primary and secondary controllers 16 and 22 are identical to one another in this embodiment. The primary controller 16 includes a GBIC 56, a PSOC (Serial Optical Converter for PCI bus) 62, and a cache 66. Similarly, the secondary controller 22 includes a GBIC 72, a PSOC 76, and a cache 82. Also included in each of the primary and secondary controllers 16 and 22 is fail-over software (not shown), which detects whether there is a failure in itself, the other controller, or the data storage device 24. The Sun Energizer, which is commercially available from Sun MicroSystems, Inc. in Mountain View, CA, may be used as the fail-over software. The fail-over software detection results are sent to the control circuit 14 so as to respectively control the multiplexers 46 and 52 via control lines 86 and 92. Each GBIC is a conventional interface converter and is commercially available, for example, from the Vixel Corporation in Lynnwood, WA. The retime circuit 42 is also a conventional circuit that adjusts (aligns) data pulses and converts pulse edges to discrete (individual) boundaries. Each PSOC includes a buffer for storing input data received from the host and transfers the data from its buffer to its cache in accordance with, for example, the Arbitrated Loop standards. As an alternative to a PSOC, an ISP2100 Intelligent Fibre Channel Processor may be used, and this is commercially available from the QLogic Corporation in Costa Mesa, CA. Input data in the form of fiber channel frames are sent to the primary and secondary controllers 16 and 22 and the data storage device 24 via a fiber channel loop in accordance with the Arbitrated Loop standards, for example. Both primary and secondary controllers 16 and 22 have the same address.

(0009) The normal operation is illustrated in FIG. 1. In FIG. 1, the multiplexer 52 is set by the control circuit 14 so that the primary controller 16 and the data storage device 24 are active on the fiber channel loop. Fiber channel frames supplied from the host are sent to the primary controller 16, and the primary controller 16 then responds on the loop by returning status information or the like. Frames addressed to the data storage device 24 are passed

DOCKET ALARM
Find authenticated court documents without watermarks at docketalarm.com.

through the PSOC 62 via the cache 66. In the normal operation, the data on the loop is also received by the secondary controller 22 and the data storage device 24. However, the multiplexer 46 is set by the control circuit 14 so that the secondary controller 22 cannot respond on the loop. Since both primary and secondary controllers have the same address, this effectively places the secondary controller 22 in a "wire tap (monitoring)" mode. That is, the secondary controller 22 "listens" to the messages directed to the primary controller 16. Since both primary and secondary controllers 16 and 22 receive the same data, both caches 66 and 82 are filled simultaneously in response to write commands from the host. The data flow is illustrated by the arrows in FIG. 1.

(0010) Synchronization between the PSOCs 62 and 76 is necessary for several reasons. A main reason is to prevent a data over-run state from occurring in the buffers of the PSOCs. The synchronization is accomplished via a communication link 86. When there is space available in a buffer of the PSOC 76, the PSOC 76 sends a request for additional data to the PSOC 62. If the PSOC 62 also has space available in its buffer, the PSOC 62 will notify the host about the additional space that is available. In addition, at the end of a received command, the secondary controller 22 returns a pending status of the command to the primary controller 16. The primary controller 16 will, at the end [sic: misspelling corrected] of the command received by itself, return a pending status of the command to the host. In addition, when a command has been processed by secondary the controller 22 so that the secondary controller 22 is ready to receive additional data, the secondary controller 22 will send a request to the primary controller 16. When the primary controller 16 has also processed the command, it will send a request to the host for additional data.

(0011) FIG. 2 illustrates a fail-over operation in which a failure occurs in the primary controller 16. Such a failure is detected by the fail-over software in each of the primary and secondary controllers 16 and 22. Based on the fail-over detection results, the control circuit 14 sets the multiplexers 52 and 46 so that the primary controller 16 is disabled from responding on the loop and so that the secondary controller 22, which is similarly connected to the data storage device 24, is active on the loop. Since both primary and secondary controllers 16 and 22 have the same address and both have access to the data storage device 24, the host on the loop does not detect the change of the controller. In addition, since the controller 16 is disabled, there is no need to fill its cache 66. The data flow is illustrated by the arrows in FIG. 2.

