

(19) World Intellectual Property
Organization
International Bureau



(43) International Publication Date
20 January 2005 (20.01.2005)

PCT

(10) International Publication Number
WO 2005/006178 A2

- (51) International Patent Classification⁷: **G06F 7/00**
- (21) International Application Number:
PCT/US2004/019872
- (22) International Filing Date: 18 June 2004 (18.06.2004)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:
10/611,163 30 June 2003 (30.06.2003) US
- (71) Applicant (for all designated States except US): **INTEL CORPORATION** [US/US]; 2200 Mission College Boulevard, Santa Clara, CA 95052 (US).
- (72) Inventors: **CLAYTON, Michele**; 12518 North 136th Place, Scottsdale, AZ 85259 (US). **BROWN, Mark**; 529 East Horeshoe Avenue, Gilbert, AZ 85296 (US). **SKERRY, Brian**; 272 West Temple court, Gilbert, AZ 85233 (US). **CORRADO, Francis**; 11 Sheffield Road, Newton, MA 02460 (US). **WOOTTEN, Jonathan**; 19302 East Via De Palmas, Queen Creekl, AZ 85242 (US).
- (74) Agent: **VINCENT, Lester, J.**; Blakely Sokoloff Taylor & Zafman, 12400 Wilshire Boulevard, 7th Floor, Los Angeles, CA 90025 (US).
- (81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.
- (84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).
- Published:**
— without international search report and to be republished upon receipt of that report
- For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.



WO 2005/006178 A2

(54) Title: INTERMEDIATE STATION

(57) Abstract: In one embodiment, a method is provided that may include determining, at least in part, at an intermediate station one or more locations in storage devices of one or more respective portions of check data and/or other data. The check data may be generated, at least in part, at the intermediate station based at least in part upon the other data. This method also may include determining, at least in part, at the intermediate station, based at least in part upon forwarding information, one or more ports of the intermediate station via which to forward one or more packets. The one or more packets may be generated at the intermediate station based at least in part upon one or more other packets received via one or more other ports of the intermediate station. The other packets may comprise the check data and/or other data.

INTERMEDIATE STATION

FIELD

[0001] This disclosure relates to an intermediate station.

5

BACKGROUND

[0002] In one conventional data storage arrangement, a computer node includes a host processor and a host bus adapter (HBA). The HBA includes ports via which the HBA communicates with data storage devices via respective point-to-point communication links
10 coupled to the ports. In this conventional arrangement, an expander may be used to increase the number of data storage devices with which and/or the number of communication links via which, the HBA may communicate. The expander may permit one-to-many communication between one of the ports of the HBA and a plurality of data storage devices and/or communication links.

[0003] In this conventional arrangement, the data storage devices comprise a redundant array of independent disks (RAID). The host processor and/or HBA perform RAID-related operations (e.g., operations involved in maintaining and/or implementing the RAID). Unfortunately, this may use undesirably large amounts host processor and/or HBA processing resources to carry out such operations.

20

BRIEF DESCRIPTION OF THE DRAWINGS

[0004] Features and advantages of embodiments of the claimed subject matter will become apparent as the following Detailed Description proceeds, and upon reference to the Drawings, wherein like numerals depict like parts, and in which:

[0005] Figure 1 is diagram that illustrates a system embodiment.

[0006] Figure 2 is a flowchart that illustrates operations that may be performed according to an embodiment.

[0007] Although the following Detailed Description will proceed with reference being made to illustrative embodiments of the claimed subject matter, many alternatives, modifications, and variations thereof will be apparent to those skilled in the art. Accordingly, it is intended that the claimed subject matter be viewed broadly, and be defined only as set forth in the accompanying claims.

DETAILED DESCRIPTION

10

[0008] Figure 1 illustrates a system embodiment 100. System 100 may include a host processor 12 coupled to a chipset 14. Host processor 12 may comprise, for example, an Intel® Pentium® IV microprocessor that is commercially available from the Assignee of the subject application. Of course, alternatively, host processor 12 may comprise another type of microprocessor, such as, for example, a microprocessor that is manufactured and/or commercially available from a source other than the Assignee of the subject application, without departing from this embodiment.

15

[0009] Chipset 14 may comprise a host bridge/hub system that may couple host processor 12, a system memory 21 and a user interface system 16 to each other and to a bus system 22. Chipset 14 may also include an input/output (I/O) bridge/hub system (not shown) that may couple the host bridge/bus system to bus 22. Chipset 14 may comprise one or more integrated circuit chips, such as those selected from integrated circuit chipsets commercially available from the assignee of the subject application (e.g., graphics memory and I/O controller hub chipsets), although one or more other integrated circuit chips may also, or alternatively be used, without departing from this embodiment. User

20

25

interface system 16 may comprise, e.g., a keyboard, pointing device, and display system that may permit a human user to input commands to, and monitor the operation of, system 100.

[0010] Bus 22 may comprise a bus that complies with the Peripheral Component Interconnect (PCI) Express™ Base Specification Revision 1.0, published July 22, 2002, available from the PCI Special Interest Group, Portland, Oregon, U.S.A. (hereinafter referred to as a “PCI Express™ bus”). Alternatively, bus 22 instead may comprise a bus that complies with the PCI-X Specification Rev. 1.0a, July 24, 2000, available from the aforesaid PCI Special Interest Group, Portland, Oregon, U.S.A. (hereinafter referred to as a “PCI-X bus”). Also alternatively, bus 22 may comprise other types and configurations of bus systems, without departing from this embodiment.

