

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
19 April 2001 (19.04.2001)

PCT

(10) International Publication Number
WO 01/28179 A2

(51) International Patent Classification⁷: **H04L 29/00**

3PQ (GB). **MEYER, Jonathan**; 2A White Rose Lane, Farnham, Surrey GU10 3NG (GB). **ASTON, Christopher, J.**; 40 The Crescent, Wycombe, Buckinghamshire HP13 6JR (GB). **WINFIELD, John**; 126 Elgar Road, Reading, Berkshire RG2 0BN (GB).

(21) International Application Number: PCT/EP00/10277

(22) International Filing Date: 12 October 2000 (12.10.2000)

(25) Filing Language: English

(74) Agents: **FROUD, Clive** et al.; Elkington and Fife, Prospect House, 8 Pembroke Road, Sevenoaks, Kent TN13 1XR (GB).

(26) Publication Language: English

(30) Priority Data:
09/418,558 14 October 1999 (14.10.1999) US

(81) Designated State (*national*): JP.

(71) Applicant: **SYNTAXIA NETWORKS, LTD.** [—/GB];
Cutbush Court, Unit 3, Lower Earley, Reading RG6 4UW (GB).

(84) Designated States (*regional*): European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

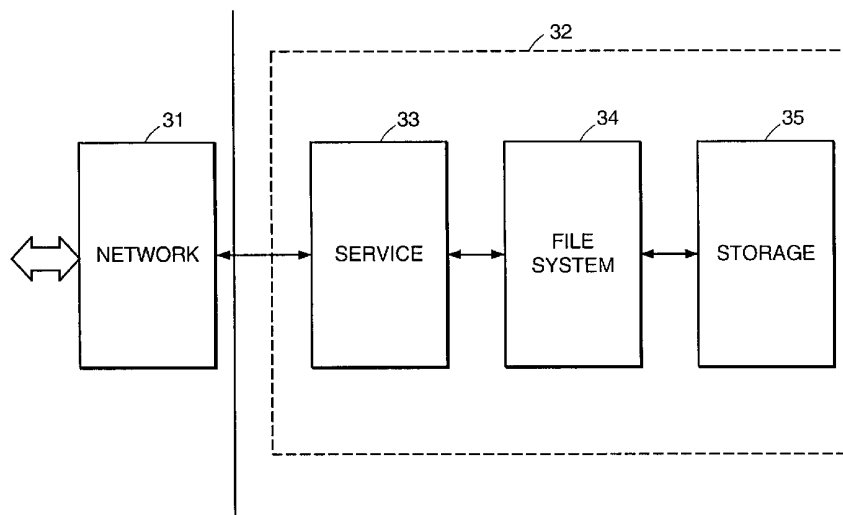
(72) Inventors: **BARRALL, Geoffrey, S.**; 9 Beauchief Close, Lower Earley, Berkshire RG46 4HY (GB). **WILLIS, Trevor**; 9 Ickfield Close, Aylesbury, Buckinghamshire HP22 6HG (GB). **BENHAM, Simon**; 4 Rockeby Close, Bracknell, Berkshire RG12 2NA (GB). **COOPER, Michael**; 102 Liverpool Road, Reading, Berkshire RG1

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.

(54) Title: APPARATUS AND METHOD FOR HARDWARE IMPLEMENTATION OR ACCELERATION OF OPERATING SYSTEM FUNCTIONS



(57) Abstract: An apparatus in one embodiment handles service requests over a network, wherein the network utilizes a protocol. In this aspect, the apparatus includes: a) a network subsystem for receiving and transmitting network service requests using the network protocol; and b) a service subsystem, coupled to the network subsystem, for satisfying the network service requests. Also in this aspect, at least one of the network subsystem and the service subsystem is hardware-implemented; the other of the network subsystem and the service subsystem may optionally be hardware-accelerated. A variety of related embodiments are also provided, including file servers and web servers.



WO 01/28179 A2

Apparatus and Method for Hardware Implementation or
Acceleration of Operating System Functions

Technical Field

5 The present invention relates to operating system functions and hardware implementation or acceleration of such functions.

Background Art

Operating systems in computers enable the computers to communicate with external resources. The operating system typically handles direct control of
10 items associated with computer usage including keyboard, display, disk storage, network facilities, printers, modems, etc. The operating system in a computer is typically designed to cause the central processing unit (CPU) to perform tasks including the managing of local and network file systems, memory, peripheral device drivers, and processes including application processes. Placing
15 responsibility for all of these functions on the CPU imposes significant processing burdens on it, particularly when the operating system is sophisticated, as, for example, in the case of Windows NT (available from Microsoft Corporation, Redmond, Washington), Unix (available from many sources, including from SCO Software, Santa Cruz, California, and, in a version
20 called "Linux", from Red Hat Software, Cambridge, Massachusetts), and NetWare (available from Novell, Provo, Utah). The more the burden is placed on the CPU to run processes other than those associated with applications, the less CPU time is available to run applications with the result that performance of the applications may be degraded. In addition, the throughput of devices
25 external to the CPU is subject to the limitations imposed by the CPU when the operating system places responsibility for managing these devices on the CPU. Furthermore, reliability of the overall software-hardware system, including the CPU, running the operating system, in association with the devices, will depend, among other things, on the operating system. Owing to the inherent
30 complexity of the operating system, unforeseen conditions may arise which may undermine stability of the overall software-hardware system.

Summary of the Invention

Certain aspects of the present invention enumerated in this summary are the subjects of other applications filed on the same date herewith. In one aspect of the invention, there is provided an apparatus for handling service requests
5 over a network, wherein the network utilizes a protocol. In this aspect, the apparatus includes:

- a. a network subsystem for receiving and transmitting network service requests using the network protocol; and
- b. a service subsystem, coupled to the network
10 subsystem, for satisfying the network service requests.

Also in this aspect, at least one of the network subsystem and the service subsystem is hardware-implemented; the other of the network subsystem and the service subsystem may optionally be hardware-accelerated. Alternatively, or in addition, the service subsystem may be hardware-accelerated.

15 In a related embodiment, the service requests include one of reading and writing data to long-term electronic storage; optionally, the network subsystem is hardware-accelerated. Also, optionally, the long-term storage is network disk storage accessible to computers over the network. Alternatively, the long-term storage is local disk storage that is accessible to a local computer but not to any
20 other computers over the network. Also optionally, the long-term storage may be associated with the provision of E-Mail service over the network; or it may provide access to web pages over the network.

Similarly, the service requests may involve access of data in a storage system, and the service subsystem may include a hardware-implemented
25 module for managing storage of the data in the storage system. Thus in one embodiment, such apparatus is a file server, wherein the data in the storage system are arranged in files, the service requests may involve requests for files in the storage system, and the service subsystem also includes a hardware-implemented module for managing a file system associated with the storage
30 system.

In another related aspect, aspect the protocol includes a file system

protocol, and the file system protocol defines operations including file read and file write. The apparatus may be a web server, wherein the data in the storage system may include web pages, and the service requests may involve requests for web pages in the storage system. Similarly the protocol may include IP. In a further related aspect, the storage system has a storage protocol and the service subsystem includes a hardware-implemented module for interfacing with the storage system.

In another aspect, a subsystem for receiving and transmitting data over a network, the network using a protocol having at least one of layers 3 and 4. In this aspect, the subsystem includes:

a receiver that receives encapsulated data from the network and de-encapsulates such data in accordance with the protocol; and

a transmitter that encapsulates data in accordance with the protocol and transmits the encapsulates data over the network.

At least one of the receiver and the transmitter is hardware-implemented; alternatively, or in addition, at least one of the receiver and the transmitter is hardware-accelerated. In a further embodiment, the network uses the TCP/IP protocol. In a related embodiment, the data are received over the network in packets, each packet having a protocol header, and the subsystem also includes a connection identifier that determines a unique connection from information contained within the protocol header of each packet received by the receiver. In another related embodiment, encapsulated data are associated with a network connection, and the subsystem further includes a memory region, associated with the network connection, that stores the state of the connection.

In another related aspect, there is provided a service subsystem for interfacing a storage arrangement with a network over which may be generated a storage access request. The service subsystem of this aspect includes:

- a. a service module that receives network service requests and fulfills such service requests and in doing so may issue data storage access requests;
- b. a file system module, coupled to the service module,

that receives data storage access requests from the service module and fulfills such storage access requests and in doing so may issue storage arrangement access requests; and

- 5 c. a storage module, coupled to the file system module, that receives storage arrangement access requests from the file system module and controls the storage arrangement to fulfill such storage arrangement access requests.

At least one of the modules is hardware-implemented; alternatively, or in addition, at least one of the modules is hardware-accelerated. In a related
10 embodiment, the service module includes:

- i. a receive control engine that receives network service requests, determines whether such requests are appropriate, and if so, responds if information is available, and otherwise issues a data storage access request; and
15 ii. a transmit control engine that generates network service responses based on instructions from the receive control engine, and, in the event that there is a data storage access response to the data storage access request, processes the data storage access response.

At least one of the engines is hardware-implemented; alternatively, or in addition, at least one of the engines is hardware-accelerated. In other related
20 embodiments, the service subsystem is integrated directly in the motherboard of a computer or integrated into an adapter card that may be plugged into a computer.

In another aspect, there is provided a service module that receives
25 network service requests and fulfills such service requests. The service module includes:

- a. a receive control engine that receives network service requests, determines whether such requests are appropriate, and if so, responds if information is available, and otherwise issues a data storage access request;
30 and
b. a transmit control engine that generates network

service responses based on instructions from the receive control engine, and, in the event that there is a data storage access response to the data storage access request, processes the data storage access response;

At least one of the engines is hardware-implemented; alternatively, or in addition at least one of the engines is hardware-accelerated. In related 5 embodiments, the network service requests are in the CIFS protocol, the SMB protocol, the HTTP protocol, the NFS protocol, the FCP. protocol, or the SMTP protocol. In a further related embodiment, the service module includes an authentication engine that determines whether a network request received by 10 the receiver has been issued from a source having authority to issue the request. In yet a further embodiment, the authentication engine determines whether a network request received by the receiver has been issued from a source having authority to perform the operation requested. Also, the service module may be integrated directly in the motherboard of a computer or integrated into an 15 adapter card that may be plugged into a computer.

In another aspect, there is provided a file system module that receives data storage access requests and fulfills such data storage access requests. The file system module includes:

a receiver that receives and interprets such data storage access requests 20 and in doing so may issue storage device access requests; and

a transmitter, coupled to the receiver, that constructs and issues data storage access responses, wherein such responses include information when appropriate based on responses to the storage device access requests.

At least one of the receiver and the transmitter is hardware-implemented; 25 alternatively, or in addition, at least one of the receiver and the transmitter is hardware-accelerated. In a further embodiment, the storage device access requests are consistent with the protocol used by a storage device to which the module may be coupled. In yet further embodiments, the protocol is NTFS, HPFS, FAT, FAT16, or FAT32. In another related embodiment, the file system 30 module also includes a file table cache, coupled to the receiver, that stores a table defining the physical location of files in a storage device to which the

module may be coupled. In various embodiments, the protocol does not require files to be placed in consecutive physical locations in a storage device. In other embodiments, the file system module is integrated directly in the motherboard of a computer or into an adapter card that may be plugged into a computer.

5 In another aspect, there is provided a storage module that receives storage device access requests from a request source and communicates with a storage device controller to fulfill such storage access requests. The storage module includes

a. a storage device request interface that receives such
10 storage device access requests and translates them into a format suitable for the storage device controller; and

b. a storage device acknowledge interface that takes the responses from the storage device controller and translates such responses into a format suitable for the request source.

15 At least one of the storage device request interface and the storage device acknowledge interface is hardware-implemented; alternatively, or in addition, at least one of the storage device request interface and the storage device acknowledge interface is hardware-accelerated. In a further embodiment, the storage module also includes a cache controller that maintains a local copy of a
20 portion of data contained on the storage device to allow fast-read access to the portion of data. In other related embodiments, the storage device request interface and the storage device acknowledge interface are coupled to a port; and the port permits communication with the storage device controller over a fiber-optic channel or utilizing a SCSI-related protocol. In further embodiments,
25 the storage module is integrated directly in the motherboard of a computer or integrated into an adapter card that may be plugged into a computer.

In another aspect, there is provided a system for interfacing a storage arrangement with a line on which may be placed a storage access request. The system in this aspect includes:

30 a. a service receive block, coupled to the storage arrangement, that processes the storage access request, generates where

necessary an access to the storage arrangement, and causes the generation of a response;

b. a file table cache, coupled to the receive block, that stores a table defining the physical location of files in the storage arrangement;
5 and

c. a service transmit block, coupled to the service receive block, for transmitting the response.

At least one of the service receive block and the service transmit block is hardware-implemented; alternatively, or in addition, at least one of the service
10 receive block and the service transmit block is hardware-accelerated. In a further embodiment, the system also includes response information memory, coupled to each of the service receive block and the service transmit block, which memory stores information present in a header associated the request, which information is used by the service transmit block in constructing the response.

15 In another related embodiment, the storage access request is a network request. In yet another embodiment, the storage access request is a generated by a local processor to which the line is coupled.

In another aspect, there is provided a process for handling storage access requests from multiple clients. The process includes:

20 testing for receipt of a storage access request from any of the clients; and
testing for completion of access to storage pursuant to any pending request.

In this embodiment, testing for receipt of a storage access request and testing for completion of access to storage are performed in a number of threads
25 independent of the number of clients. In a further embodiment, the process also includes, conditioned on a positive determination from testing for receipt of a request, processing the request that gave rise to the positive determination and initiating storage access pursuant to such request. In a related embodiment, the process also includes, conditioned on a positive determination from testing for
30 completion of access to storage pursuant to a pending request, sending a reply to the client issuing such pending request. In yet another related embodiment,

the number of threads is fewer than three. In fact the entire process may be embodied in a single thread.

In another aspect, there is provided a scalable apparatus for handling service requests over a network, wherein the network utilizes a protocol. The apparatus of this aspect includes:

a first plurality of network subsystems for receiving and transmitting network service requests using the network protocol; and

a second plurality of service subsystems, for satisfying the network service requests.

Each one of the network subsystems and the service subsystems being one of hardware-implemented or hardware-accelerated. In addition, the apparatus includes an interconnect coupling each of the first plurality of network subsystems to each of the second plurality of service subsystems. In a related embodiment, the interconnect is a switch or the interconnect is a bus.

In a related aspect, there is provided a scalable service subsystem for interfacing a storage arrangement with a network over which may be generated a storage access request. The service subsystem includes:

a first plurality of service modules that receive network service requests and fulfill such service requests and in doing so may issue data storage access requests; and

a second plurality of file system modules that receive data storage access requests and fulfill such storage access requests and in doing so may issue storage arrangement access requests;

Each one of the service modules and the file system modules is one of hardware-implemented or hardware-accelerated. The service subsystem also includes an interconnect coupling each of the first plurality of service modules to each of the second plurality of file system modules. In related embodiments, the interconnect is a switch or the interconnect is a bus. In a further embodiment, the scalable service subsystem includes a third plurality of storage modules that receive storage arrangement access requests controls the storage arrangement to fulfill such storage arrangement access requests, and each one of the storage

modules is one of hardware-implemented or hardware-accelerated; also the service subsystem includes a second interconnect coupling each of the file system modules to each of the storage modules. Similarly, in related embodiments, each of the interconnect and the second interconnect is a switch or each of the interconnect and the second interconnect is a bus.

Brief Description of the Drawings

The foregoing features of the invention will be more readily understood by reference to the following detailed description, taken with reference to the accompanying drawings, in which:

10 Fig. 1 is a schematic representation of an embodiment of the present invention configured to provide network services, such as a file server or a web server;

Fig. 2 is a block diagram of the embodiment illustrated in Fig. 1;

Fig. 3 is a block diagram of the embodiment configured as a file server;

15 Fig. 4 is a block diagram of the embodiment configured as a web server;

Fig. 5 is the network subsystem of the embodiments of Figs. 2-4;

Fig. 6 is a block diagram of the network subsystem of Fig. 5;

Fig. 7 is a block diagram of the receive module of the network subsystem of Fig. 6;

20 Fig. 8 is a block diagram of the transmit module of the network subsystem of Fig. 6;

Fig. 9 is a block diagram illustrating use of the network subsystem of Fig. 5 as a network interface adaptor for use with a network node, such as a workstation or server;

25 Fig. 10 is a block diagram of a hardware-implemented combination of the SMB service module 33 and file system module 34 of Fig. 3 for use in an embodiment such as illustrated in Fig. 3;

Fig. 11 is a block diagram of a hardware-accelerated combination of the SMB service module 33 and file system module 34 of Fig. 3 for use in an embodiment such as illustrated in Fig. 3;

30 Fig. 12A is a block diagram of a hardware-implemented service module

such as item 33 or 43 in Fig. 3 or Fig. 4 respectively;

Fig. 12B is a block diagram of a hardware-implemented file module such as item 34 or 44 in Fig. 3 or Fig. 4 respectively;

Fig. 12C is a detailed block diagram of the hardware-implemented service subsystem of Fig. 10, which provides a combined service module and file module;

Fig. 13 is a detailed block diagram of the hardware-accelerated service subsystem of Fig. 11;

Fig. 14 is a flow chart representing a typical prior art approach, implemented in software, for handling multiple service requests as multiple threads;

Fig. 15 is a flow chart showing the handling of multiple service requests, for use in connection with the service subsystem of Fig. 2 and, for example, the embodiments of Figs. 12 and 13;

Fig. 16 is a block diagram illustrating use of a file system module, such as illustrated in Fig. 3, in connection with a computer system having file storage;

Fig. 17A is a block diagram of data flow in the storage module of Fig. 3;

Fig. 17B is a block diagram of control flow in the storage module of Fig. 3;

Fig. 18 is a block diagram illustrating use of a storage module, such as illustrated in Fig. 3, in connection with a computer system having file storage.

Fig. 19 is a block diagram illustrating scalability of embodiments of the present invention, and, in particular, an embodiment wherein a plurality of network subsystems and service subsystems are employed utilizing expansion switches for communication among ports of successive subsystems and/or modules.

Detailed Description of Specific Embodiments

For the purpose of the present description and the accompanying claims, the following terms shall have the indicated meanings unless the context otherwise requires:

A "hardware-implemented" subsystem means a subsystem wherein major subsystem functions are performed in dedicated hardware that operates

outside the immediate control of a software program. Note that such a subsystem may interact with a processor that is under software control, but the subsystem itself is not immediately controlled by software. "Major" functions are the ones most frequently used.

5 A "hardware-accelerated" subsystem means one wherein major subsystem functions are carried out using a dedicated processor and dedicated memory, and, additionally (or alternatively), special purpose hardware; that is, the dedicated processor and memory are distinct from any central processor unit (CPU) and memory associated with the CPU.

10 "TCP/IP" are the protocols defined, among other places, on the web site of the Internet Engineering Task Force, at www.ietf.org, which is hereby incorporated herein by reference. "IP" is the Internet Protocol, defined at the same location.

 A "file" is a logical association of data.

15 A protocol "header" is information in a format specified by the protocol for transport of data associated with the user of the protocol.