(0012) FIG. 3 illustrates an alternative preferred embodiment of the invention in which two hosts, host 1 and host 2, are communicating with data storage devices 124 and 125 via a switching circuit set 110 and controllers 116 and 122 on two fiber channel loops. In this embodiment, the switching circuit set 110 is coupled to control circuits 114 and 115 and to the controllers 116 and 122. Each of the controllers 116 and 122 is coupled to both of the data storage devices 124 and 125. The switching circuit set 110 includes two switching circuits 111 and 112, each of which is identical to the switching circuit 12 in FIG. 1. The control circuits 114 and 115 are equivalent to the control circuit 14 in FIG. 1. Further, each of the controllers 116 and 122 is equivalent to the combination of primary and secondary controllers 16 and 22 in FIG. 1. In addition, each of the controllers 116 and 122 includes fail-over software (not shown) such as the Sun Energizer so as to detect whether there is failure in itself, the other controller, or the data storage devices 124 and 125. The fail-over detection results are sent to the control circuits 114 and 115 to control the multiplexers in the switching circuits 111 and 112. In this embodiment, the controller 116 functions as a primary controller (primary 1) for host 1 and a secondary controller (secondary 2) for host 2. Similarly, the controller 122 functions as a primary controller (primary 2) for host 2 and a secondary controller (secondary 1) for host 1. Primary 1 and secondary 1 have the same address, but only one is enabled at a time. Similarly, primary 2 and secondary 2 have the same address, but only one is enabled at a time.

(0013) In the normal operation illustrated in FIG. 3, the multiplexers in the switching circuits 111 and 112 are respectively set by the control circuits 114 and 115 so that each of the controllers 116 and 122 functions only as a primary controller for the respective hosts 1 and 2. In the normal operation, the functions of the secondary controllers in each of the controllers 116 and 122 are disabled by the respective multiplexers. The flow of data is indicated by the arrows in FIG. 3 in a similar manner as in FIG. 1.

(0014) FIG. 4 illustrates a fail-over operation for the embodiment in FIG. 3. If the fail-over software in either of the controllers 116 or 122 detects a failure in one controller – for example, in the controller 122 – the fail-over detection results are sent to the control circuits 114 and 115. In such a case, the multiplexers in the switching circuits 111 and 112 are switched by the control circuits

DOCKET ALARM
Find authenticated court documents without watermarks at docketalarm.com.

DOCKET ALARM

Explore Litigation Insights

Docket Alarm provides insights to develop a more informed litigation strategy and the peace of mind of knowing you're on top of things.

Real-Time Litigation Alerts

Keep your litigation team up-to-date with **real-time alerts** and advanced team management tools built for the enterprise, all while greatly reducing PACER spend.

Our comprehensive service means we can handle Federal, State, and Administrative courts across the country.

Advanced Docket Research

With over 230 million records, Docket Alarm's cloud-native docket research platform finds what other services can't. Coverage includes Federal, State, plus PTAB, TTAB, ITC and NLRB decisions, all in one place.

Identify arguments that have been successful in the past with full text, pinpoint searching. Link to case law cited within any court document via Fastcase.

Analytics At Your Fingertips

Learn what happened the last time a particular judge, opposing counsel or company faced cases similar to yours.

Advanced out-of-the-box PTAB and TTAB analytics are always at your fingertips.

API

Docket Alarm offers a powerful API (application programming interface) to developers that want to integrate case filings into their apps.

LAW FIRMS
Build custom dashboards for your attorneys and clients with live data direct from the court.

Automate many repetitive legal tasks like conflict checks, document management, and marketing.

FINANCIAL INSTITUTIONS
Litigation and bankruptcy checks for companies and debtors.

E-DISCOVERY AND LEGAL VENDORS
Sync your system to PACER to automate legal marketing.