[0011] System embodiment 100 may comprise storage devices 28A and 27. Storage devices 28A may comprise one or more mass storage devices communicatively coupled to an I/O controller circuit card 20 via one or more communication media 44A. As used herein, a “storage device” means an apparatus or medium into, and from which, data and/or commands may be stored and retrieved, respectively. Also as used herein, a “mass storage device” means a storage device that is capable of non-volatile storage of data and/or commands, and, for example, may include, without limitation, magnetic, optical, and/or semiconductor storage devices. In this embodiment, card 20 may comprise, for example, an HBA. Storage devices 27 may comprise a plurality of storage devices 28B . . . 28N that may be communicatively coupled to an intermediate station 40 via one or more respective communication media 44C . . . 44N. Each of the storage devices 28B . . . 28N may comprise a respective mass storage device. Intermediate station 40 may be communicatively coupled to card 20 via one or more communication media 44B. Of

course, the number of storage devices 28A and 27, and the number of communication media 44A, 44B, 44C . . . 44N may vary without departing from this embodiment.

[0012] As used herein, a “communication medium” means a physical entity through which electromagnetic radiation may be transmitted and/or received. Also as used herein, an
5 “intermediate station” means a first node in a network that is capable of forwarding, in response at least in part to receipt by the first node of one or more packets from a second node in the network, one or more other packets to a third node in the network. Also as used herein, a “packet” means a sequence of one or more signals that encode one or more symbols and/or values. As used herein, “forwarding” one or more packets by, from, or via
10 a first device means transmitting by, from, or via, respectively, the first device the one or more packets to a second device.

[0013] In this embodiment, storage devices 27 may comprise one or more redundant arrays of independent disks (RAID) 29 that may be implemented using mass storage devices 28B . . .
15 . 28N. The RAID level that may be implemented by RAID 29 may be 0, 1, or greater than 1. Depending upon, for example, the RAID level implemented in RAID 29, the number of mass storage devices 28B . . . 28N comprised in RAID 29 may vary so as to permit the number of mass storage devices 28A . . . 28N to be at least sufficient to implement the RAID level implemented in RAID 29.

[0014] Processor 12, system memory 21, chipset 14, bus 22, and circuit card slot 30 may be
20 comprised in a single circuit board, such as, for example, a system motherboard 32. A host computer system 110 may comprise system motherboard 32.

[0015] In this embodiment, card 20 may exchange data and/or commands with one or more mass storage devices 28A and intermediate station 40, via one or more media 44A and one
25 or more media 44B, respectively, using, e.g., Serial Advanced Technology Attachment (S-ATA) protocol and/or Serial Attached Small Computer Systems Interface (SAS) protocol.

Of course, alternatively, I/O controller card 20 may exchange data and/or commands with one or more mass storage devices 28A and intermediate station 40 using other and/or additional communication protocols, without departing from this embodiment.

[0016] In this embodiment, intermediate station 40 may exchange data and/or commands with storage devices 27, via communication media 44C . . . 44N, using, e.g. an S-ATA and/or SAS protocol. Of course, alternatively, station 40 may exchange data and/or commands with devices 27 using other and/or additional communication protocols, without departing from this embodiment.

[0017] In accordance with this embodiment, if an S-ATA protocol is used by controller card 20 to exchange data and/or commands with one or more mass storage devices 28A and intermediate station 40, and/or if an S-ATA protocol is used by intermediate station 40 to exchange data and/or commands with mass storage devices 27, the S-ATA protocol may comply or be compatible with the protocol described in "Serial ATA: High Speed Serialized AT Attachment," Revision 1.0, published on August 29, 2001 by the Serial ATA Working Group. If an SAS protocol is used by controller card 20 to exchange data and/or commands with one or more mass storage devices 28A and intermediate station 40, and/or if an SAS protocol is used by intermediate station 40 to exchange data and/or commands with mass storage devices 27, the SAS protocol may comply or be compatible with the protocol described in "Information Technology - Serial Attached SCSI (SAS)," Working Draft American National Standard of International Committee For Information Technology Standards (INCITS) T10 Technical Committee, Project T10/1562-D, Revision 2b, published 19 October 2002, by American National Standards Institute (hereinafter termed the "SAS Standard") and/or later-published versions of the SAS Standard.

[0018] Depending upon, for example, whether bus 22 comprises a PCI Express™ bus or a PCI-X bus, circuit card slot 30 may comprise, for example, a PCI Express™ or PCI-X bus compatible or compliant expansion slot or interface 36. Interface 36 may comprise a bus connector 37 may be electrically and mechanically mated with a mating bus connector 34
5 that may be comprised in a bus expansion slot or interface 35 in circuit card 20.

[0019] As used herein, “circuitry” may comprise, for example, singly or in any combination, hardwired circuitry, programmable circuitry, state machine circuitry, and/or memory that may comprise program instructions that may be executed by programmable circuitry. In this embodiment, circuit card 20 may comprise operative circuitry 38 which may comprise
10 computer-readable memory 39. Memory 39 may comprise one or more of the following types of memories: semiconductor firmware memory, programmable memory, non-volatile memory, read only memory, electrically programmable memory, random access memory, flash memory, magnetic disk memory, and/or optical disk memory. Either additionally or alternatively, memory 39 may comprise other and/or later-developed types
15 of computer-readable memory.