 A "SCSI-related" protocol includes SCSI, SCSI-2, SCSI-3, Wide SCSI, Fast SCSI, Fast Wide SCSI, Ultra SCSI, Ultra2 SCSI, Wide Ultra2 SCSI, or any similar or successor protocol. SCSI refers to "Small Computer System Interface", which
20 is a standard for parallel connection of computer peripherals in accordance with the American National Standards Institute (ANSI), having a web URL address at www.ansi.org.

 Reference to "layers 3 and 4" means layers 3 and 4 in the Open System Interconnection ("OSI") seven-layer model, which is an ISO standard. The ISO
25 (International Organization for Standardization) has a web URL address at www.iso.ch.

 Fig. 1 is a schematic representation of an embodiment of the present invention configured to handle service requests over a network. Thus this embodiment includes configurations in which there is provided a file server or a
30 web server. The embodiment 11 of the present invention is coupled to the network 10 via the network interface 13. The network 10 may include, for

example, communications links to a plurality of workstations. The embodiment 11 here is also coupled to a plurality of storage devices 12 via storage interconnect 14. The embodiment 11 may be hardware implemented or hardware accelerated (or utilize a combination of hardware implementation and hardware acceleration).

Fig. 2 is a block diagram of the embodiment illustrated in Fig. 1. The network subsystem 21 receives and transmits network service requests and responses. The network subsystem 21 is coupled to the service subsystem 22, which satisfies the network service requests. The network subsystem 21, the service subsystem 22, or both subsystems may be either hardware implemented or hardware accelerated.

Fig. 3 is a block diagram of the embodiment of Fig. 1, more particularly configured as a file server. The network subsystem 31 receives and transmits network service requests and responses. The network subsystem 31 is coupled to the service subsystem 32. The service subsystem includes three modules: the service module 33, the file system module 34, and the storage module 35. The service module 33 analyzes network service requests passed to the service subsystem 32 and issues, when appropriate, a corresponding storage access request. The network service request may be conveyed in any of a variety of protocols, such as CIFS, SMB, NFS, or FCP. The service module 33 is coupled to the file system module 34. If the network service request involves a storage access request, the file system module 34 converts requests for access to storage by converting the request into a format consistent with the file storage protocol (for example, HTFS, NTFS, FAT, FAT16, or FAT32) utilized by the storage medium. The storage module 35 converts the output of the file system module 34 into a format (such as SCSI) consistent with the bus requirements for directly accessing the storage medium to which the service subsystem 32 may be connected.

Fig. 4 is similar to Fig. 3, and is a block diagram of the embodiment of Fig. 1 configured as a web server. The network subsystem 41 receives and transmits network service requests and responses. The network subsystem 41 is

coupled to the service subsystem 42. The service subsystem includes three modules: the service module 43, the file system module 44, and the storage module 45. The service module 43 analyzes network service requests passed to the service subsystem 32 and issues, when appropriate, a corresponding storage access request. Here, the network service requests is typically in the HTTP
5 protocol. The service module 43 is coupled to the file system module 44, which is coupled to the storage module 45; the file system module 44 and the storage module 45 operate in a manner similar to the corresponding modules 34 and 35 described above in connection with Fig. 3.

10 Fig. 5 is the network subsystem and service subsystem of the embodiments of Figs. 2-4. The network subsystem 51 receives encapsulated data from the network receive interface 54 and de-encapsulates the data in accordance with the TCP/IP or other protocol bus 53. The network subsystem 51 is also coupled to the PCI bus 53 to provide to a local processor (which is also
15 coupled to the PCI bus) to access data over the network. The network subsystem 51 also transmits the data to the service subsystem 52, and the data to be transmitted may come from the network receive interface 54 or the local processor via the PCI bus 53. The service subsystem 52, in turn, operates in a manner similar to the service subsystems 22, 32, and 42 Figs. 2, 3, and 4
20 respectively.

Fig. 6 is a detailed block diagram of the network subsystem 51 of Fig. 5. The network subsystem of Fig. 6 includes a receiver module 614 (which includes a receiver 601, receive buffer memory 603, and receive control memory 604) and a transmitter module 613 (which includes transmitter 602, transmit buffer
25 memory 605, and transmit control memory 606). The processor 611 is used by both the receiver module 614 and the transmitter module 613. The receiver 601 receives and interprets encapsulated data from the network receive interface 607. The receiver 601 de-encapsulates the data using control information contained in the receive control memory 604 and transmit control memory 606
30 and stores the de-encapsulated data in the receive buffer memory 603, from where it is either retrieved by the processor 611 via PCI bus 613 or output to the

receive fast path interface 606. Memory 612 is used by processor 611 for storage of data and instructions.

The transmitter 602 accepts transmit requests from transmit fast path interface 610 or from the processor 611 via PCI bus 613. The transmitter 602
5 stores the data in transmit buffer memory 605. The transmitter 602 encapsulates the transmit data using control information contained in the transmit control memory 606 and transmits the encapsulated data over the network via the network transmit interface 609.

Fig. 7 is a block diagram of the receive module 614 of the network
10 subsystem of Fig. 6. Packets are received by the receive engine 701 from the network receive interface 607. The receive engine 701 analyzes the packets and determines whether the packet contains an error, is a TCP/IP packet, or is not a TCP/IP packet. A packet is determined to be or not to be a TCP/IP packet by
15 examination of the network protocol headers contained in the packet. If the packet contains an error then it is dropped.

If the packet is not a TCP/IP packet then the packet is stored in the receive buffer memory 603 via the receive buffer memory arbiter 709. An indication that a packet has been received is written into the processor event queue 702. The processor 713 can then retrieve the packet from the receive
20 buffer memory 603 using the PCI bus 704 and the receive PCI interface block 705.

If the packet is a TCP/IP packet then the receive engine 701 uses a hash table contained in the receive control memory 604 to attempt to resolve the network addresses and port numbers contained within the protocol headers in
25 the packet into a number which uniquely identifies the connection to which this packet belongs, i.e., the connection identification. If this is a new connection identification, then the packet is stored in the receive buffer memory 603 via the receive buffer memory arbiter 708. An indication that a packet has been received is written into the processor event queue 702. The processor 713 can
30 then retrieve the packet from the receive buffer memory 603 using the PCI bus 704 and the receive PCI interface block 703. The processor can then establish a

new connection if required as specified in the TCP/IP protocol, or it can take other appropriate action.

If the connection identification already exists, then the receive engine 701 uses this connection identification as an index into a table of data which contains information about the state of each connection. This information is called the "TCP control block" ("TCB"). The TCB for each connection is stored in the transmit control memory 606. The receive engine 701 accesses the TCB for this connection via the receiver TCB access interface 710. It then processes this packet according to the TCP/IP protocol and adds the resulting bytes to the received byte stream for this connection in the receive buffer memory 603. If data on this connection is destined for the processor 713 then an indication that some bytes have been received is written into the processor event queue 702. The processor can then retrieve the bytes from the receive buffer memory 603 using the PCI bus 704 and the receive PCI interface block 703. If data on this connection is destined for the fast path interface 608, then an indication that some bytes have been received is written into the fast path event queue 705. The receive DMA engine 706 will then retrieve the bytes from the receive buffer memory 603 and output them to the fast path interface 608.

Some packets received by the receive engine 701 may be fragments of IP packets. If this is the case then the fragments are first reassembled in the receive buffer memory 603. When a complete IP packet has been reassembled, the normal packet processing is then applied as described above.

According to the TCP protocol, a connection can exist in a number of different states, including SYN_SENT, SYN_RECEIVED and ESTABLISHED. When a network node wishes to establish a connection to the network subsystem, it first transmits a TCP/IP packet with the SYN flag set. This packet is retrieved by the processor 713 since it will have a new connection identification. The processor 713 will then perform all required initialization including setting the connection state in the TCB for this connection to SYN_RECEIVED. The transition from SYN_RECEIVED to ESTABLISHED is performed by the receive engine 701 in accordance with the TCP/IP protocol. When the processor 713 wished to establish

a connection to a network node via the network subsystem, it first performs all required initialization including setting the connection state in the TCB for this connection to SYN_SENT. It then transmits a TCP/IP packet with the SYN flag set. The transition from SYN_SENT to ESTABLISHED is performed by the receive engine 701 in accordance with the TCP/IP protocol.

If a packet is received which has a SYN flag or FIN flag or RST flag set in the protocol header, and if this requires action by the processor 713, then the receive engine 701 will notify the processor of this event by writing an entry into the processor event queue 702. The processor 713 can then take the appropriate action as required by the TCP/IP protocol.

As a result of applying the TCP/IP protocol to the received packet it is possible that one or more packets should now be transmitted on this connection. For example, an acknowledgment of the received data may need to be transmitted, or the received packet may indicate an increased window size thus allowing more data to be transmitted on this connection if such data is available for transmission. The receive engine 701 achieves this by modifying the TCB accordingly and then requesting a transmit attempt by writing the connection identification into the transmit queue 802 in Fig. 8 via the receiver transmit queue request interface 711.

Received data is stored in discrete units (buffers) within the receive buffer memory 603. As soon as all the data within a buffer has been either retrieved by the processor 713 or outputted to the fast path interface 608 then the buffer can be freed, i.e., it can then be reused to store new data. A similar system operates for the transmit buffer memory 605, however, in the transmit case, the buffer can only be freed when all the data within it has been fully acknowledged, using the TCP/IP protocol, by the network node which is receiving the transmitting data. When the protocol header of the packet indicates that transmitted data has been acknowledged, then the receive engine 701 indicates this to the free transmit buffers block 805 in Fig. 8 via the receiver free transmit buffers request interface 712.

Additionally, it is possible for the receive engine 701 to process the upper layer protocol (ULP) that runs on top of TCP/IP as well as TCP/IP itself. In this

case, event queue entries are written into the processor event queue 702 and the fast path event queue 705 only when a complete ULP protocol data unit (PDU) has been received; only complete ULP PDUs are received by the processor 713 and outputted to the fast path interface 608. An example of a ULP is NetBIOS. The enabling of ULP processing may be made on a per-connection basis; *i.e.*, some connections may have ULP processing enabled, and others may not.

Fig. 8 is a block diagram of the transmit module 613 of the network subsystem of Fig. 6. Data to be transmitted over the network using TCP/IP is inputted to the transmit DMA engine 807. This data is either input from the transmit fast path interface 610 or from the processor 713 via PCI bus 704 and the transmit PCI interface 808. In each case, the connection identification determining which TCP/IP connection should be used to transmit the data is also input. As mentioned above, each connection has an associated TCB which contains information about the state of the connection.

The transmit DMA engine stores the data in the transmit buffer memory 605, adding the inputted bytes to the stored byte stream for this connection. At the end of the input it modifies the TCB for the connection accordingly and it also writes the connection identification into the transmit queue 802.

The transmit queue 802 accepts transmit requests in the form of connection identifications from three sources: the received transmit queue request interface 711, the timer functions block 806, and the transmit DMA engine 807. As the requests are received they are placed in a queue. Whenever the queue is not empty, a transmit request for the connection identification at the front of the queue is passed to the transmit engine 801. When the transmit engine 801 has completed processing the transmit request this connection identification is removed from the front of the queue and the process repeats.

The transmit engine 801 accepts transmit requests from the transmit queue 802. For each request the transmit engine 801 applies the TCP/IP protocol to the connection and transmit packets as required. In order to do this it accesses the TCB for the connection in the transmit control memory 606, via the transmit control memory arbiter 803, and it retrieves the stored byte stream for the

connection from the transmit buffer memory 605 via the transmit buffer memory arbiter 804.

The stored byte stream for a connection is stored in discrete units (buffers) within the transmit buffer memory 605. As mentioned above, each buffer can only
5 be freed when all the data within it has been fully acknowledged, using the TCP/IP protocol, by the network node which is receiving the transmitting data. When the protocol header of the packet indicates that transmitted data has been acknowledged then the receive engine 701 indicates this to the free transmit buffers block 805 via the receiver free transmit buffers request interface 712. The free
10 transmit buffers block 805 will then free all buffers which have been fully acknowledged and these buffers can then be reused to store new data.

TCP/IP has a number of timer functions which require certain operations to be performed at regular intervals if certain conditions are met. These functions are implemented by the timer functions block 806. At regular intervals the timer
15 functions block 806 accesses the TCBs for each connection via the transmit control memory arbiter 803. If any operation needs to be performed for a particular connection, then the TCB for that connection is modified accordingly and the connection identification is written to the transmit queue 802.

Additionally it is possible for the transmit DMA engine 807 to process the
20 upper layer protocol that runs on top of TCP/IP. In this case, only complete ULP protocol data units are inputted to the transmit DMA engine 807, either from the processor 713 or from the transmit fast path interface 610. The transmit DMA engine 807 then attaches the ULP header at the front of the PDU and adds the "pre-pended" ULP header and the inputted bytes to the stored byte stream for the
25 connection. As discussed in connection with Fig. 7 above, an example of a ULP is NetBIOS. The enabling of ULP processing may be made on a per-connection basis; *i.e.*, some connections may have ULP processing enabled, and others may not.

If the processor 713 wishes to transmit a raw packet, *i.e.*, to transmit data without the hardware's automatic transmission of the data using TCP/IP, then
30 when the processor 713 inputs the data to the transmit DMA engine 807 it uses a special connection identification. This special connection identification causes the

transmit engine 801 to transmit raw packets, exactly as input to the transmit DMA engine 807 by the processor 713.

Fig. 9 is a block diagram illustrating use of the network subsystem of Fig. 5 as a network interface adaptor for use with a network node, such as a workstation or server. In this embodiment, the network subsystem 901 is integrated into an adapter card 900 that is plugged into a computer. The adapter card 900 is coupled to the network via the network interface 904. The adapter card 900 is also coupled to the computer's microprocessor 910 via the PCI bus 907 and the PCI bridge 912. The PCI bus 907 may also be used by the computer to access peripheral devices such as video system 913. The receive module 902 and transmit module 903 operate in a manner similar to the receive module 614 and transmit module 613 of Fig. 6. Alternately or in addition, the adapter card 900 may be connected, via single protocol fast receive pipe 906 and single protocol fast transmit pipe 908, to a service module comparable to any of items 22, 32, 42, or 52 of Figs. 2, 3, 4, or 5 respectively, for providing rapid access to a storage arrangement by a remote node on the network or by the microprocessor 910.

Fig. 10 is a block diagram of a hardware-implemented combination of the SMB service module 33 and file system module 34 of Fig. 3 for use in an embodiment such as illustrated in Fig. 3. In the embodiment of Fig. 10, SMB requests are received on the input 105 to the service receive block 101. Ultimately processing by this embodiment results in transmission of a corresponding SMB response over the output 106. A part of this response includes a header. To produce the output header, the input header is stored in SMB response information memory 103. The block 101 processes the SMB request and generates a response. Depending on the nature of the request, the block 101 may access the file table cache 104 and issue a disk access request; otherwise the response will be relayed directly the transmit block 102. The service transmit block 102 transmits the response, generated by block 101, over the output 106. In the event that a disk access request has been issued by block 101, then upon receipt over line 108 of a disk response, the transmit block 102 issues the appropriate SMB response over line 106. Both the receive and transmit modules 101 and 102 are optionally in

communication with the host system over PCI bus 109. Such communication, when provided, permits a host system to communicate directly with the embodiment instead of over a network, so as to give the host system rapid, hardware-implemented file system accesses, outside the purview of a traditional operating system.

5 Fig. 11 is a block diagram of a hardware-accelerated combination of the SMB service module 33 and file system module 34 of Fig. 3 for use in an embodiment such as illustrated in Fig. 3. The operation is analogous to that described above in connection with Fig. 10 with respect to similarly numbered
10 blocks and lines 105, 107, 108, and 106. However, the dedicated file system processor 110, in cooperation with dedicated memory 111 operating over dedicated bus 112 control the processes of blocks 101 and 102. Additionally these items provide flexibility in handling of such processes, since they can be reconfigured in software.

15 Fig. 12A is a block diagram of a hardware-implemented service module such as item 33 or 43 in Fig. 3 or Fig. 4 respectively. The service module 1200 receives network service requests, fulfills such service requests, and may issue data storage access requests. The service module 1200 includes a receiver 1201 coupled to a transmitter 1202 and a data storage access interface 1203, which is also coupled
20 to both the receiver 1201 and the transmitter 1202. The receiver 1201 receives and interprets network service requests. On receipt of a service request, the receiver 1201 either passes the request to the data storage access interface 1203 or passes information fulfilling the network service request to the transmitter 1202. If the request is passed to the data storage access interface 1203, the data storage access
25 interface 1203 constructs and issues data storage access requests. The data storage access interface 1203 also receives replies to the data storage access requests and extracts information required to fulfill the original network service request. The information is then passed to the transmitter 1202. The transmitter 1202 processes information passed to it from the receiver 1201 or the data storage access interface
30 1203 and constructs and issues network service replies.

Fig. 12B is a block diagram of a hardware-implemented file module such as

item 34 or 44 in Fig. 3 or Fig. 4 respectively. The file system module 1210 receives data storage access requests, fulfills such data service access requests, and may issue storage device access requests. The file system module 1210 includes a receiver 1211 coupled to a transmitter 1212 and a data storage device access
5 interface 1213 which is also coupled to both the receiver 1211 and the transmitter 1212. The receiver 1211 receives and interprets data storage access requests and either passes the request to the data storage device access interface 1213 or passes information fulfilling the data storage access request to the transmitter 1212. If the request is passed to the data storage device access interface 1213, the data storage
10 device access interface 1213 constructs and issues data storage device access requests. The data storage device access interface 1213 also receives replies to the data storage device access requests and extracts information required to fulfill the original data storage access request. The information is then passed to the transmitter 1212. The transmitter 1212 processes information passed to it from the
15 receiver 1211 or the data storage device access interface module 1213 and constructs and issues data storage access replies.

Fig. 12C is a detailed block diagram of the hardware-implemented service subsystem of Fig. 10, which provides a combined service module and file module. Dashed line 129 in Fig. 12C shows the division between functions of this
20 implementation. To the left of line 129 is the service module portion; to the right of line 129 is the file system module portion. (It will be understood, however, that the double-headed arrow connecting the SMB receive control engine 121 and the SMB transmit control engine 122 properly provides two-way communication between the engines 121 and 122 for each of the service module portion and the file system
25 module portion.)

In Fig. 12C, SMB frames are received from the network subsystem via the network receive interface 121f and are passed to the SMB frame interpretation engine 121b. Here the frame is analyzed and a number of tasks are performed. The first section of the header is copied to the SMB response info control 123, which
30 stores relevant information on a per connection basis in the SMB response info memory 103. The complete frame is written into buffers in the receive buffer

memory 121c and the receive control memory 121d is updated. Relevant parts of the SMB frame header are passed to the SMB receive control engine 121.

The SMB receive control engine 121 of Fig. 12C parses the information from the header and, where appropriate, requests file access permission from the authentication engine 124. For SMB frames where a file access has been requested, the SMB receive control engine 121 extracts either file path information or the file identification from the SMB frame header and requests the MFT control engine 125 for the physical location of the required file data.

The MFT control engine 125 can queue requests from the SMB receive control engine 121 and similarly the SMB receive control engine 121 can queue responses from the MFT control engine 125. This allows the two engines to operate asynchronously from each other and thus allows incoming SMB frames to be processed while MFT requests are outstanding.