[0020] Machine-readable firmware program instructions may be stored in memory 39. These instructions may be accessed and executed by operative circuitry 38. When executed by circuitry 38, these instructions may result in card 20 and/or circuitry 38 performing the operations described herein as being performed by card 20 and/or circuitry 38.

[0021] Slot 30 and card 20 are constructed to permit card 20 to be inserted into slot 30. When card 20 is properly inserted into slot 30, connectors 34 and 36 become electrically and mechanically coupled to each other. When connectors 34 and 36 are so coupled to each other, card 20 becomes electrically coupled to bus 22 and may exchange data and/or

commands with system memory 21, host processor 12, and/or user interface system 16 via bus 22 and chipset 14.

[0022] Alternatively, without departing from this embodiment, operative circuitry 38 may not be comprised in card 20, but instead, may be comprised in other structures, systems, and/or devices. These other structures, systems, and/or devices may be, for example, comprised in motherboard 32, coupled to bus 22, and exchange data and/or commands with other components (such as, for example, system memory 21, host processor 12, and/or user interface system 16) in system 100.

[0023] In this embodiment, intermediate station 40 may comprise a standalone node coupled to, and geographically displaced from, at least in part, host system 110 and RAID 29. In this embodiment, station 40 may comprise operative circuitry 50 that may comprise, for example, packet forwarding mechanism 51. In this embodiment, mechanism 51 may comprise, for example, an expander 52 that, in this embodiment, unless stated to the contrary herein, may comply and/or be compatible with the definition of, and perform operations performed by a SAS expander, for example, as provided in the SAS Standard. Expander 52 may comprise one or more I/O ports 54, one or more I/O ports 56, forwarding/routing circuitry 60, and computer-readable memory 64. Memory 64 may comprise one or more of the following types of memories: semiconductor firmware memory, programmable memory, non-volatile memory, read only memory, electrically programmable memory, random access memory, flash memory, magnetic disk memory, and/or optical disk memory. Either additionally or alternatively, memory 64 may comprise other and/or later-developed types of computer-readable memory. One or more I/O ports 54 may be coupled via one or more media 44B to card 20. One or more I/O ports 56 may comprise a plurality of I/O ports that may be respectively coupled via media

44C . . . 44N to devices 28B . . . 28N. Circuitry 60 may be coupled to one or more ports 54, memory 64, and ports 56.

[0024] Alternatively, without departing from this embodiment, operative circuitry 50 may not be comprised in a standalone intermediate station that is geographically displaced, at least
5 in part, from host system 110, but instead, may be comprised in other structures, systems, and/or devices that may be, for example, comprised in motherboard 32 and coupled to bus 22, and exchange data and/or commands with other components (such as, for example, system memory 21, host processor 12, and/or user interface system 16) in system 110. Further alternatively, without departing from this embodiment, these other structures,
10 systems, and/or devices may be, for example, comprised in another circuit card (not shown) that may be coupled to bus 22 via another circuit card slot (not shown) that may be comprised in host system 110.

[0025] Memory 64 may store, for example, machine executable program instructions 66 and packet forwarding and/or routing tables 69. Tables 69 may comprise forwarding
15 information 68. As used herein, "forwarding information" means data and/or one or more parameters based upon which, at least in part, a determination or selection of one or more devices via which to forward one or more packets may be made, as least in part.

Instructions 66, and/or portions thereof, may be retrieved from memory 64 and executed by circuitry 60, circuitry 62, one or more ports 54, and/or ports 56. The execution of
20 instructions 66 and/or portions thereof by circuitry 60, circuitry 62, one or more ports 54, and/or ports 56 may result in circuitry 60, circuitry 62, one or more ports 54, and/or ports 56 carrying out the functions and/or operations described herein as being carried out by circuitry 60, circuitry 62, one or more ports 54, and/or ports 56, respectively.

[0026] As is described in below, circuitry 60 may be capable of maintaining in, and retrieving
25 from memory 64 forwarding information 68 and/or tables 69. Based at least in upon the

forwarding information 68 and/or tables 69, circuitry 60 may be capable of determining, at least in part, the manner in which one or more packets may be forwarded by expander 52. Additionally, as is also described below, circuitry 62 may be capable of performing one or more RAID-related operations that may use, for example, as operands, check data and/or other data (e.g., data from which the check data has been and/or may be generated, at least in part) comprised in one or more packets received by expander 54 from card 20 and/or RAID 29. The check data and/or other data generated as a result, at least in part, of the one or more RAID-related operations, may be comprised in one or more other packets that may be generated and forwarded by expander 52 to RAID 29 or card 20, respectively. As used herein, "check data" means first data generated based at least in part upon second data and from which the second data may be regenerated at least in part. In this embodiment, check data may comprise RAID parity data. Circuitry 62 may comprise, for example, arithmetic circuitry (not shown) that may be capable of performing one or more arithmetic and/or logical operations using and/or involving check data and/or other data from which the check data has been and/or may be generated. These one or more arithmetic and/or logical operations may comprise, for example, logical exclusive-or operations that may generate RAID parity data from initial user data, and/or that may regenerate the initial user data from such RAID parity data.

[0027] Alternatively, although not shown in the Figures, one or more portions of forwarding information 68 and/or tables 69 may be stored and/or maintained elsewhere in system 100, such as, for example, in memory 39 and/or memory 21. In this alternative arrangement, circuitry 60 may be capable of providing one or more commands, for example, to card 20 and/or memory 21 that may result in the retrieving of these one or more portion of information 68 and/or tables 69 from memory 39 and/or memory 21 and the providing of these one or more portions of information 68 and/or tables 69 to circuitry 60.

[0028] For reasons described below, circuitry 60 and/or circuitry 62 may be capable of determining, in accordance, at least in part with conventional RAID techniques, one or more respective locations in mass storage devices 27 in RAID 29 in which to store and/or from which may be retrieved one or more respective portions of check data and/or user data from which the check data has been generated, at least in part, by circuitry 62, based at least in part upon the user data. For example, if the RAID level implemented in RAID 29 is equal to zero, the check data may be absent from RAID 29, and these locations may be comprise, for example, addresses and/or block locations in devices 27 of one or more stripes (not shown) of user data. If the RAID level implemented in RAID 29 is greater than one, the check data may comprise parity data and these locations may comprise, for example, addresses and/or block locations in devices 27 of one or more stripes (not shown) of such check data and/or user data. Alternatively, if the RAID level implemented in RAID 29 is equal to one (i.e., RAID 29 implements data mirroring), the check data may comprise a copy of the user data and these locations may comprise, for example, addresses and/or block locations of the user data and the redundant copy of the user data in respective mirrored volumes (not shown) in RAID 29.

[0029] Although all of the storage devices 27 are shown in Figure 1 as being comprised in RAID 29, without departing from this embodiment, a subset of storage devices 27 may not be comprised in RAID 29.

[0030] With reference now being made to Figure 2, operations 200 will be described that may be performed in accordance with an embodiment. After, for example, a reset of system 100, card 20 may transmit to intermediate station 40, via one or more media 44B, one or more packets 70 that may comprise one or more requests 72 to store user data 74 comprised in the one or more packets 70 in RAID 29. One or more ports 54 may receive one or more packets 70, and may provide them to circuitry 60.

[0031] In response, at least in part, to the receipt by intermediate station 40 of one or more packets 70, circuitry 60 at intermediate station 40 may determine, at least in part, one or more respective locations (e.g., locations 106 and 108) in plurality of storage devices 27 to store one or more respective portions (e.g., portions 107 and 109) of check data and/or other data (e.g., user data) based, at least in part, upon which circuitry 62 at intermediate station 40 may generate, at least in part, the check data, as illustrated by operation 202 in Figure 2. Also in response, at least in part to the receipt by intermediate station 40 of one or more packets 70, circuitry 60 at intermediate station 40 may determine, at least in part, based, at least in part, upon forwarding information 68 stored, at least in part, at the intermediate station, one or more ports comprised in ports 56 via which to forward to storage devices (e.g., storage devices 28B and 28N) in the plurality of storage devices 27 one or more other packets (e.g., packets 88 and 96) that may be generated by circuitry 60, based at least in part upon one or more packets 70, as illustrated by operation 204 in Figure 2. Packets 88 and 96 may comprise respective portions (e.g., stripes) of the check data and/or other data (e.g., user data), collectively or singly referred to by 90 and 98 in Figure 1), to be stored at locations 106 and 108 in storage devices 27. Of course, depending upon the RAID level implemented in RAID 29, the number and respective contents of packets forwarded via ports 56 to storage devices 27, as well as, the number of storage devices to which such packets may be propagated, and the number of portions of check data and/or user data that may be stored in devices 27, may vary, without departing from this embodiment, in order to allow said RAID level to be implemented. As used herein, a "portion" of data may comprise some or all of the data, or a copy of some or all of the data.

[0032] In this embodiment, as part of operations 202 and/or 204, circuitry 60 may examine header information, such as, for example, destination addresses, that may be comprised

one or more requests 72. Also as part of operations 202 and/or 204, circuitry 60 may correlate this header information with forwarding information 68 to determine which of the ports 56 via which to forward packets 88 and 96, via media 44C and 44N, to storage devices 28B and 28N. For example, in this embodiment, forwarding information 68 may
5 comprise one or more tables (not shown) that may correlate the destination addresses that may be expected to be comprised in header information in one or more requests 72, with locations (e.g., addresses and/or blocks) in each of the storage devices 27 to which to store portions of user data that may be comprised in one or more packets 70 and/or portions of check data that may be generated by circuitry 62, based at least in part upon the user data,
10 in order to permit one or more requests 72 to be satisfied while maintaining RAID 29. These one or more tables comprised in forwarding information 68 may also correlate these locations in storage devices 27 with ports in ports 56 via which the storage devices that comprise these locations may be coupled to expander 52. Additionally, forwarding information 68 may correlate with each of these locations in storage 27 the type of data
15 (e.g., check data, or user data from which the one or more portions of check data may be generated, at least in part) that may be stored at these locations, in order to permit RAID 29 to be maintained and/or implemented.