The MFT control engine 125 processes requests from the SMB receive control engine 121. Typically for SMB OPEN commands, a request will require a disk access to obtain the necessary physical file location information. Where this is necessary, the MFT control engine 125 passes a request to the compressed SCSI frame generation engine 121a which will generate the necessary compressed SCSI request. The compressed SCSI protocol ("CSP") relates to a data format from which a SCSI command may be generated in the manner described in connection with Fig. 17A and other figures below. Because compressed SCSI data are not derived from SCSI but are rather the source from which SCSI data may be derived, we sometimes refer to compressed SCSI data as "proto-SCSI" data. The relevant proto-SCSI response will be passed back to the MFT control engine 125, where it will be processed, the MFT cache 104 will be updated, and the physical file information will be passed back to the SMB receive control engine 121.

Typically, for a SMB READ or WRITE command with respect to a recently accessed small file, the file information will be present in the MFT cache 104. Thus no disk access will be required.

When the SMB receive control engine 121 has received the response from an MFT request and a disk access for file data is required, as would be necessary for

typical READ or WRITE commands, one or more proto-SCSI requests are passed to the proto-SCSI frame generation engine 121a.

The proto-SCSI frame generation engine 121a will construct the proto-SCSI headers and, where necessary, for example, for WRITE commands, program the file data DMA engine 121e to pull the file data out of the receive buffer memory 121c. The proto-SCSI frame is then passed to the proto-SCSI module via proto-SCSI transmit interface 121g. Where no disk access is required, an SMB response request is passed directly to the SMB transmit control engine 122.

Proto-SCSI frames are received from the proto-SCSI module and via proto-SCSI receive interface 122f are passed to the proto-SCSI frame interpretation engine 122b. Here the frame is analyzed and a number of tasks are performed. MFT responses are passed back to the MFT control engine 125. All other frames are written into buffers in the receive buffer memory 121c and the receive control memory 121d is updated. Relevant parts of the proto-SCSI frame header are passed to the SMB transmit control engine 122.

Each SMB connection has previously been assigned a unique identification. All proto-SCSI frames include this identification and the SMB transmit control engine 122 uses this unique identification to request state information from the SMB receive control engine 121 and update this where necessary. When all necessary information for an SMB response has been received from the proto-SCSI module, the SMB transmit control engine 122 passes a request to the SMB frame generation engine 122a.

The SMB frame generation engine 122a constructs the SMB response frame from data contained in the SMB response info memory 103 and file data stored in the SMB transmit buffer memory 122c. It then passes the frame to the SMB transmit interface 106 which in turn forwards it to the network subsystem.

Fig. 13 is a detailed block diagram of the hardware-accelerated service subsystem of Fig. 11. Incoming SMB frames from the IP block are provided over input 105 are written, via the SMB receive FIFO 1317, into free buffers in the SMB receive buffer memory 121c. The SMB receive buffer memory 121c includes in one embodiment a series of receive buffers that are 2Kb long and thus one SMB frame

may straddle a number of receive buffers. As frames are written into SMB receive buffer memory 121c, SMB receive buffer descriptors are updated in the SMB receive control memory 121d.

5 A 32-bit connection identification and a 32-bit frame byte count are passed to the SMB block from the IP block at the start of the frame. These two fields are written to the first two locations of the receive buffer in receive buffer memory 121c.

10 While the frame is being stored, the SMB header is also written to the SMB response info memory 103 for later use by the SMB transmit process. The unique connection identification passed to the SMB block by the IP block is used as a pointer to the appropriate info field in the SMB response info memory 103. This memory is arranged as blocks of 16 words, one block for each unique connection identification. With a 128Mb SDRAM fitted, this allows 2M connections. At present just the first 32 bytes of the SMB frame are written to each info field.

15 When a complete frame has been written to the receive buffer memory 121c, an SMB buffer locator is written to the SMB receive event queue 1314 and an interrupt to the host processor 1301 is generated. The SMB buffer locator contains information pertaining to the SMB frame including a buffer pointer and a 'last' bit. The buffer pointer points to the buffer in receive buffer memory 121c which contains the start of the SMB frame. The 'last' bit indicates whether this buffer also
20 contains the end of the SMB frame (i.e., whether the SMB frame is less than 2Kb in length).

25 The host processor 1301 can read the SMB buffer locator in the SMB receive event queue 1314 by reading an appropriate SMB receive event register associated with the event queue 1314. From the buffer pointer read from the SMB buffer locator the host processor 1301 can determine the address of the first buffer of the SMB frame in the receive buffer memory 121c and can thus read the SMB header and the first part of the frame.

30 If the SMB frame is longer than 2Kb and it is necessary to read more than the first 2Kb of the SMB frame, then the receive buffer descriptor associated with this receive buffer should be read from the receive control memory 121d. This receive buffer descriptor will contain a pointer to the next buffer of the SMB frame. This

next buffer will similarly have a receive buffer descriptor associated with it unless the previous buffer's descriptor contained a 'last' bit indicating that the receive buffer it pointed to contained the end of the SMB frame.

5 After reading the received SMB frame, if none of the data contained within the frame is to be used further, then the buffers of the received frame are made available for use again by writing pointers to them to the receive free buffers queue, which is contained in the receive buffer control memory 121d, by writing to an associated receive return free buffers register.

10 To transmit a proto-SCSI frame the host processor 1301 firstly obtains a pointer to a free SMB receive buffer by reading from the receive fetch free buffer register. This action will pull a pointer to a free buffer from the free buffers queue contained in the receive control memory 121d. In this buffer the start of the proto-SCSI request frame can be constructed.

15 To request the proto-SCSI transmit entity to transfer the proto-SCSI frame to the proto-SCSI entity, the host processor 1301 writes a buffer locator and buffer offset pair to the proto-SCSI transmit event queue 1315 by writing them to the receive proto-SCSI event register associated with the proto-SCSI transmit event queue 1315.

20 The buffer locator contains a pointer to the buffer containing data for the proto-SCSI frame. The buffer offset contains an offset to the start of the data within the buffer and a length field. The buffer locator also contains a last bit to indicate whether further buffer locator/buffer offset pairs will be written to the proto-SCSI transmit event queue 1315 containing pointers to more data for this proto-SCSI frame.

25 If the proto-SCSI frame is to include data from another SMB receive buffer, as would be typical for a SMB WRITE command, then the host processor 1301 must write another buffer locator/buffer offset pair describing this SMB receive buffer to the proto-SCSI transmit event queue 1315. If the data to be included in the proto-SCSI frame straddles more than one SMB receive buffer, then the proto-SCSI
30 transmit entity can use the buffer pointers in the associated SMB receive buffer descriptor located in receive control memory 121d to link the data together. If the

extra data is from a SMB receive frame, then these descriptors will have been filled in previously by the SMB receive entity.

Because data from SMB receive buffers may be used for more than one proto-SCSI frame, then freeing up the SMB receive buffers after they have been
5 used is not a simple process. SMB receive buffers that contain sections of a received SMB frame that are not involved in the proto-SCSI transmit can be freed by writing them back to the free buffers queue contained in the receive control memory via the associated receive return free buffer register. SMB receive buffers that contain data to be included in proto-SCSI frames can not be freed in the same way as they can
10 not be freed until the data within them has been transmitted.

So, after the buffer locator/buffer offset pairs to the various proto-SCSI frames which will contain the SMB data have been written to the proto-SCSI transmit event queue 1315, pointers to the original SMB receive buffers are also written to the proto-SCSI transmit event queue 1315. These pointers are marked to
15 indicate that they are to be freed back to the free buffers queue contained in the receive control memory. As the proto-SCSI transmit event queue 1315 is handled in sequence, then the SMB receive buffers will only be freed after any data within them has been transmitted.

Incoming proto-SCSI frames from the IP block are written, via the proto-SCSI receive FIFO 1327, into free buffers in the SMB transmit buffer memory 122c. The
20 SMB transmit buffers are 2Kb long and thus one proto-SCSI frame may straddle a number of transmit buffers. As frames are written into SMB transmit buffer memory 122c, SMB transmit buffer descriptors are updated in the SMB transmit control memory 122d.

25 When a complete frame has been written to the SMB transmit buffer memory 122c, an SMB buffer locator is written to the proto-SCSI receive event queue 1324 and an interrupt to the host processor 1301 is generated. The SMB buffer locator contains information pertaining to the proto-SCSI frame including a buffer pointer and a 'last' bit. The buffer pointer points to the buffer in transmit buffer memory
30 121c which contains the start of the proto-SCSI frame. The 'last' bit indicates

whether this buffer also contains the end of the proto-SCSI frame (i.e., whether the frame is less than 2Kb in length).

The host processor 1301 can read the buffer locator in the proto-SCSI receive event queue 1324 by reading an appropriate proto-SCSI receive event register associated with the event queue 1324. From the buffer pointer read from the buffer locator the host processor 1301 can determine the address of the first buffer of the proto-SCSI frame in the transmit buffer memory 122c and can thus read the header and the first part of the frame.

If the proto-SCSI frame is longer than 2Kb and it is necessary to read more than the first 2Kb of the frame then the transmit descriptor associated with this transmit buffer should be read from the receive control memory 121d. The descriptor will contain a pointer to the next buffer of the proto-SCSI frame. This next buffer will similarly have a transmit descriptor associated with it unless the previous buffer's descriptor contained a 'last' bit indicating that the buffer it pointed to contained the end of the proto-SCSI frame.

After reading the received proto-SCSI frame, if none of the data contained within the frame is to be used further, then the buffers of the received frame should be returned to the transmit free buffers queue contained in the transmit control memory 122d by writing to the transmit return free buffers register associated with it.

To transmit an SMB frame, the host processor first obtains a pointer to a free SMB transmit buffer in transmit buffer memory 122c from the transmit free buffer queue contained in the transmit control memory 122d by reading from an associated register. In this buffer the start of the SMB response frame can be constructed.

The 32-bit connection identification and a 32-bit SMB transmit control field are placed before the SMB frame in the buffer. The SMB transmit control field includes a 24-bit frame byte count and a pre-pend header bit. If the pre-pend header bit is set, then after the connection identification and SMB transmit control field have been passed to the IP block then the SMB header stored in the response info memory 103 will be automatically inserted.

To request the SMB transmit entity to transfer the SMB frame to the SMB entity, the host processor 1301 writes a buffer locator and buffer offset pair to the SMB transmit event queue 1325 by writing them to an associated transmit SMB transmit event register.

5 The buffer locator contains a pointer to the buffer containing data for the SMB frame. The buffer offset contains an offset to the start of the data within the buffer and a length field. The buffer locator also contains a last bit to indicate whether further buffer locator/buffer offset pairs will be written containing pointers to more data for this SMB frame.

10 If the SMB frame is to include data from another SMB transmit buffer in buffer memory 122c, then the host processor 1301 must write another buffer locator/buffer offset pair describing this SMB transmit buffer to the SMB transmit event queue 1325. If the data to be included in the SMB frame straddles more than one SMB transmit buffer, then the SMB transmit entity can use the buffer pointers
15 in the associated transmit buffer descriptor to link the data together. If the extra data is from a proto-SCSI receive frame, then these descriptors will have been filled in previously by the proto-SCSI receive entity.

Because data from SMB transmit buffers in transmit buffer memory 122c may be used for more than one SMB frame, then freeing up the SMB transmit buffers
20 after they have been used is not a simple process. SMB transmit buffers that contain sections of a received proto-SCSI frame that are not involved in the SMB transmit can be freed by writing them back to the transmit free buffers queue contained in the transmit control memory via the associated transmit return free buffers register. SMB transmit buffers that contain data to be included in SMB frames cannot be
25 freed in the same way, as they cannot be freed until the data within them has been transmitted.

So, after the buffer locator/buffer offset pairs to the various SMB frames which will contain the proto-SCSI data have been written to the SMB transmit event queue 1325, pointers to the original SMB transmit buffers are also written to the
30 SMB transmit event queue 1325. These pointers are marked to indicate that they are to be freed back to the transmit free buffers queue. As the SMB transmit event

queue 1325 is handled in sequence, then the SMB transmit buffers will only be freed after any data within them has been transmitted.

Fig. 14 is a flow chart representing a typical prior art approach, implemented in software, for handling multiple service requests as multiple threads. In a traditional multiple-threaded architecture there is typically at least one thread to service each client. Threads are started and ended as clients attach and detach from the server. Each client may have a thread on the server to handle service requests and a thread to handle disk requests. The service process 1400 includes a repeated loop in which there is testing for the presence of a client connection request in box 1401; if the test is in the affirmative, the process initiates in box 1402, the client process 1430. When the client process 1430 requires disk access in box 1435, it first requests the appropriate disk process to access the disk and then sleeps in box 1436 until the disk access completes. The disk process 1402 then wakes up the client process 1430 to allow it to send the reply in box 1437 to the client issuing the service request. Thus there are at least two process switches for each client request requiring disk access. Implementing these multiple threaded processes in hardware poses problems because normally, they are handled by a multi-tasking operating system.

Fig. 15 is a flow chart showing the handling of multiple service requests in a single thread, for use in connection with the service subsystem of Fig. 2 and, for example, the embodiments of Figs. 12 and 13. In the single-threaded architecture one service process 1500 handles requests from multiple clients in a single thread and one disk process 1502 handles all the requests from the service process 1500. The prior art approach of using a separate process for each client making a request has been eliminated and its function has been here handled by the single service process 1500. Additionally, these two processes, the service process and the disk process, may be contained within the same thread, as illustrated, or may be shared between two separate threads to facilitate load balancing.

The single-threaded service process of Fig. 15 can have disk requests from multiple clients outstanding simultaneously. The single thread includes a main loop with two tests. The first test in box 1501 is whether a request has been received

from a client. The second test in box 1508 is whether a previously initiated disk access request has been completed. In consequence, as a disk access has been determined in box 1508 to have been completed, the service process in box 1507 will send the appropriate reply back to the client. Once the service process 1500 has handled a disk access request via box 1501 and caused the request to be processed in box 1502, and caused, in box 1504, the initiation of a disk access, the service process is free to handle another request from another client via box 1501 without having to stop and wait for the previous disk access to complete. Upon a determination in box 1508 that the disk access has been completed, the disk process in box 1507 will inform the service process of the result and the service process will send the response to the client. Thus the service and disk processes will be constantly running as long as there are requests being sent from clients.

Fig. 16 is a block diagram illustrating use of a file system module, such as illustrated in Fig. 3, in connection with a computer system having file storage. (An implementation analogous to that of Fig. 16 may be used to provide a storage module, such as illustrated in Fig. 3, in connection with a computer system having file storage.) In this embodiment, the file system module 1601 is integrated into a computer system, which includes microprocessor 1605, memory 1606, and a peripheral device, such as video 1609, as well as disk drive 1610, accessed via a disk subsystem 1602, which is here a conventional disk drive controller. The file system module 1601 is coupled to the disk subsystem 1602. The file system module 1601 is also coupled to the computer multiprocessor 1605 and the computer memory 1606 via the PCI bridge 1604 over the PCI bus 1607. The PCI bus 1607 also couples the microprocessor 1605 to the computer peripheral device 1609. The receive engine 1610 of the file system module processes disk access requests from the microprocessor 1605 in a manner similar to that described above with respect to Figs. 10, 11, 12B, and 13. Also the transmit engine 1611 provides responses to such disk access requests in a manner similar to that described above with respect to Figs. 10, 11, 12B, and 13.

Fig. 17A is a block diagram of data flow in the storage module of Fig. 3. It should be noted that while in Figs. 17A and 17B a Tachyon XL fiber optic channel

controller, available from Hewlett Packard Co., Palo Alto, California, has been used as the I/O device, embodiments of the present invention may equally use other I/O devices. Proto-SCSI requests are received over the proto-SCSI input 1700 by the proto-SCSI request processor 1702. The information relating to this request is stored in a SEST information table, and if this is a WRITE request, then the WRITE data, which are also provided over the proto-SCSI input 1700, are stored in the WRITE buffer memory 1736.

The exchange request (ERQ) generator 1716 takes the information from the WRITE buffer memory 1736. If all the buffers to be written are currently cached, or the data to be written completely fill the buffers to be written, then the WRITE can be performed immediately. The data to be written are copied from WRITE buffer memory 1736 to the appropriate areas in the cache memory 1740. The Fiber Channel I/O controller 1720 is then configured to write the data to the appropriate region of disk storage that is in communication with the controller 1720. Otherwise a READ from the disk must be done before the WRITE to obtain the required data from the appropriate disk.

The proto-SCSI acknowledge generator 1730 is responsible for generating the proto-SCSI responses. There are three possible sources which can generate proto-SCSI responses, each of which supplies a SEST index: the processor 1738, Fiber Channel I/O controller 1720, and the cache memory 1740. For all transfers an identification which allows the proto-SCSI request to be tied up with the acknowledge, along with status information, are returned the proto-SCSI acknowledge interface 1734.

Fig. 17B is a detailed block diagram showing control flow in the storage module of Fig. 3. When the proto-SCSI requests are received over the proto-SCSI input 1700 by the proto-SCSI request processor 1702, it is assigned a unique identifier (called the SEST index). The information relating to this request is stored in a SEST information table, and if this is a WRITE request then the WRITE data which is also provided on the proto-SCSI input 1700 is stored in the WRITE buffer memory 1736. The SEST index is then written into the proto-SCSI request queue 1704.

The cache controller 1706 takes entries out of the proto-SCSI request queue 1704 and the used buffer queue 1708. When an entry is taken out of the proto-SCSI request queue 1704 the information relating to this SEST index is read out of the SEST information table. The cache controller 1706 then works out which disk blocks are required for this transfer and translates this into cache buffer locations using a hash lookup of the disk block number and the disk device to be accessed. If any of the buffers in the write buffer memory 1736 required for this transfer are currently being used by other transfers, then the SEST index is put into the outstanding request queue 1710 awaiting completion of the other transfers. Otherwise if this is a READ transfer and all of the required buffers are in the cache then the SEST index is put into the cached READ queue 1712. Otherwise the SEST index is written into the storage request queue 1714. A possible enhancement to this algorithm is to allow multiple READs of the same buffer to be in progress provided that the buffer is currently cached.

When an entry is taken out of the used buffer queue 1708, a check is made as to whether any requests were waiting for this buffer to become available. This is done by searching through the outstanding request queue, starting with the oldest requests. If a request is found which was waiting for this buffer to become available then the buffer is allocated to that request. If the request now has all the buffers required for this transfer, then the SEST index is written into the storage request queue 1714 and this request is removed from the outstanding request queue 1710. Otherwise the request is left in the outstanding request queue 1710.

The exchange request (ERQ) generator 1716 takes entries out of the storage request queue 1714 and the partial WRITE queue 1718. When a SEST index is read out of either queue then the information relating to this SEST index is read out of the SEST information table. If it is a READ transfer then the Fiber Channel I/O controller 1720 is configured to read the data from the appropriate disk. If it is a WRITE transfer and all the buffers to be written are currently cached, or the data to be written completely fills the buffers to be written, then the WRITE can be performed immediately. The data to be written is copied from WRITE buffer memory 1736 to the appropriate areas in the cache buffers. The Fiber Channel I/O

controller 1720 is then configured to write the data to the appropriate disk. Otherwise, as mentioned above with respect to Fig. 17A, we need to do a READ from the disk before we do a WRITE and initiate a READ of the required data from the appropriate disk.