[0033] As part of operations 202 and/or 204, circuitry 62 may generate, based at least in part upon user data 74 and conventional RAID techniques associated with the RAID level
20 implemented by RAID 29, one or more respective portions of user data 74 and/or check data to be stored in storage devices 27 that may satisfy one or more requests 70 while maintaining RAID 29. In this embodiment, as part of operation 202, based at least in part upon forwarding information 68 and the header information comprised in one or more requests 72, circuitry 60 may determine, at least in part, one or more respective locations
25 106 and 108 in storage devices 27 to store one or more respective portions 90 and 98 in

order to satisfy one or more requests 72 while maintaining RAID 29. Also in this embodiment, as part of operation 204, based at least in part upon forwarding information 68 and the header information comprised in one or more requests 72, circuitry 60 may determine, at least in part, ports comprised in ports 56 via which packets 88 and 96 may be
5 forwarded to storage devices 28B and 28N, in order to satisfy one or more requests 72 while maintaining RAID 29.

[0034] After executing operations 202 and 204, circuitry 60 may generate packets 88 and 96 such that each of the respective packets 88 and 96 may comprise a respective one of these one or more respective portions of user data 74 and/or check data generated by circuitry
10 62. For example, in this embodiment, if RAID 29 implements mirroring, one or more packets 88 may comprise user data 90 that may be identical to user data 74, and one or more packets 96 may comprise a copy 98 of user data 90. One or more packets 88 may also comprise addressing information that may address one or more locations 106 determined by circuitry 60 as a result of operation 202. Additionally, one or more packets
15 96 also may comprise addressing information that may address one or more locations 108 determined by circuitry 60 as a result of operation 202.

[0035] In this embodiment, after generating packets 88 and 96, circuitry 60 transmit packets 88 and 96 to the ports comprised in ports 56, determined by circuitry 60 as a result of operation 204, via which packets 88 and 96 may be forwarded to storage devices 28B and
20 28N in order to satisfy one or more requests 72 and maintain RAID 29. These ports in ports 56 may forward one or more packets 88 to storage device 28B via one or more media 44C, and may forward one or more packets 96 to storage device 28N via one or more media 44N.

[0036] After storage device 28B receives one or more packets 88, device 28B may store copy
25 107 of one or more respective portions 90 at one or more locations 106 addressed by the

addressing information in one or more packets 88. Additionally, after storage device 28N receives one or more packets 96, device 28N may store copy 109 of one or more portions 98 at one or more locations 108 addressed by the addressing information in one or more packets 96.

[0037] Alternatively or additionally, card 20 may transmit to intermediate station 40 one or more packets (not shown) that may comprise one or more retrieval requests (not shown) to retrieve user data 74 from RAID 29. One or more ports 54 may receive these one or more requests, and may provide them to circuitry 60.

[0038] In response, at least in part, to the receipt by intermediate station 40 of these one or
10 more retrieval requests, circuitry 60 may determine (e.g., as a result of operation 202), at least in part, based at least in part upon forwarding information 68 and header information comprised in the one or more packets that comprise the one or more requests, one or more
15 respective locations (e.g., locations 106 and 108) in storage devices 27 from which to retrieve one or more respective portions (e.g., portions 107 and 109) of user data 74,
and/or check data generated, at least in part, by circuitry 62 based, at least in part, upon
user data 74. Also in response, at least in part, to the receipt by intermediate station 40 of
these one or more requests, circuitry 60 at intermediate station 40 may determine, at least
in part, based, at least in part, upon forwarding information 68 stored, at least in part, at
intermediate station 40 and header information comprised in the one or more packets that
20 comprise the one or more requests, one or more ports comprised in ports 56 via which to forward to one or more of the storage devices 27 one or more additional packets that may be generated by circuitry 60, based at least in part upon the one or more requests. These one or more additional packets may request that storage devices 27 retrieve these one or more respective portions of user data 74, and/or check data that was generated based at
25 least in part upon user data 74, and forward them to intermediate station 40. Of course,

depending upon the RAID level implemented in RAID 29, the number of additional packets forwarded via ports 56 to storage devices 27, as well as, the number of storage devices to which such additional packets may be propagated, and the number of portions of check data and/or user data to be retrieved from devices 27, may vary, without

5 departing from this embodiment, in order to allow said RAID level to be implemented.

[0039] In this embodiment, as part of operation 202, based at least in part upon forwarding information 68 and the header information comprised in the one or more packets that comprise the one or more retrieval requests, circuitry 60 may transmit these one or more additional packets to one or more of the storage devices 27 via one or more of the ports 56
10 and media 44C . . . 44N. In this embodiment, this may result in storage device 28B retrieving one or more portions 107 from one or more locations 106, and transmitting to intermediate station 40 one or more packets 92 that comprise a copy 94 of one or more portions 107. Alternatively, in this embodiment, this may result in storage device 28N retrieving one or more portions 109 from one or more locations 108, and transmitting to
15 intermediate station 40 one or more packets 102 that comprise a copy 104 of one or more portions 109.