5 The IMQ processor 1722 takes messages from the inbound message queue 1724. This is a queue of transfers which the Fiber Channel I/O controller 1720 has completed or transfers which have encountered a problem. If there was a problem with the Fiber Channel transfer then the IMQ processor 1722 will pass the message on to the processor via the processor message queue 1726 to allow it to do the
10 appropriate error recovery. If the transfer was acceptable, then the SEST information is read out for this SEST index. If this transfer was a READ transfer at the start of a WRITE transfer, then the SEST index is written into the partial WRITE queue 1718. Otherwise it is written into the storage acknowledge queue 1728.

 As mentioned with respect to Fig. 17A, the proto-SCSI acknowledge
15 generator 1730 is responsible for generating the proto-SCSI responses. Again, there are three possible sources which can generate proto-SCSI responses, each of which supplies a SEST index.

 The processor acknowledge queue 1732 is used by the processor 1738 to pass requests which generated errors and which had to be sorted out by the processor
20 1738 back to the hardware once they have been sorted out. The storage acknowledge queue 1728 is used to pass back Fiber Channel requests which have completed normally. The cached READ queue 1712 is used to pass back requests where all the READ data required is already in the cache and so no Fiber Channel accesses are required.

25 When there is an entry in any of these queues the SEST index is read out. The SEST information for this index is then read. For all transfers an identification which allows the proto-SCSI request to be tied up with the acknowledge, along with status information, are returned across the proto-SCSI acknowledge interface 1734. For a READ the read data is also returned across the proto-SCSI acknowledge
30 interface 1734.

Once the proto-SCSI transfer has been completed, the addresses of all the buffers associated with this transfer are written into the used buffer queue 1708. Any WRITE buffer memory used in this transfer is also returned to the pool of free WRITE buffer memory.

5 Fig. 18 is a block diagram illustrating use of a storage module, such as illustrated in Fig. 3, in connection with a computer system having file storage. Here the storage module 1801 acts as a fiber channel host bus adapter and driver for the computer system, which includes microprocessor 1802, memory 1803, a peripheral device, such as a video system 1805, and storage devices 1809, 1810, and 1811. The storage module 1801 is coupled to the microprocessor 1802 and the computer
10 memory 1803 via the PCI bridge 1804 over PCI bus 1807. The storage module 1801 receives requests from the PCI bus and processes the requests in the manner described above with respect to Figs. 17A and 17B. The storage module 1801 accesses the storage devices 1809, 1810, and 1811 via the storage device access
15 interface 1808.

Fig. 19 is a block diagram illustrating scalability of embodiments of the present invention, and, in particular, an embodiment wherein a plurality of network subsystems and service subsystems are employed utilizing expansion switches for establishing communication among ports of successive subsystems
20 and/or modules. To allow extra network connections, to increase the bandwidth capabilities of the unit, and to support a larger number of storage elements, in this embodiment, expansion switches 1901, 1902, 1903 are used to interface a number of modules together. The expansion switch routes any connection from a module on one side of the expansion switch to any module on the other side. The expansion
25 switch is non-blocking, and may be controlled by an intelligent expansion switch control module that takes in a number of inputs and decides upon the best route for a particular connection.

In the embodiment of Fig. 19, the overall system shown utilizes a plurality of network subsystems shown in column 1921 including network subsystem 1904 and
30 similar subsystems 1908 and 1912. There are also a plurality of service subsystems, which are here realized as a combination of file access modules (in column 1922), file

system modules (in column 1923), and storage modules (in column 1924). Between each column of modules (and between the network subsystems column and the file access modules column) is a switch arrangement, implemented as the file access protocol expansion switch 1901, the storage access expansion switch 1902, and the proto-SCSI protocol expansion switch 1903. At the file access protocol level, the expansion switch 1901 dynamically allocates incoming network connections from the network subsystem 1904 to particular file access modules 1905 depending on criteria including the existing workload of each of the file access modules 1905.

At the storage access protocol level, the expansion switch 1902 dynamically allocates incoming file access connections from the file access modules 1905 to particular file system modules 1906 depending on criteria including the existing workload of the file system modules 1906.

At the proto-SCSI protocol level, the expansion switch 1903 dynamically allocates incoming file system connections to particular storage modules 1907 depending on criteria including the physical location of the storage element.

Alternatively, the items 1901, 1902, and 1903 may be implemented as buses, in which case each module in a column accepting an input signal communicates with other modules in the column to prevent duplicate processing of the signal, thereby freeing the other modules to handle other signals. Regardless whether the items 1901, 1902, and 1903 are realized as buses or switches, it is within the scope of the present invention to track the signal processing path through the system, so that when a response to a file request is involved, the appropriate header information from the corresponding request is available to permit convenient formatting of the response header.

Claims:

1. Apparatus for handling service requests over a network, wherein the network utilizes a protocol, the apparatus comprising:
- a. a network subsystem for receiving and transmitting network service requests using the network protocol; and
 - b. a service subsystem, coupled to the network subsystem, for satisfying the network service requests;
- wherein at least one of the network subsystem and the service subsystem is hardware-implemented.
- 10 2. Apparatus for handling service requests over a network, wherein the network utilizes a protocol, the apparatus comprising:
- a. a network subsystem for receiving and transmitting network service requests using the network protocol; and
 - b. a service subsystem, coupled to the network subsystem, for satisfying
- 15 the network service requests wherein one of the network subsystem and the service subsystem is hardware-implemented and the other of the network subsystem and the service subsystem is hardware accelerated.
- 20 3. Apparatus for handling service requests over a network, wherein the network utilizes a protocol, the apparatus comprising:
- a. a hardware-implemented network subsystem for receiving and transmitting network service requests using the network protocol; and
 - b. a hardware-implemented service subsystem, coupled to the network subsystem, for satisfying the network service requests.
- 25 4. Apparatus for handling service requests over a network, wherein the network utilizes a protocol, the apparatus comprising:
- a. a network subsystem for receiving and transmitting network service requests using the network protocol; and
 - b. a hardware-accelerated service subsystem, coupled to the network
- 30 subsystem, for satisfying the network service requests.

5. Apparatus for handling service requests over a network, wherein the network utilizes a protocol, the apparatus comprising:
- a. a hardware-accelerated network subsystem for receiving and transmitting network service requests using the network protocol; and
 - 5 b. a service subsystem, coupled to the network subsystem, for satisfying the network service requests;
- wherein the service requests include one of reading and writing data to long-term electronic storage.
6. Apparatus according to any of claims 1 through 4, wherein the service requests include one of reading and writing data to long-term electronic storage.
- 10 7. Apparatus according to any of claims 5 or 6, wherein the long-term storage is network disk storage accessible to computers over the network.
8. Apparatus according to any of claims 5 or 6, wherein the long-term storage is local disk storage that is accessible to a local computer but not to any other
- 15 computers over the network.
9. Apparatus according to any of claims 5 or 6, wherein the long-term storage is associated with the provision of E-Mail service over the network.
10. Apparatus according to any of claims 5 or 6, wherein the long-term storage provides access to web pages over the network.
- 20 11. Apparatus according to any of claims 1 through 5, wherein the service requests may involve access of data in a storage system, and the service subsystem also includes a hardware-implemented module for managing storage of the data in the storage system.
12. Apparatus according to claim 11, such apparatus being a file server, wherein
- 25 the data in the storage system are arranged in files, the service requests may involve requests for files in the storage system, and the service subsystem also includes a hardware-implemented module for managing a file system associated with the storage system.
13. Apparatus according to claim 12, wherein the protocol includes a file system
- 30 protocol, and the file system protocol defining operations including file read and file write.

14. Apparatus according to claim 11, such apparatus being a web server, wherein the data in the storage system may include web pages, and the service requests may involve requests for web pages in the storage system.
15. Apparatus according to any of claims 1, 2, 3, 4, 5, 11, 12, or 14, wherein the
5 protocol includes IP.
16. Apparatus according to any of claims 11 or 12, wherein the storage system has a storage protocol and the service subsystem includes a hardware-implemented module for interfacing with the storage system.
17. A subsystem for receiving and transmitting data over a network, the
10 network using a protocol having at least one of layers 3 and 4, the subsystem comprising:
- a. a receiver that receives encapsulated data from the network and de-encapsulates such data in accordance with the protocol; and
 - b. a transmitter that encapsulates data in accordance with the protocol
15 and transmits the encapsulates data over the network;
- wherein at least one of the receiver and the transmitter is hardware-implemented.
18. A subsystem for receiving and transmitting data over a network, the network using a protocol having at least one of layers 3 and 4, the subsystem comprising:
- 20 a. a receiver that receives encapsulated data from the network and de-encapsulates such data in accordance with the protocol; and
 - b. a transmitter that encapsulates data in accordance with the protocol and transmits the encapsulates data over the network;
- wherein at least one of the receiver and the transmitter is hardware-accelerated.
- 25 19. A subsystem according to any of claims 17 or 18 , wherein the protocol is TCP/IP.
20. A subsystem according to claim 19, wherein the data are received over the network in packets, each packet having a protocol header, the subsystem further comprising:

a connection identifier that determines a unique connection from information contained within the protocol header of each packet received by the receiver.

21. A subsystem according to any of claims 19 through 20, wherein encapsulated data are associated with a network connection, the subsystem further comprising:
5 a memory region, associated with the network connection, that stores the state of the connection.

22. A service subsystem for interfacing a storage arrangement with a network over which may be generated a storage access request, the subsystem comprising:
10 a. a service module that receives network service requests and fulfills such service requests and in doing so may issue data storage access requests;
b. a file system module, coupled to the service module, that receives data storage access requests from the service module and fulfills such storage access requests and in doing so may issue storage arrangement access requests;
15 c. a storage module, coupled to the file system module, that receives storage arrangement access requests from the file system module and controls the storage arrangement to fulfill such storage arrangement access requests;
wherein at least one of the modules is hardware-implemented.

23. A service subsystem for interfacing a storage system with a network over
20 which may be generated a storage access request, the subsystem comprising:
a. a service module that receives network service requests and fulfills such service requests and in doing so may issue data storage access requests;
b. a file system module, coupled to the service module, that receives data storage access requests from the service module and fulfills such storage access
25 requests and in doing so may issue storage arrangement access requests;
c. a storage module, coupled to the file system module, that receives storage arrangement access requests from the file system module and controls the storage arrangement to fulfill such storage arrangement access requests;
wherein at least one of the modules is hardware-accelerated.

30 24. A service subsystem according to any of claims 22 and 23, wherein the service module includes:

- i. a receive control engine that receives network service requests, determines whether such requests are appropriate, and if so, responds if information is available, and otherwise issues a data storage access request; and
- ii. a transmit control engine that generates network service responses based on instructions from the receive control engine, and, in the event that there is a data storage access response to the data storage access request, processes the data storage access response;
- wherein at least one of the engines is hardware-implemented.
25. A service subsystem according to any of claims 22 and 23, wherein the service module includes:
- i. a receive control engine that receives network service requests, determines whether such requests are appropriate, and if so, responds if information is available, and otherwise issues a data storage access request; and
- ii. a transmit control engine that generates network service responses based on instructions from the receive control engine, and, in the event that there is a data storage access response to the data storage access request, processes the data storage access response;
- wherein at least one of the engines is hardware-accelerated.
26. A service subsystem according to any of claims 22 and 23, wherein the service subsystem is integrated directly in the motherboard of a computer.
27. A service subsystem according to any of claims 22 and 23, wherein the service subsystem is integrated into an adapter card that may be plugged into a computer.
28. A service module that receives network service requests and fulfills such service requests, such module comprising:
- a. a receive control engine that receives network service requests, determines whether such requests are appropriate, and if so, responds if information is available, and otherwise issues a data storage access request; and
- b. a transmit control engine that generates network service responses based on instructions from the receive control engine, and, in the event that there is

a data storage access response to the data storage access request, processes the data storage access response;

wherein at least one of the engines is hardware-implemented.

29. A service module that receives network service requests and fulfills such
5 service requests, such module comprising:
- a. a receive control engine that receives network service requests, determines whether such requests are appropriate and if so, responds if information is available, and otherwise issues a data storage access request; and
 - b. a transmit control engine that generates network service responses
10 based on instructions from the receive control engine, and, in the event that there is a data storage access response to the data storage access request, processes the data storage access response;
wherein at least one of the engines is hardware-accelerated.
30. A service module according to any of claims 28 or 29, wherein the network
15 service requests are in the CIFS protocol.
31. A service module according to any of claims 28 or 29, wherein the network service requests are in the SMB protocol.
32. A service module according to any of claims 28 or 29, wherein the network service requests are in the HTTP protocol.
- 20 33. A service module according to any of claims 28 or 29, wherein the network service requests are in the NFS protocol.
34. A service module according to any of claims 28 or 29, wherein the network service requests are in the FTP protocol.
35. A service module according to any of claims 28 or 29, wherein the network
25 service requests are in the SMTP protocol.
36. A service module according to any of claims 28 or 29, further comprising:
an authentication engine that determines whether a network request received by the receiver has been issued from a source having authority to issue the request.
- 30 37. A service module according to any of claims 28 through 36, wherein the authentication engine determines whether a network request received by the

receiver has been issued from a source having authority to perform the operation requested.

38. A service module according to any of claims 28 through 37 wherein the service module is integrated directly in the motherboard of a computer.

5 39. A service module according to any of claims 28 through 37 wherein the service module is integrated into an adapter card that may be plugged into a computer.

40. A file system module that receives data storage access requests and fulfills such data storage access requests, the module comprising:

10 a. a receiver that receives and interprets such data storage access requests and in doing so may issue storage device access requests;

b. a transmitter, coupled to the receiver, that constructs and issues data storage access responses, wherein such responses include information when appropriate based on responses to the storage device access requests;

15 wherein at least one of the receiver and the transmitter is hardware-implemented.

41. A file system module that receives data storage access requests and fulfills such data storage access requests, the module comprising:

a. a receiver that receives and interprets such data storage access requests and in doing so may issue storage device access requests;

20 b. a transmitter, coupled to the receiver, that constructs and issues data storage access responses, wherein such responses include information when appropriate based on responses to the storage device access requests;

wherein at least one of the receiver and the transmitter is hardware-accelerated.

25 42. A file system module according to any of claims 40 or 41, wherein the storage device access requests are consistent with the protocol used by a storage device to which the module may be coupled.

43. A file system module according to claim 42, wherein the protocol is NTFS.

44. A file system module according to claim 42, wherein the protocol is HPFS.

45. A file system module according to claim 42, wherein the protocol is FAT.

30 46. A file system module according to claim 42, wherein the protocol is FAT16.

47. A file system module according to claim 42, wherein the protocol is FAT32.

48. A file system module according to claim 42, further comprising:
file table cache, coupled to the receiver, that stores a table defining the physical
location of files in a storage device to which the module may be coupled.
49. A file system module according to claim 48, wherein the protocol does not
5 require files to be placed in consecutive physical locations in a storage device.
50. A file system module according to claim 42, wherein the file system module
is integrated directly in the motherboard of a computer.
51. A file system module according to claim 42, wherein the file system module
is integrated into an adapter card that may be plugged into a computer.
- 10 52. A storage module that receives storage device access requests from a request
source and communicates with a storage device controller to fulfill such storage
access requests, the module comprising:
- 15 a. a storage device request interface that receives such storage device
access requests and translates them into a format suitable for the storage device
controller; and
- b. a storage device acknowledge interface that takes the responses from
the storage device controller and translates such responses into a format suitable for
the request source;
- wherein at least one of the storage device request interface and the storage device
20 acknowledge interface is hardware-implemented.
53. A storage module that receives storage device access requests from a request
source and communicates with a storage device controller for a storage device to
fulfill such storage access requests, the module comprising:
- 25 a. a storage device request interface that receives such storage device
access requests and translates them into a format suitable for the storage device
controller; and
- b. a storage device acknowledge interface that takes the responses from
the storage device controller and translates such responses into a format suitable for
the request source;
- 30 wherein at least one of the storage device request interface and the storage device
acknowledge interface is hardware-accelerated.

54. A storage module according to any of claims 52 or 53, further comprising:
a cache controller that maintains a local copy of a portion of data contained on the
storage device to allow fast-read access to the portion of data.
55. A storage module according to any of claims 52 or 53, wherein the storage
5 device request interface and the storage device acknowledge interface are coupled
to a port permitting communication with the storage device controller over a fiber-
optic channel.
56. A storage module according to any of claims 52 or 53, wherein the storage
device request interface and the storage device acknowledge interface are coupled
10 to a port permitting communication with the storage device controller utilizing a
SCSI-related protocol.
57. A storage module according to any of claims 52 or 53, wherein the storage
module is integrated directly in the motherboard of a computer.
58. A storage module according to any of claims 52 or 53, wherein the storage
15 module is integrated into an adapter card that may be plugged into a computer.
59. A system for interfacing a storage arrangement with a line on which may be
placed a storage access request, the system comprising:
- a. a service receive block, coupled to the storage arrangement, that
processes the storage access request, generates where necessary an access to the
20 storage arrangement, and causes the generation of a response;
 - b. a file table cache, coupled to the receive block, that stores a table
defining the physical location of files in the storage arrangement;
 - c. a service transmit block, coupled to the service receive block, for
transmitting the response;
- 25 wherein at least one of the service receive block and the service transmit block is
hardware-implemented.
60. A system for interfacing a storage arrangement with a line on which may be
placed a storage access request, the system comprising:
- a. a service receive block, coupled to the storage arrangement, that
30 processes the storage access request, and generates where necessary an access to the
storage arrangement;

b. a file table cache, coupled to the receive block, that stores a table defining the physical location of files in the storage arrangement;

c. a service transmit block, coupled to the service receive block and to the storage arrangement, for constructing and transmitting a response;

5 wherein at least one of the service receive block and the service transmit block is hardware-accelerated.

61. A system according to any of claims 59 or 60, further comprising:
response information memory, coupled to each of the service receive block and the
service transmit block, which memory stores information present in a header
10 associated the request, which information is used by the service transmit block in
constructing the response.

62. A system according to any of claims 59 or 60, wherein the storage access request is a network request.

63. A system according to any of claims 59 or 60, wherein the storage access
15 request is a generated by a local processor to which the line is coupled.

64. A process for handling storage access requests from multiple clients, the process comprising:

a. testing for receipt of a storage access request from any of the clients;
and

20 b. testing for completion of access to storage pursuant to any pending request;

wherein testing for receipt of a storage access request and testing for completion of access to storage are performed in a number of threads independent of the number of clients.

25 65. A process according to claim 64, wherein the process also includes, conditioned on a positive determination from testing for receipt of a request, processing the request that gave rise to the positive determination and initiating storage access pursuant to such request.

30 66. A process according to any of claims 64 or 65, wherein the process also includes, conditioned on a positive determination from testing for completion of

access to storage pursuant to a pending request, sending a reply to the client issuing such pending request.

67. A process according to any of claims 64, 65, or 66, wherein the number of threads is fewer than three.

5 68. A process according to claim 67, wherein the number of threads is one.

69. Scalable apparatus for handling service requests over a network, wherein the network utilizes a protocol, the apparatus comprising:

a. a first plurality of network subsystems for receiving and transmitting network service requests using the network protocol;

10 b. a second plurality of service subsystems, for satisfying the network service requests;

each one of the network subsystems and the service subsystems being one of hardware-implemented or hardware-accelerated; and

15 c. an interconnect coupling each of the first plurality of network subsystems to each of the second plurality of service subsystems.

70. Apparatus according to claim 69, wherein the interconnect is a switch.

71. Apparatus according to claim 69, wherein the interconnect is a bus.

72. A scalable service subsystem for interfacing a storage arrangement with a network over which may be generated a storage access request, the subsystem
20 comprising:

a. a first plurality of service modules that receive network service requests and fulfill such service requests and in doing so may issue data storage access requests;

25 b. a second plurality of file system modules that receive data storage access requests and fulfill such storage access requests and in doing so may issue storage arrangement access requests;

each one of the service modules and the file system modules being one of hardware-implemented or hardware-accelerated; and

30 c. an interconnect coupling each of the first plurality of service modules to each of the second plurality of file system modules.

73. A scalable service subsystem according to claim 72, wherein the interconnect is a switch.
74. A scalable service subsystem according to claim 72, wherein the interconnect is a bus.
- 5 75. A scalable service subsystem according to claim 72, further comprising:
- d. a third plurality of storage modules that receive storage arrangement access requests controls the storage arrangement to fulfill such storage arrangement access requests;
- each one of the storage modules being one of hardware-implemented or hardware-
- 10 accelerated;
- e. a second interconnect coupling each of the file system modules to each of the storage modules.
76. A scalable service subsystem according to claim 72, wherein each of the interconnect and the second interconnect is a switch.
- 15 77. A scalable service subsystem according to claim 72, wherein each of the interconnect and the second interconnect is a bus.