[0040] Either the case of either alternative, one or more ports 56 may receive one or more packets 92 or 102, and may transmit them to circuitry 60. One or more packets 92 and 102 may comprise header information that may comprise one or more destination
20 addresses that may indicate that one or more packets 92 and 102 are intended to be received by host system 110 and/or card 20. The one or more tables comprised in forwarding information 68 may correlate these one or more destination addresses with one or more ports 54 that are coupled to card 20. Circuitry 60 may examine forwarding information 68, and as a result of operation 204, may determine, based at least in part upon
25 forwarding information 68, to forward one or more packets 76 to card 20 via the one or

more ports 54 that are correlated in forwarding information 68 with these one or more destination addresses. Circuitry 60 may generate one or more packets 76, based at least in part upon packets 92 or 102 previously received by one or more ports 56, such that one or more packets 76 may comprise a copy 80 of the requested data 74 obtained from packets 92 or 102. Circuitry 62 may forward one or more packets 76 to card 20 via one or more ports 54.

[0041] Alternatively, depending upon the RAID level implemented in RAID 29, if the copy in storage 27 of user data 74 has become corrupted, one or more packets 92 and 102 retrieved from storage 27 may comprise copies 94 and 104 of check data generated based at least in part upon user data 74 and stored in storage 27. In this alternative, after circuitry 60 receives one or more packets 92 and 102, circuitry 62 may regenerate, at least in part, user data 74, based at least in part upon the copies 94 and 104 of the check data from packets 92 and 102 and conventional RAID techniques associated with the RAID level implemented in RAID 29. Circuitry 60 may examine forwarding information 68, and as a result of operation 204, may determine, based at least in part upon forwarding information 68, to forward one or more packets 76 to card 20 via the one or more ports 54. Circuitry 60 may generate one or more packets 76 such that one or more packets 76 may comprise a copy 80 of the requested data 74 regenerated, at least in part, by circuitry 62 based at least in part upon the copies 94 and 104 of the check data from packets 92 and 102. Circuitry 62 may forward one or more packets 76 to card 20 via one or more ports 54.

[0042] Card 20 also may be capable of generating and transmitting to intermediate station 40 a plurality of Serial Management Protocol (SMP) messages (collectively and singly referred to herein by the numeral 82 in Figure 1). For example, one or more of these messages 82 may comprise header information 84 and one or more requests 86 to modify one or more operational parameters of RAID 29, such as, for example, partitioning of

RAID 29 into a plurality of redundant arrays of independent disks, collapsing a plurality of such arrays into a single RAID, changing assignment of storage devices to such arrays, changing one or more RAID levels implemented by such arrays, etc. In this embodiment, one more messages may comprise, for example, one or more SMP_REQUEST messages, and one or more requests 86 may be indicated by one or more values in one or more special fields in one or more messages 82. These one or more special fields may be chosen by a vendor and/or maker of intermediate station 40 so as to be, for example, uniquely associated with SMP messages intended to be received by one or more SAS expanders vended and/or made by such vendor and/or maker, while allowing such SMP messages to be compatible and/or in compliance with the SAS Standard.

[0043] At least one of one or more ports 54 may receive these one or more messages 82, as illustrated by operation 206 in Figure 2. In response, at least in part, to receipt of one or more messages 82, one or more ports 52 may transmit one or more messages 82 to circuitry 60. This may result in circuitry 60 generating and issuing to RAID 29 one or more packets (not shown) that may result in RAID 29 modifying, in accordance with conventional RAID techniques, the one or more operating parameters of RAID 29 so as to carry out the modification to these one or more operating parameters requested in one or more requests 86. Circuitry 60 may also modify forwarding information 68 and/or tables 69 so as to cause them to comport with the one or more modified operating parameters of RAID 29.

[0044] Alternatively or additionally, one or more others of these messages 82 may comprise header information 84 and one or more requests 86 to modify forwarding information 68 and/or tables 69. At least one of one or more ports 54 may receive these one or more messages 82, as illustrated by operation 208 in Figure 2. In response, at least in part, to receipt of one or more messages 82, one or more ports 52 may transmit one or more

messages 82 to circuitry 60. This may result in circuitry 60 modifying forwarding information 68 and/or tables 69 so as to cause them to comport with the one or more requests 86.

[0045] Thus, one system embodiment comprises circuitry to determine, at least in part, at an
5 intermediate station one or more respective locations in a plurality of storage devices of one or more respective portions of check data and/or other data (e.g., user data). In this system embodiment, the check data may be generated, at least in part, at the intermediate station based at least in part upon the other data. The circuitry also may be capable of determining, at least in part, at the intermediate station, based at least in part upon
10 forwarding information stored at least in part at the intermediate station, one or more ports of the intermediate station via which to forward one or more packets. The one or more packets may be generated at the intermediate station based at least in part upon one or more other packets previously received via one or more other ports of the intermediate station. The one or more other packets may comprise the check data and/or other data.
15 The circuit card may be communicatively coupled to the intermediate station via a communication medium.

[0046] Advantageously, the intermediate station comprised in this system embodiment may perform some or all of the RAID-related operations that, in the prior art, may have been performed by a host processor and/or HBA. Advantageously, this may reduce the
20 amounts host processor and/or HBA processing resources used to carry out such operations in this system embodiment, compared to the prior art. Further advantageously, the intermediate station of this system embodiment also may provide a relatively easy and transparent (from the standpoint of the host system) way in which to add RAID functionality to a storage network having storage devices, especially if such storage
25 devices are capable of communicating using SAS and/or S-ATA protocol.

[0047] The terms and expressions which have been employed herein are used as terms of description and not of limitation, and there is no intention, in the use of such terms and expressions, of excluding any equivalents of the features shown and described (or portions thereof), and it is recognized that various modifications are possible within the scope of
5 the claims.