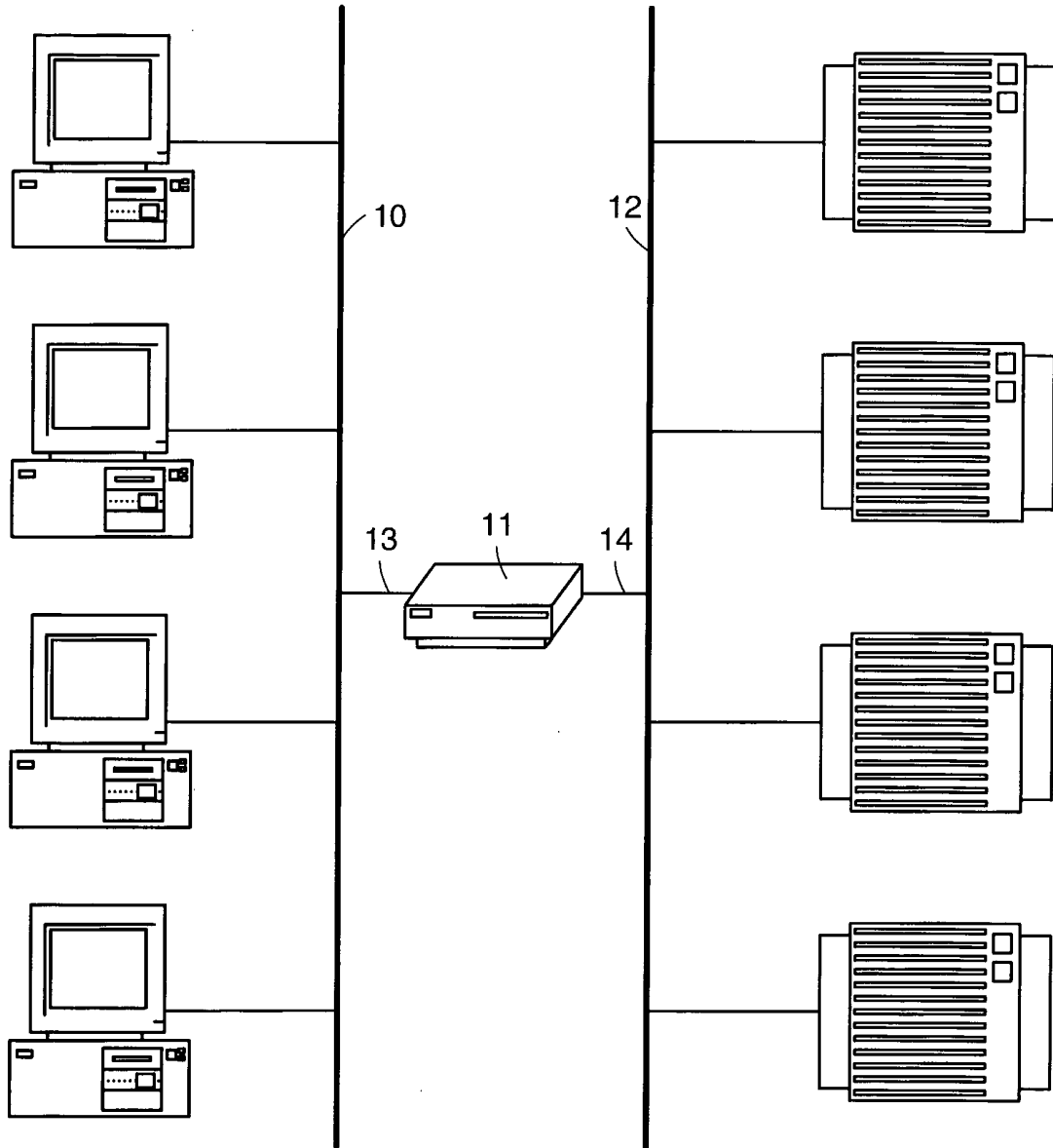


FIG. 1

2/21

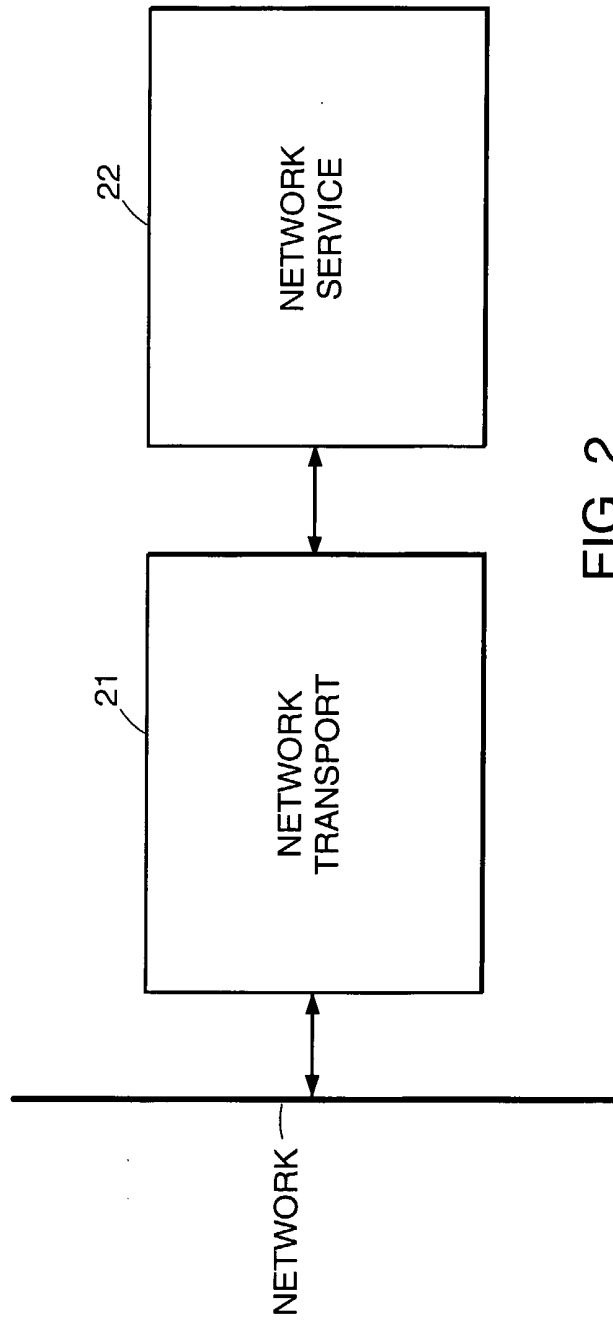


FIG. 2

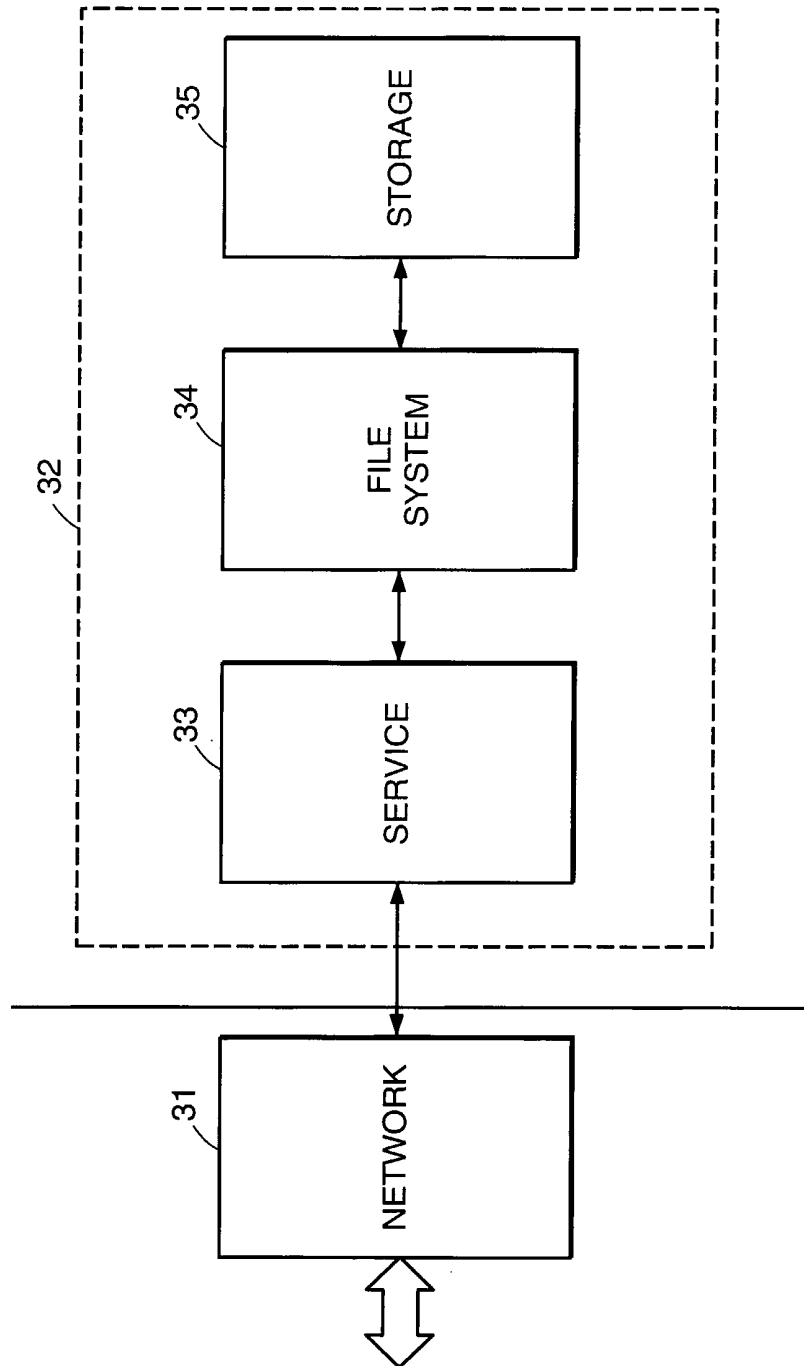


FIG. 3

4/21

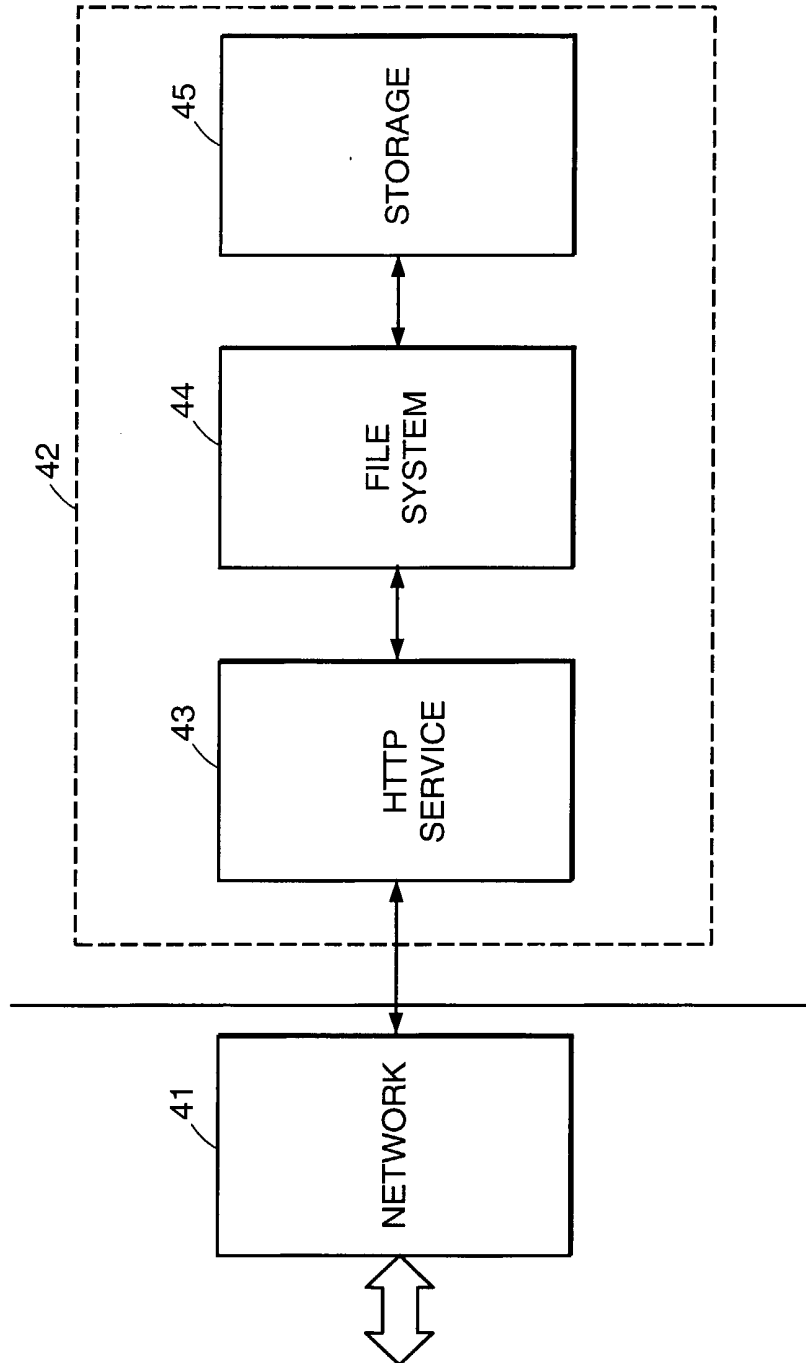


FIG. 4

5/21

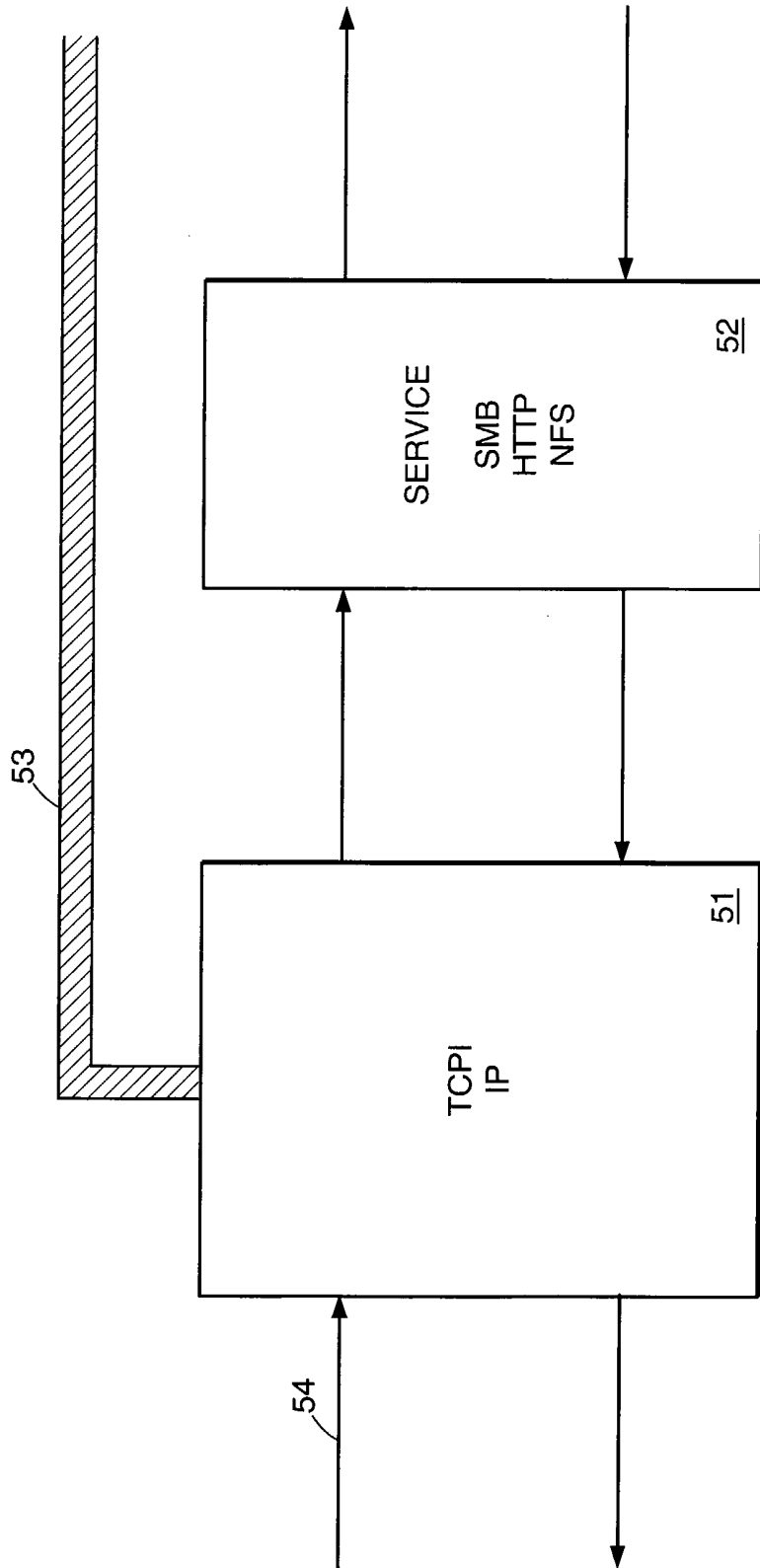


FIG. 5

6/21

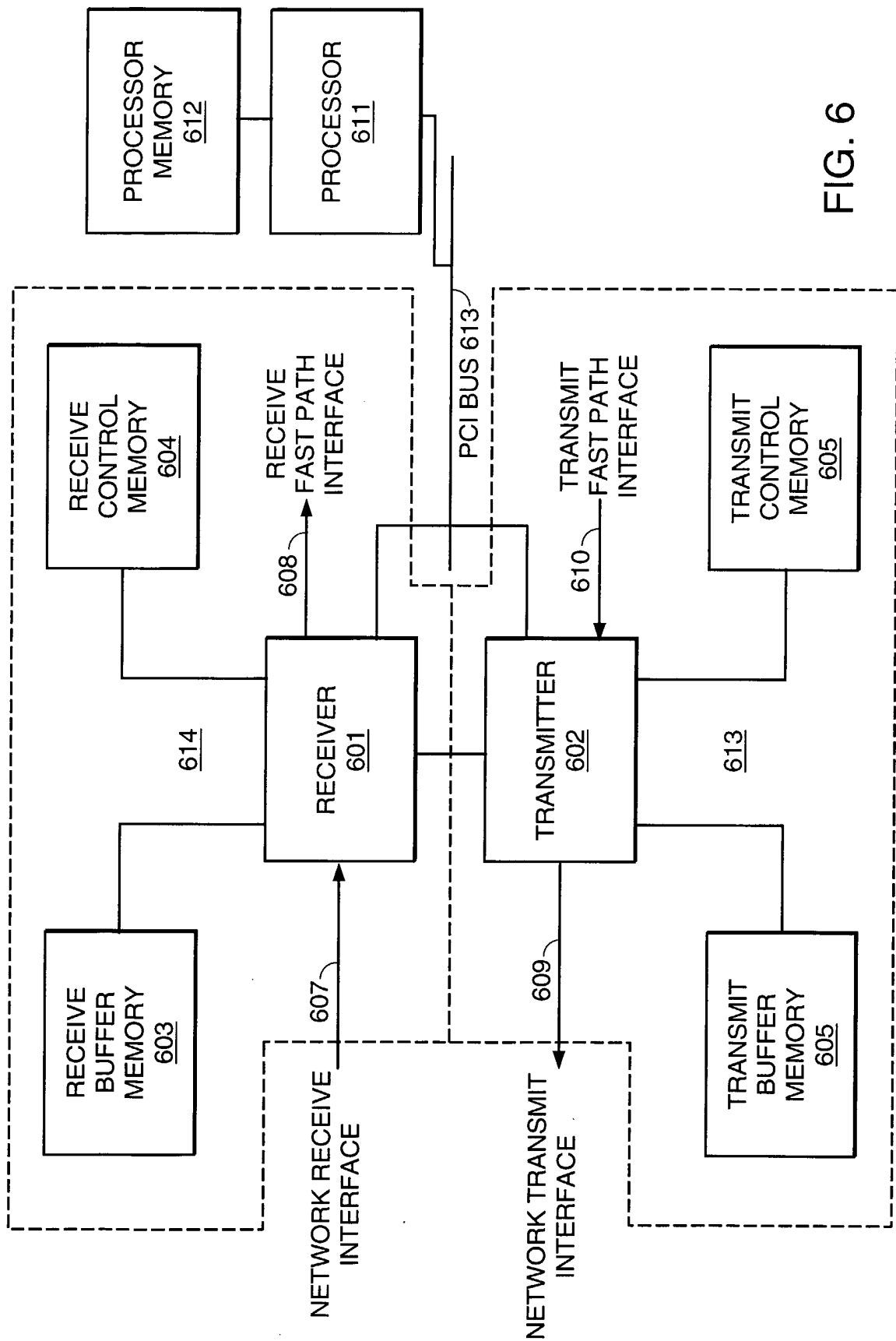


FIG. 6

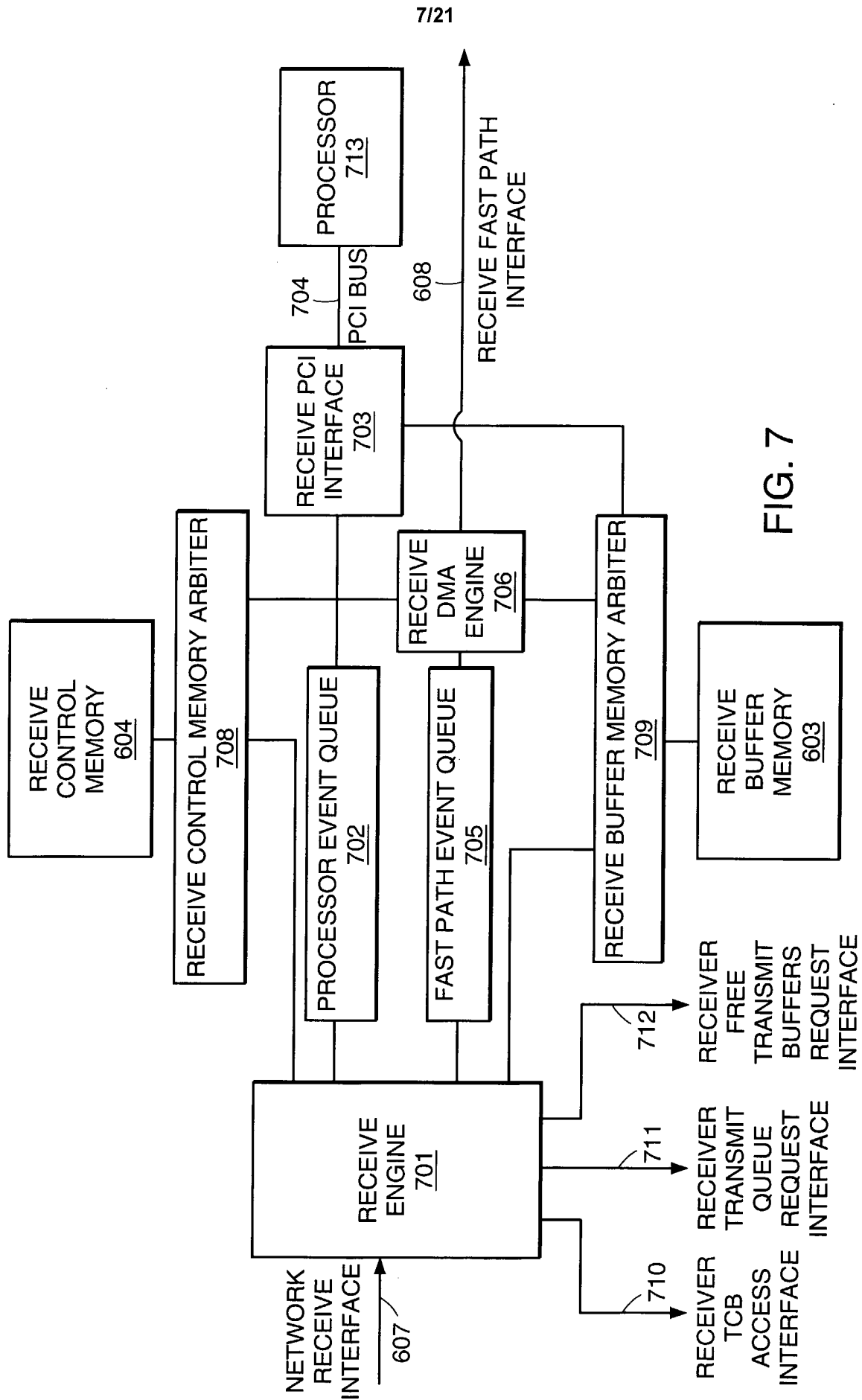


FIG. 7

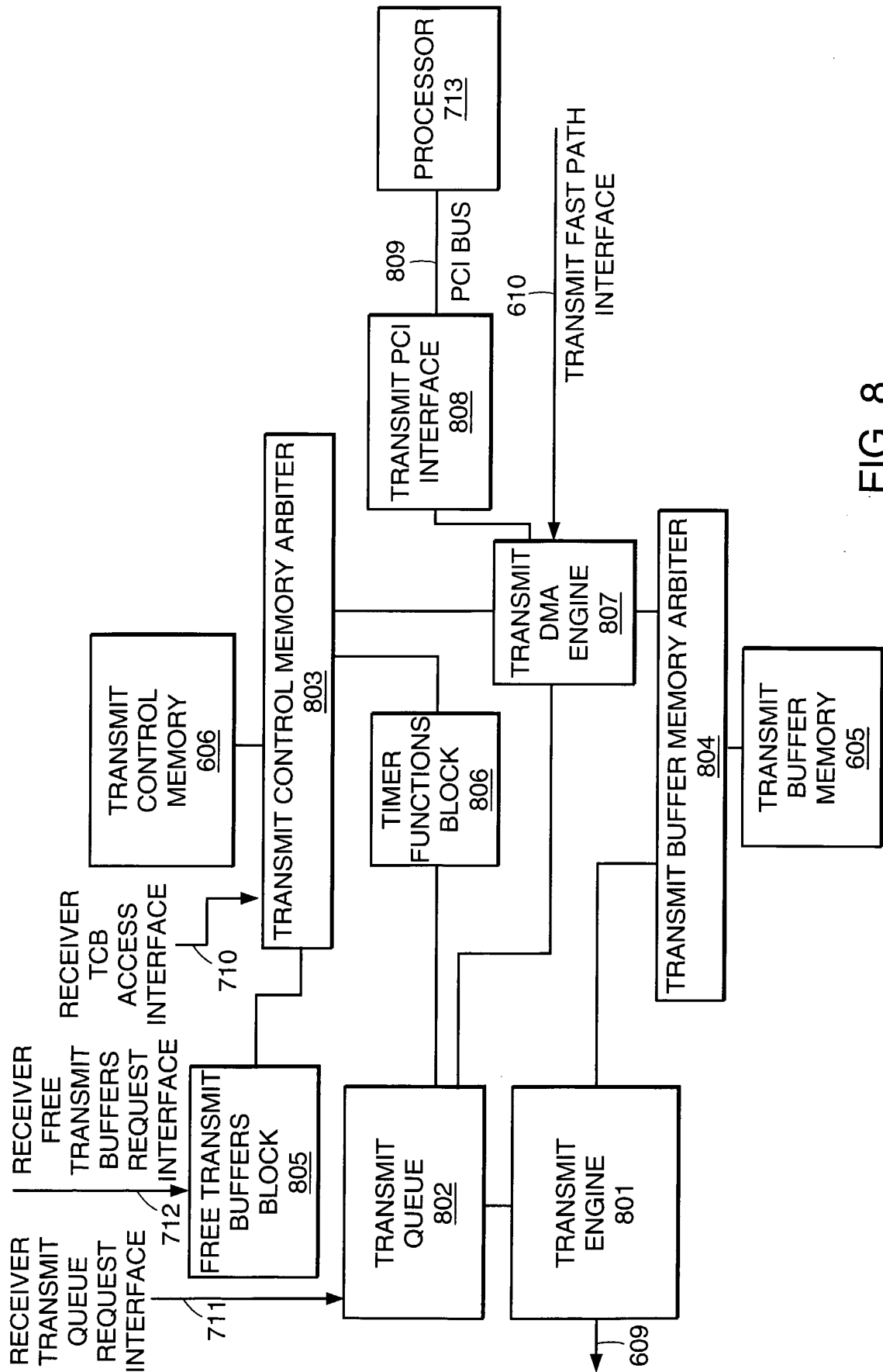


FIG. 8

9/21

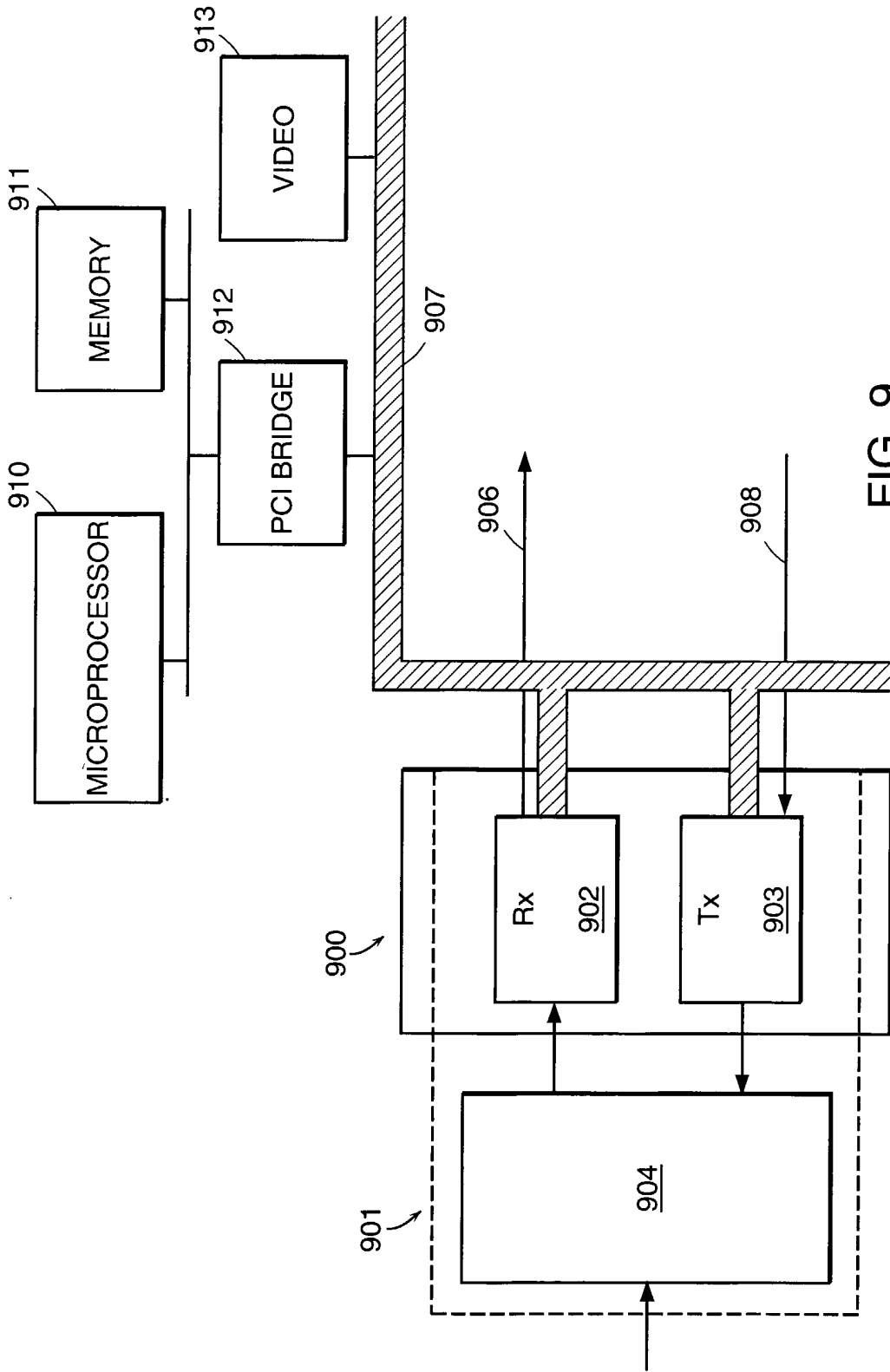


FIG. 9

10/21

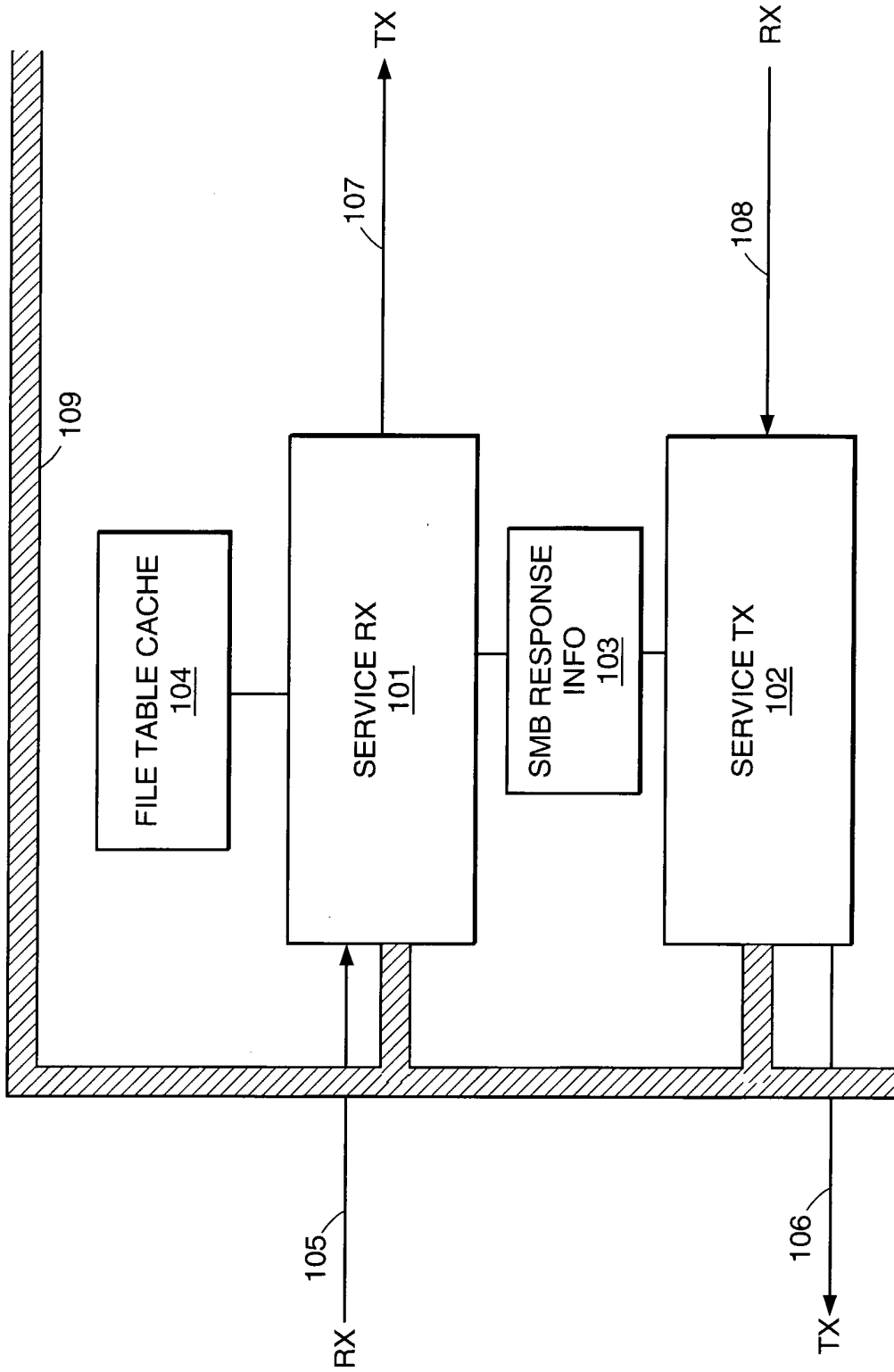


FIG. 10

11/21

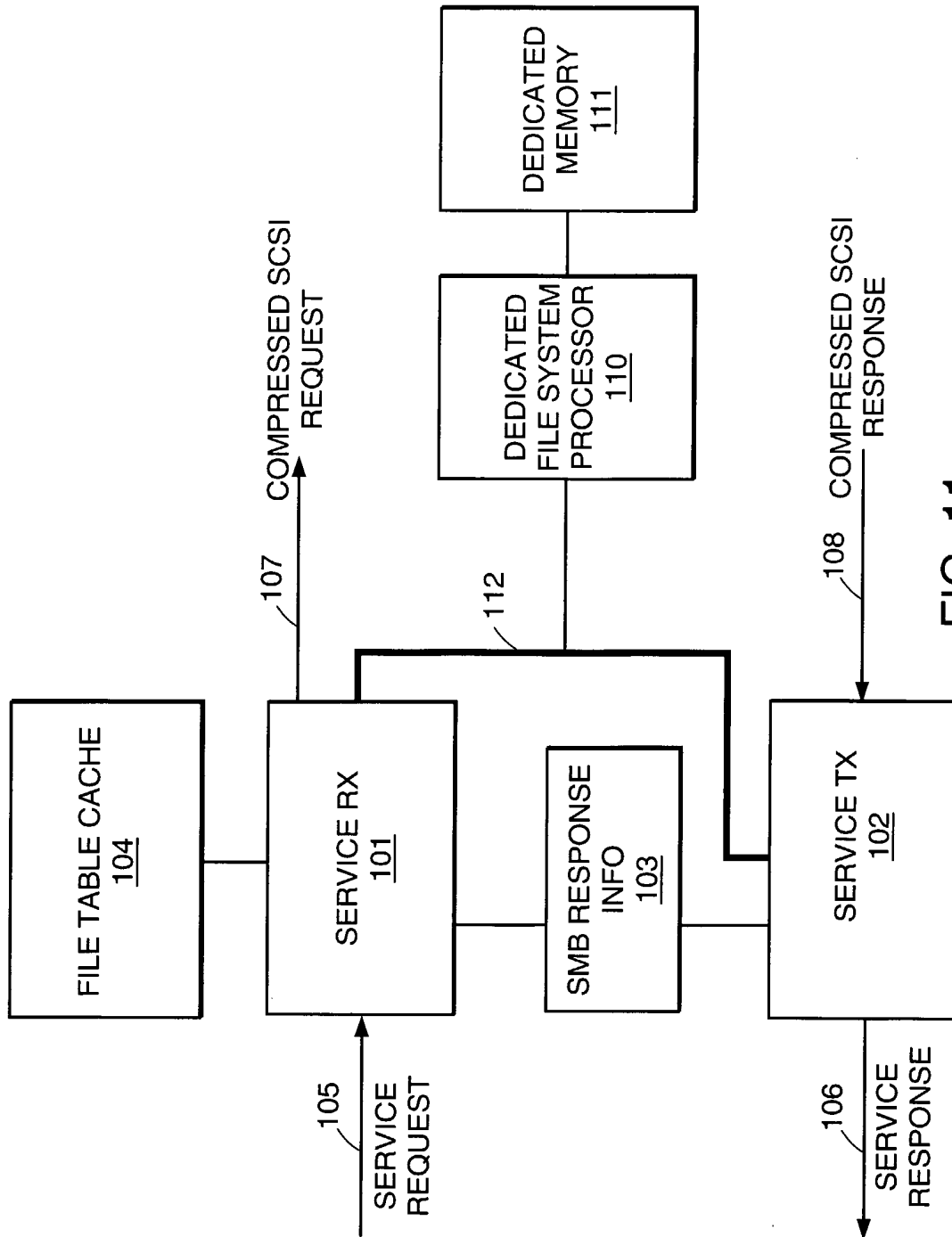


FIG. 11

12/21

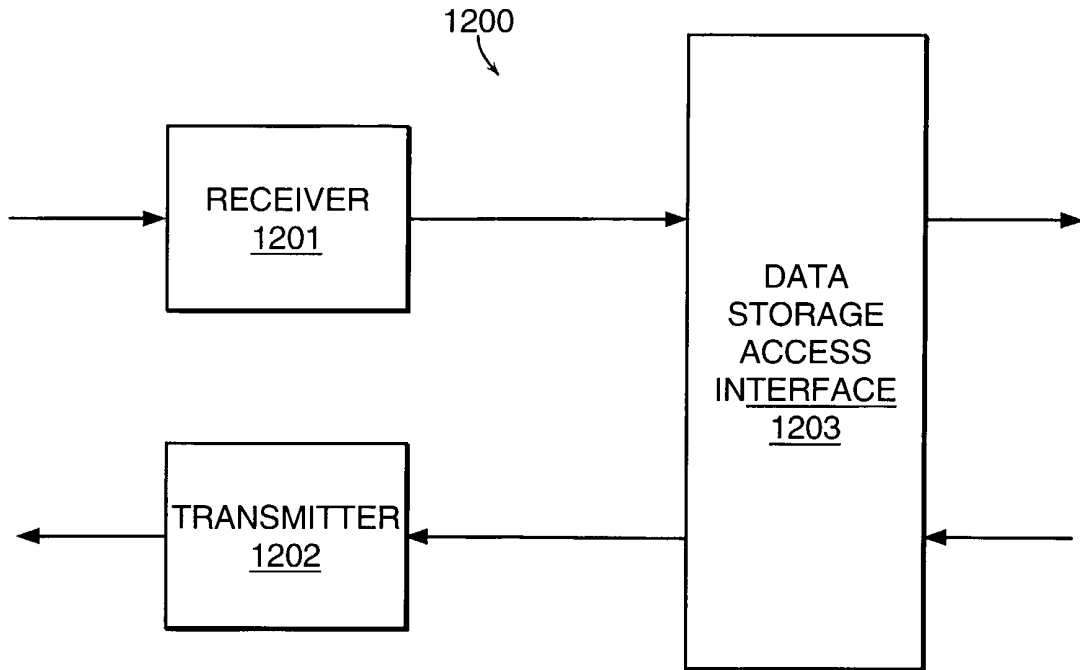


FIG. 12A

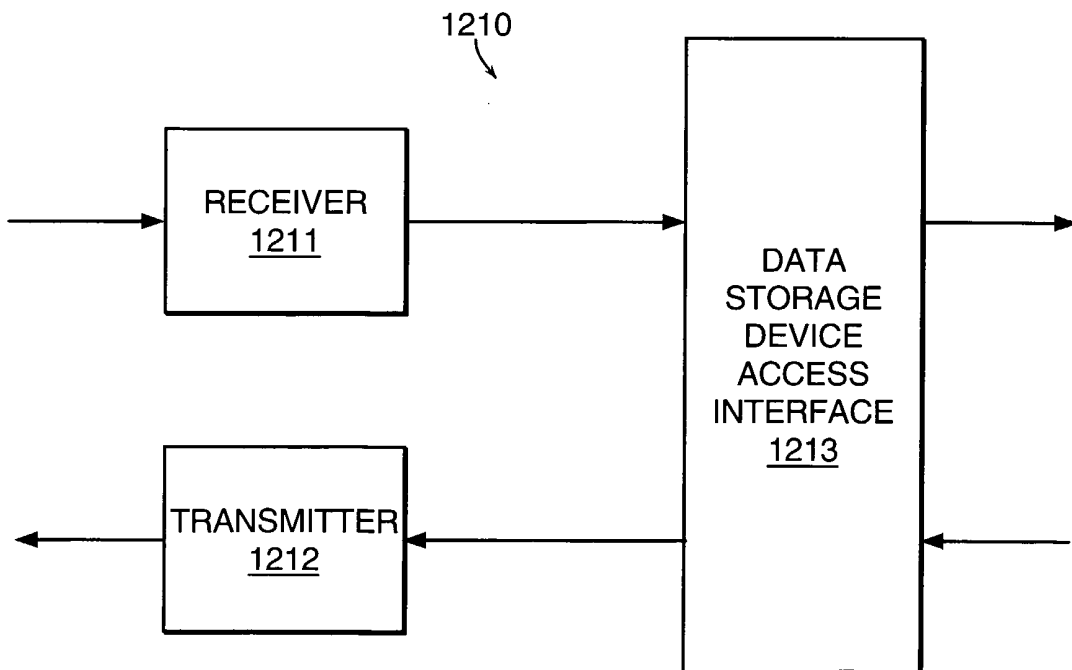
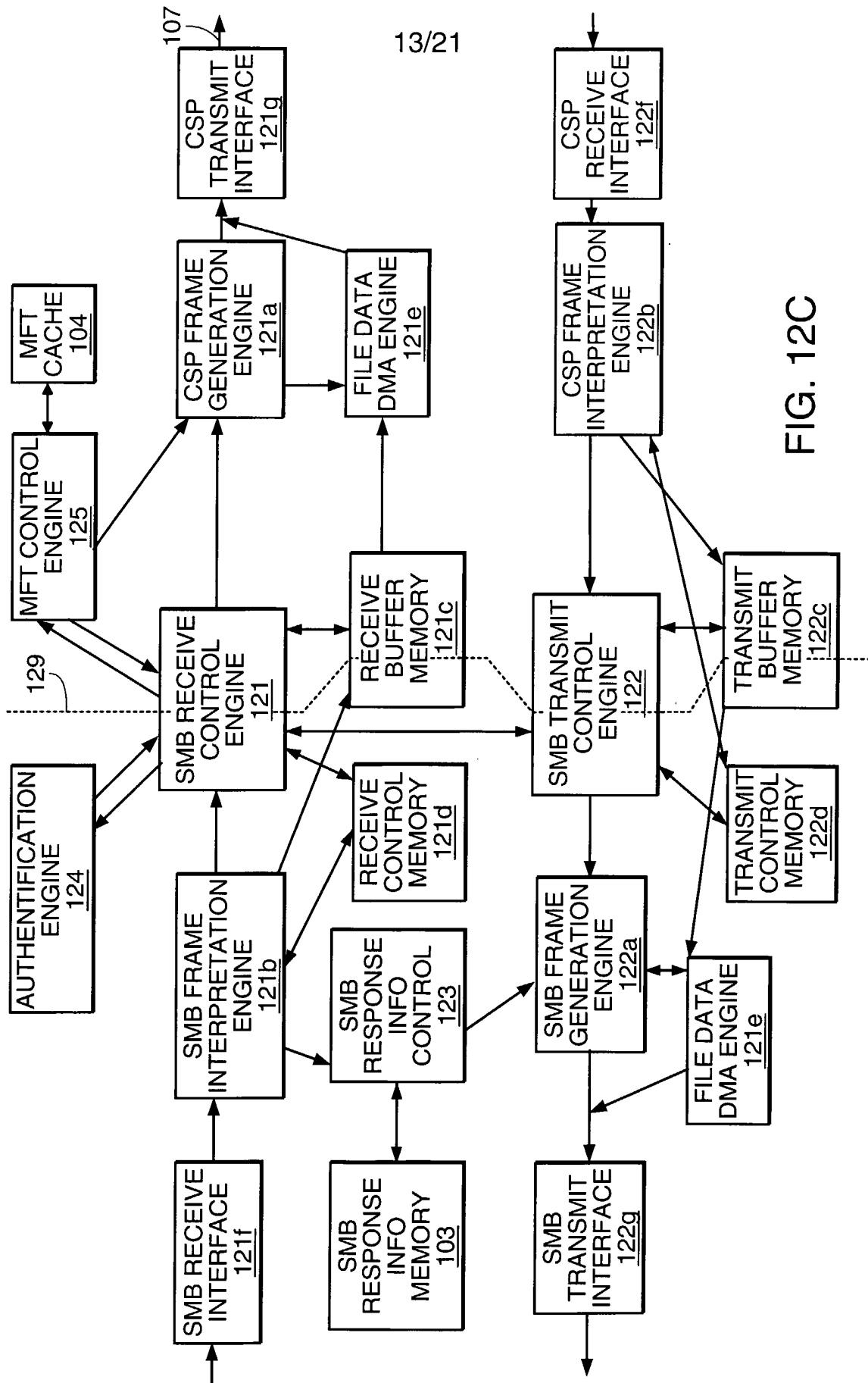