[0048] For example, instead of or in addition to comprising expander 52, intermediate station 40 may comprise a port multiplier that complies and/or is compatible with an S-ATA protocol. Host system 110 may be capable of providing one or more commands (e.g., one or more commands specific to and/or associated with a vendor and/or maker of the port
10 multiplier) that may permit forwarding and/or routing tables and/or management information to be stored and/or modified in the port multiplier. By so storing and/or modifying such tables and/or information, host system 110 may manage and/or control operation of the port multiplier.

[0049] Additionally or alternatively, expander 52 may be capable of detecting when a
15 configuration change occurs in devices 27, and when expander 52 detects the occurrence of such a configuration change, expander 52 may provide one or more messages to host computer system 110 that may indicate the occurrence and nature of the change. Examples of the types of configuration changes that may be detectable by expander 52 may include changes in the configuration of RAID 29, such as failure and/or removal of
20 one or more of mass storage devices 28B . . . 28N, and/or coupling of one or more other and/or additional mass storage devices (not shown) into RAID 29 (e.g., in a "hot plug" of one or more mass storage devices into RAID 29).

[0050] Additional modifications are also possible. Accordingly, the claims are intended to cover all such equivalents.

What is claimed is:

CLAIMS

- 5 1. A method comprising:
determining, at least in part, at an intermediate station one or more respective
locations in a plurality of storage devices of one or more respective portions of at least one
of check data and other data, the check data being generated, at least in part, at the
intermediate station based at least in part upon the other data;
- 10 determining, at least in part, at the intermediate station, based at least in part upon
forwarding information stored at least in part at the intermediate station, one or more ports
of the intermediate station via which to forward one or more packets, the one or more
packets being generated at the intermediate station based at least in part upon one or more
other packets previously received via one or more other ports of the intermediate station,
- 15 the one or more other packets comprising at least one of the check data and the other data.
2. The method of claim 1, wherein:
the intermediate station comprises a Serial Attached Small Computer Systems
Interface (SAS) expander.
3. The method of claim 1, wherein:
20 the plurality of storage devices comprises a redundant array of independent disks
(RAID); and
the one or more other ports comprises a plurality of ports coupled to the RAID.
4. The method of claim 3, wherein:
the one or more ports comprises at least one port coupled to a host computer
- 25 system.
5. The method of claim 1, further comprising:

receiving at the intermediate station via the at least one port a Serial Management Protocol (SMP) message that comprises one or more requests to modify an operational parameter of the RAID.

6. The method of claim 5, further comprising:

5 receiving at the intermediate station via the at least one port another SMP message that comprises one or more other requests to modify the forwarding information.

7. The method of claim 1, wherein:

one or more routing tables comprise the forwarding information.

8. The method of claim 1, wherein:

10 the one or more respective locations comprise a plurality of locations; and each of the storage devices comprises one or more of the plurality of locations.

9. An apparatus comprising:

circuitry to determine, at least in part, at an intermediate station one or more respective locations in a plurality of storage devices of one or more respective portions of at least one of check data and other data, the check data being generated, at least in part, at the intermediate station based at least in part upon the other data;

the circuitry also being capable of determining, at least in part, at the intermediate station, based at least in part upon forwarding information stored at least in part at the intermediate station, one or more ports of the intermediate station via which to forward one or more packets, the one or more packets being generated at the intermediate station based at least in part upon one or more other packets previously received via one or more other ports of the intermediate station, the one or more other packets comprising at least one of the check data and the other data.

10. The apparatus of claim 9, wherein:

25 the circuitry is comprised, at least in part, at the intermediate station; and

the intermediate station comprises a Serial Attached Small Computer Systems Interface (SAS) expander.

11. The apparatus of claim 9, wherein:

the plurality of storage devices comprises a redundant array of independent disks
5 (RAID); and

the one or more other ports comprises a plurality of ports coupled to the RAID.

12. The apparatus of claim 11, wherein:

the one or more ports comprises at least one port coupled to a host.

13. The apparatus of claim 12, wherein:

10 the intermediate station is also capable of receiving via the at least one port a Serial Management Protocol (SMP) message that comprises one or more requests to modify an operational parameter of the RAID.

14. The apparatus of claim 13, wherein:

the intermediate station is also capable of receiving via the at least one port another
15 SMP message that comprises one or more other requests to modify the forwarding information.

15. The apparatus of claim 9, wherein:

one or more routing tables comprise the forwarding information.

16. The apparatus of claim 9, wherein:

20 the one or more respective locations comprise a plurality of locations; and each of the storage devices comprises one or more of the plurality of locations.

17. An article comprising:

a storage medium having stored thereon instructions that when executed by a machine result in the following:

determining, at least in part, at an intermediate station one or more respective locations in a plurality of storage devices of one or more respective portions of at least one of check data and other data, the check data being generated, at least in part, at the intermediate station based at least in part upon the other data;

5 determining, at least in part, at the intermediate station, based at least in part upon forwarding information stored at least in part at the intermediate station, one or more ports of the intermediate station via which to forward one or more packets, the one or more packets being generated at the intermediate station based at least in part upon one or more other packets previously received via one or more other ports of the intermediate station,
10 the one or more other packets comprising at least one of the check data and the other data.

18. The article of claim 17, wherein:

the intermediate station comprises a Serial Attached Small Computer Systems Interface (SAS) expander.

19. The article of claim 17, wherein:

15 the plurality of storage devices comprises a redundant array of independent disks (RAID); and

the one or more other ports comprises a plurality of ports coupled to the RAID.

20. The article of claim 19, wherein:

20 the one or more ports comprises at least one port coupled to a host computer system.

21. The article of claim 20, wherein the instructions when executed by the machine also result in:

receiving at the intermediate station via the at least one port a Serial Management Protocol (SMP) message that comprises one or more requests to modify an operational
25 parameter of the RAID.

22. The article of claim 21, wherein the instructions when executed by the machine also result in:

receiving at the intermediate station via the at least one port another SMP message that comprises one or more other requests to modify the forwarding information.

5 23. The article of claim 17, wherein:

one or more routing tables comprise the forwarding information.

24. The article of claim 17, wherein:

the one or more respective locations comprise a plurality of locations; and each of the storage devices comprises one or more of the plurality of locations.

10 25. A system comprising:

circuitry to determine, at least in part, at an intermediate station one or more respective locations in a plurality of storage devices of one or more respective portions of at least one of check data and other data, the check data being generated, at least in part, at the intermediate station based at least in part upon the other data;

15 the circuitry also being capable of determining, at least in part, at the intermediate station, based at least in part upon forwarding information stored at least in part at the intermediate station, one or more ports of the intermediate station via which to forward one or more packets, the one or more packets being generated at the intermediate station based at least in part upon one or more other packets previously received via one or more
20 other ports of the intermediate station, the one or more other packets comprising at least one of the check data and the other data; and

a circuit card communicatively coupled to the intermediate station via a communication medium.

26. The system of claim 25, further comprising:

a circuit board comprising a bus and a bus interface slot, the circuit card being capable of being coupled to the bus interface slot.

27. The system of claim 26, wherein:

the circuit board comprises a processor that is coupled to the bus.

5 28. The system of claim 25, wherein:

the check data comprises one of:

parity data generated based at least in part upon the other data; and

a copy of the other data.

29. The system of claim 25, wherein:

10 the circuit card is coupled to one or more other storage devices that are independent of the plurality of storage devices.

30. The system of claim 25, wherein:

the plurality of storage devices comprises a redundant array of independent disks (RAID) that implements a RAID level greater than 1; and

15 the check data comprises parity data generated based at least in part upon the other data.

31. The method of claim 1, wherein:

the intermediate station comprises a Serial Advanced Technology Attachment port multiplier.

20 32. The apparatus of claim 9, wherein:

the circuitry is comprised, at least in part, at the intermediate station; and

the intermediate station comprises a Serial Advanced Technology Attachment port multiplier.

33. The article of claim 17, wherein:

the intermediate station comprises a Serial Advanced Technology Attachment port multiplier.

34. The method of claim 1, further comprising:

providing one or more messages from the intermediate station to a host computer
5 system that indicate that a configuration change has occurred in the plurality of storage devices.

35. The apparatus of claim 9, wherein:

the circuitry is also capable of providing to a host computer system one or more
messages that indicate that a configuration change has occurred in the plurality of storage
10 devices.

36. The article of claim 17, wherein the instructions when executed by the machine also result in:

providing one or more messages from the intermediate station to a host computer
system that indicate that a configuration change has occurred in the plurality of storage
15 devices.

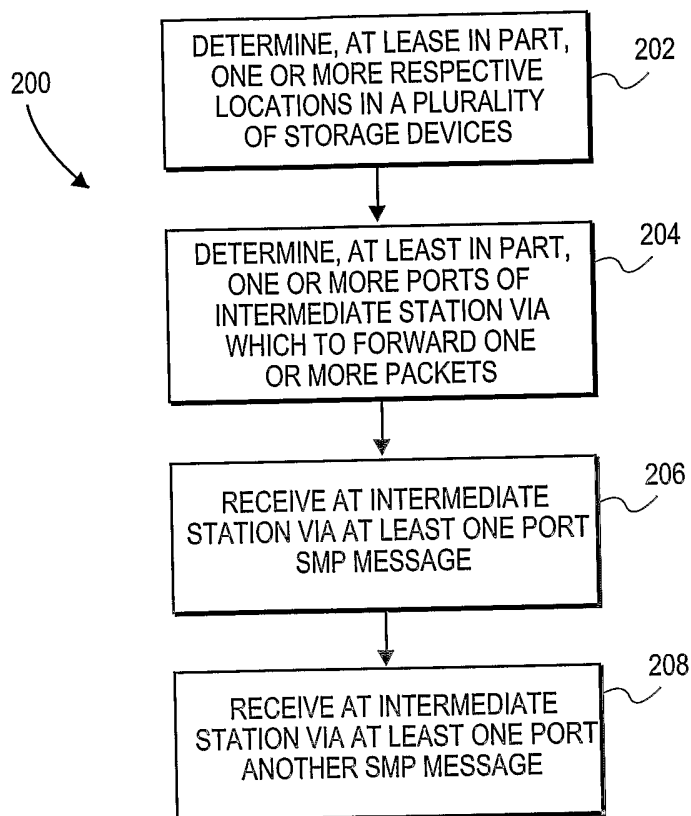


FIG. 2