FIG. 12B



13/21

FIG. 12C

14/21

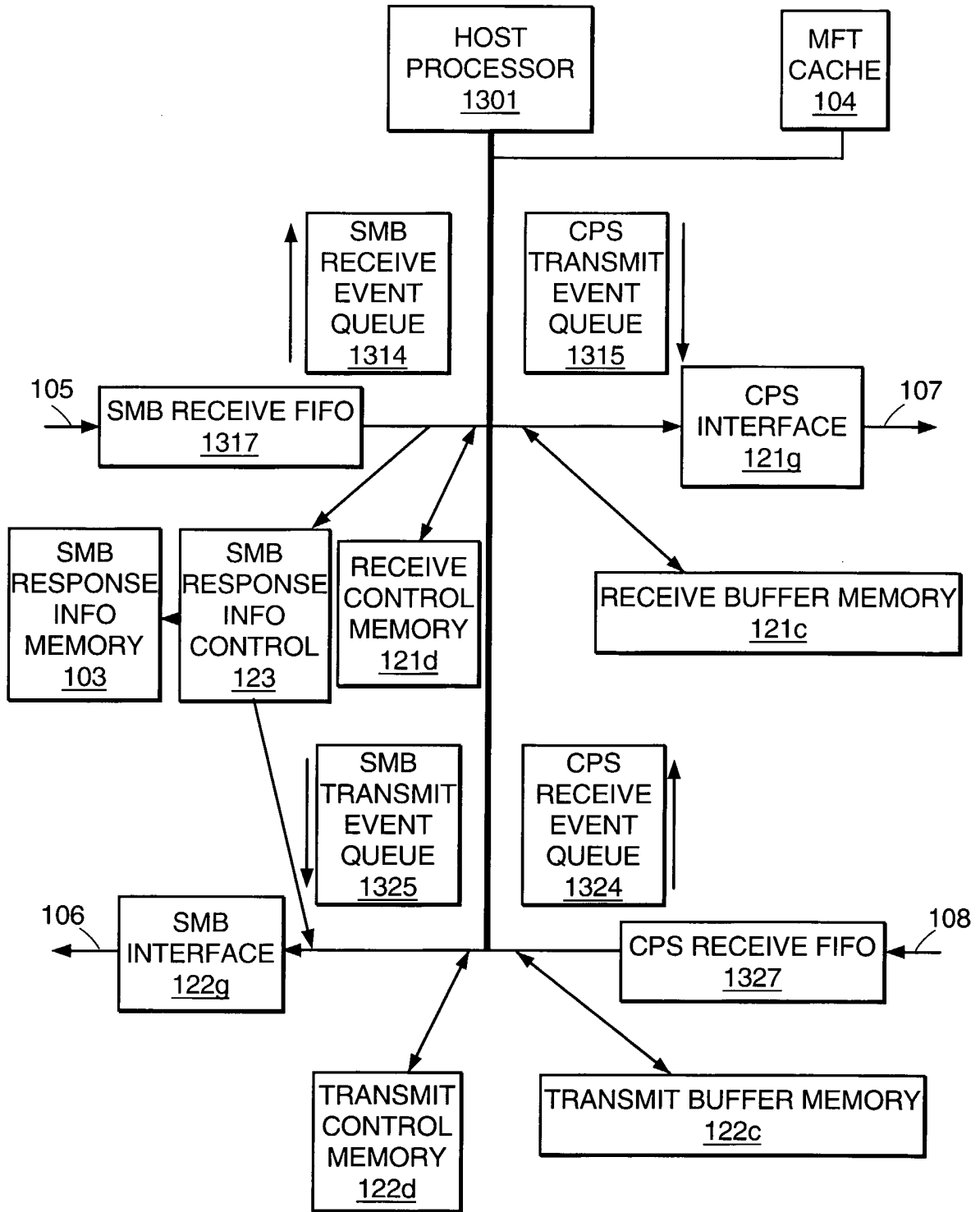


FIG. 13

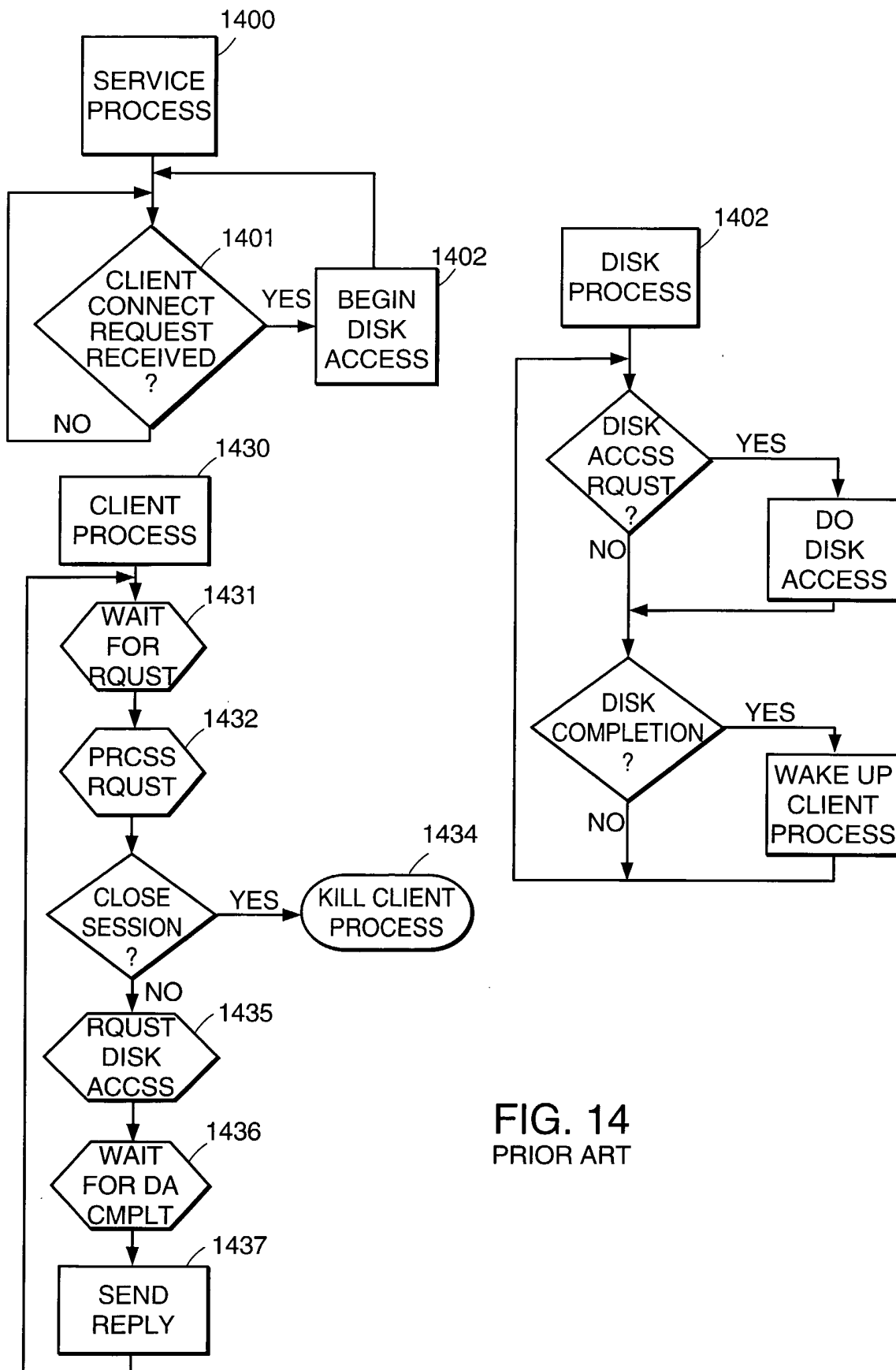


FIG. 14
PRIOR ART

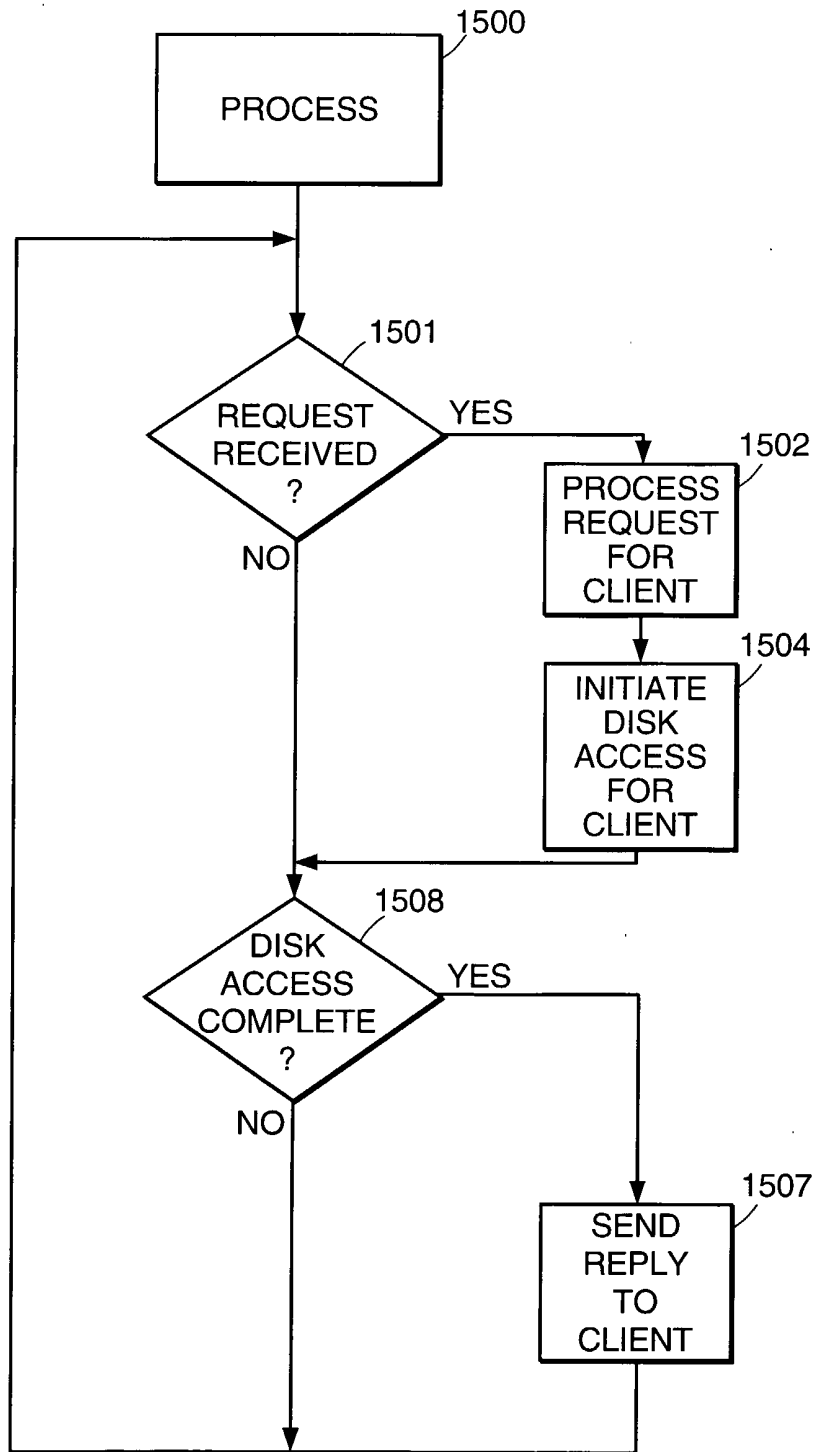
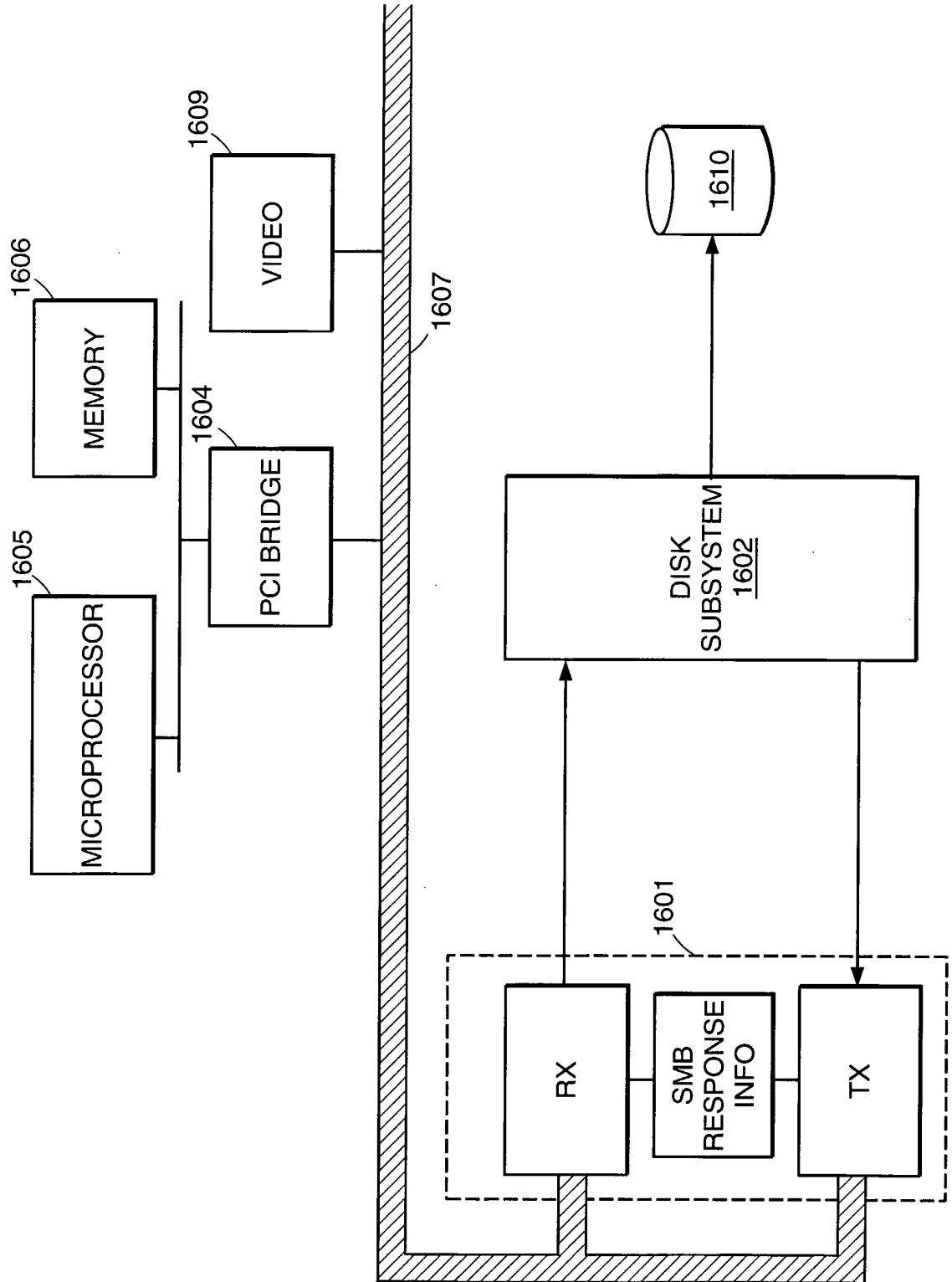


FIG. 15



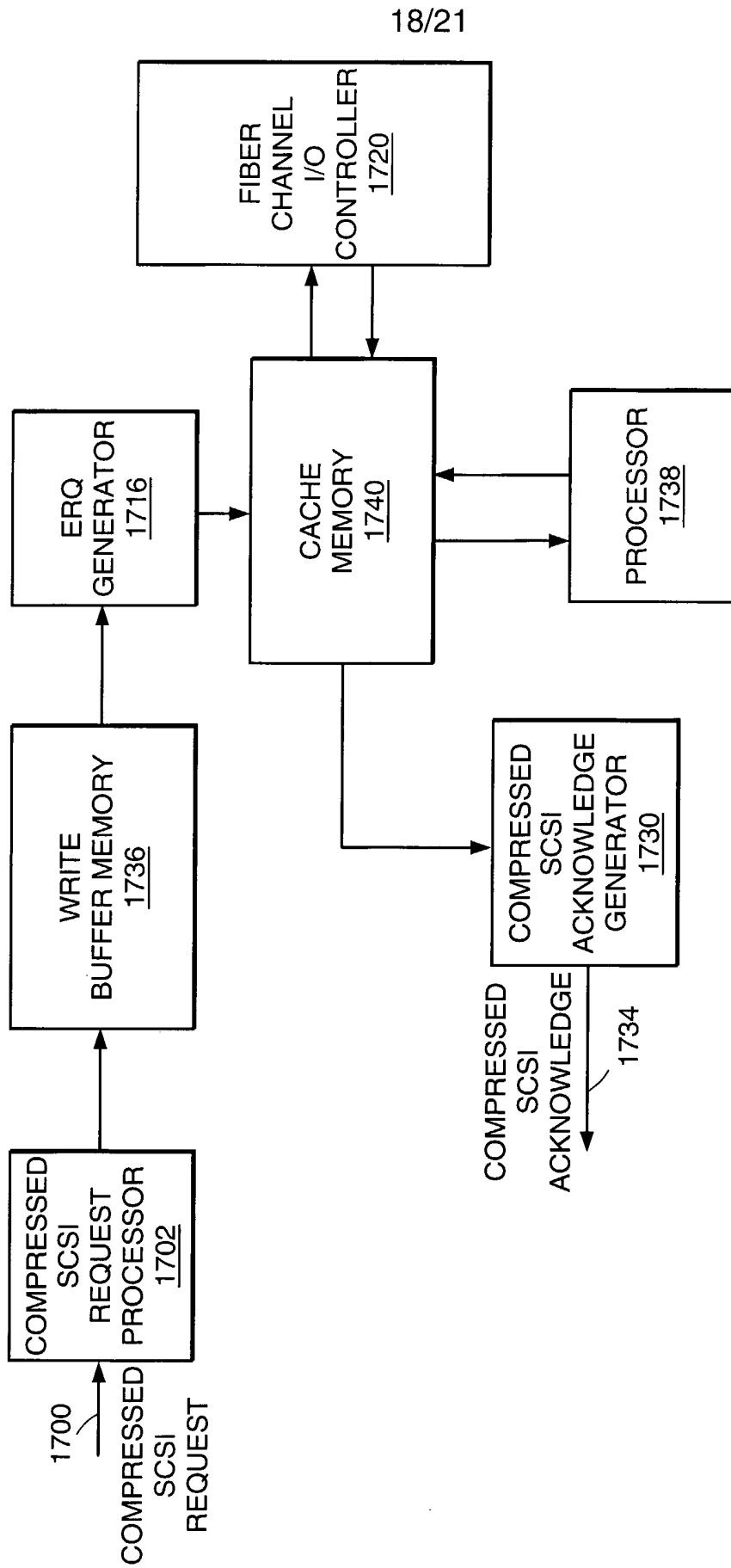


FIG. 17A

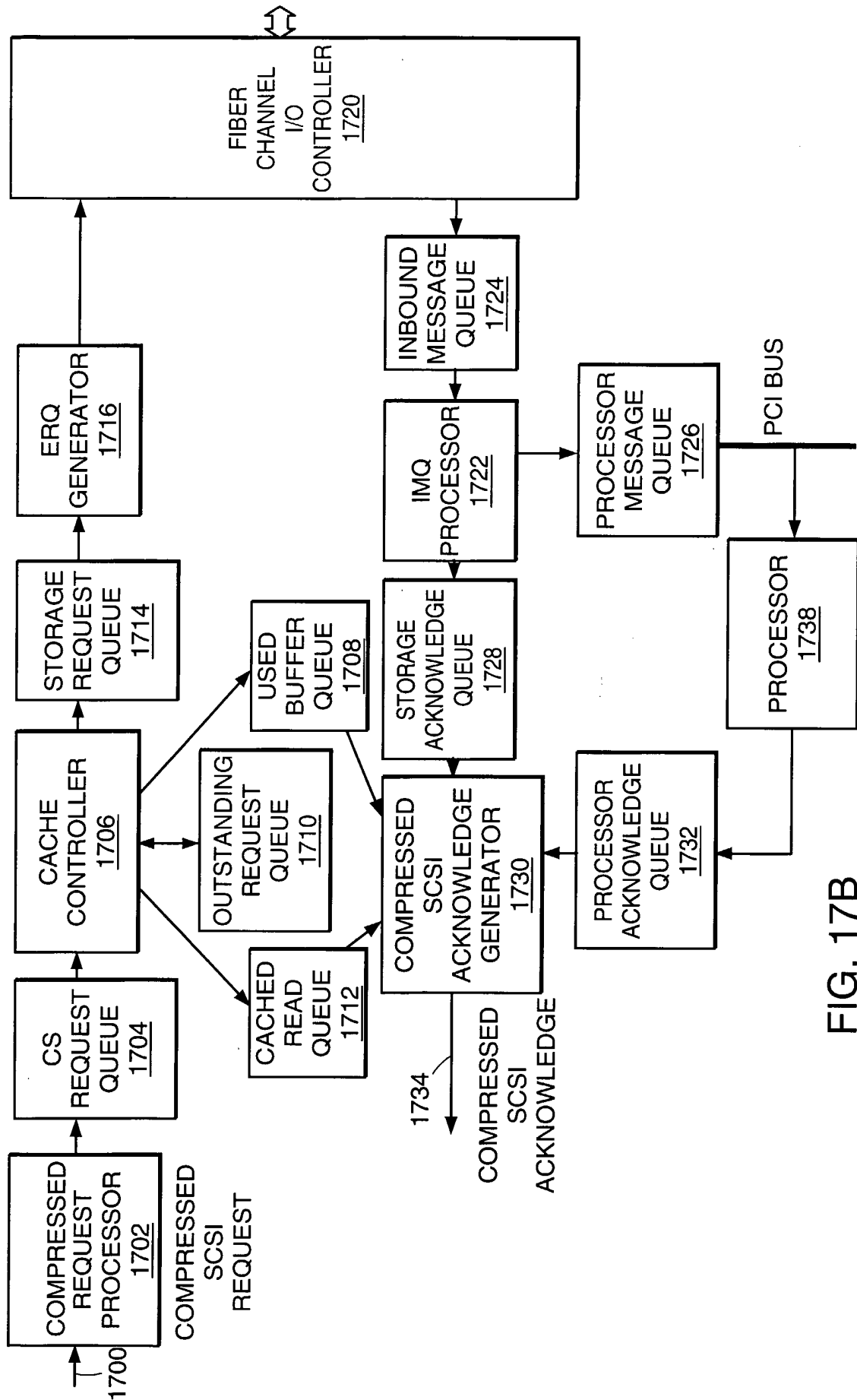


FIG. 17B

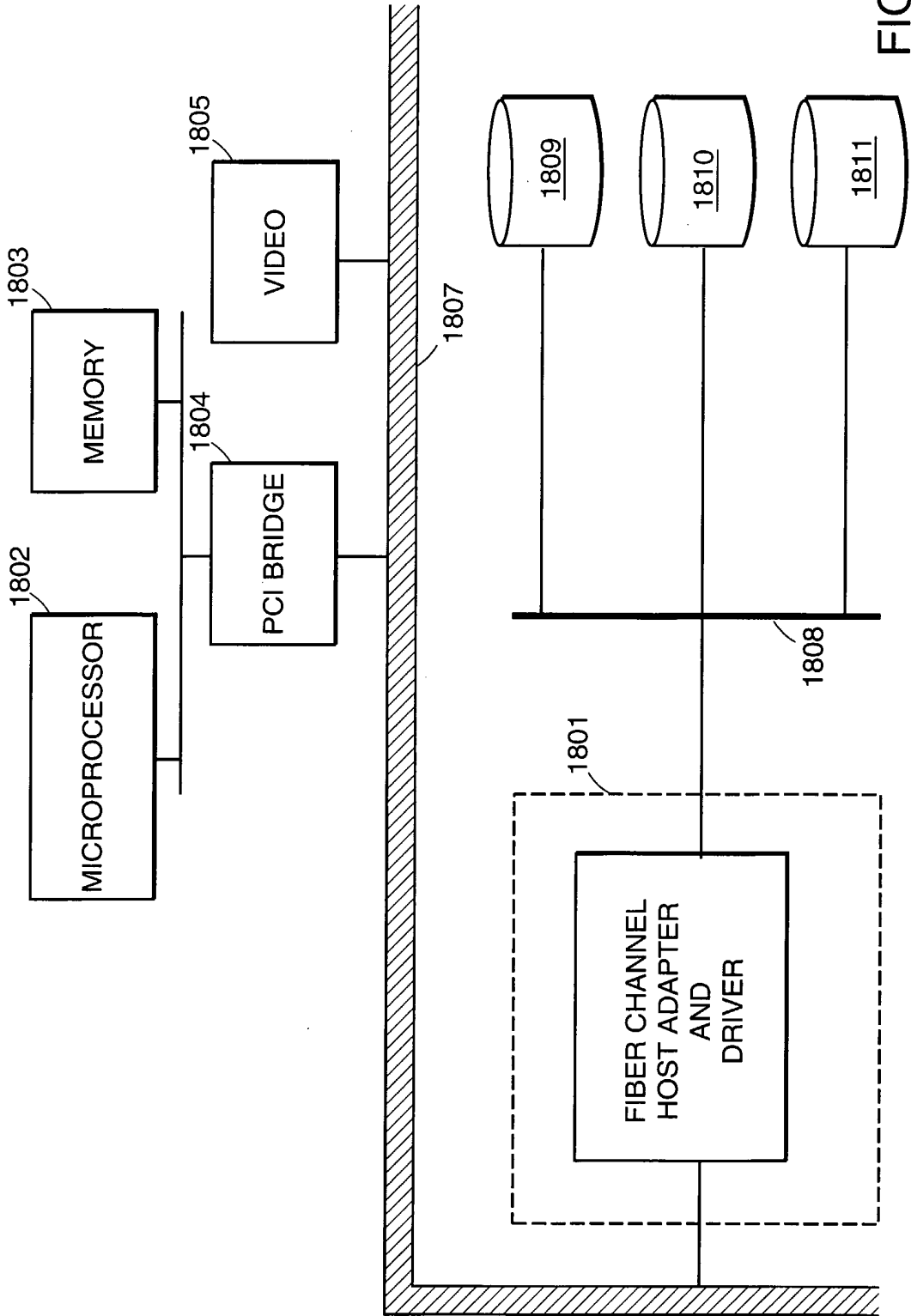


FIG. 18

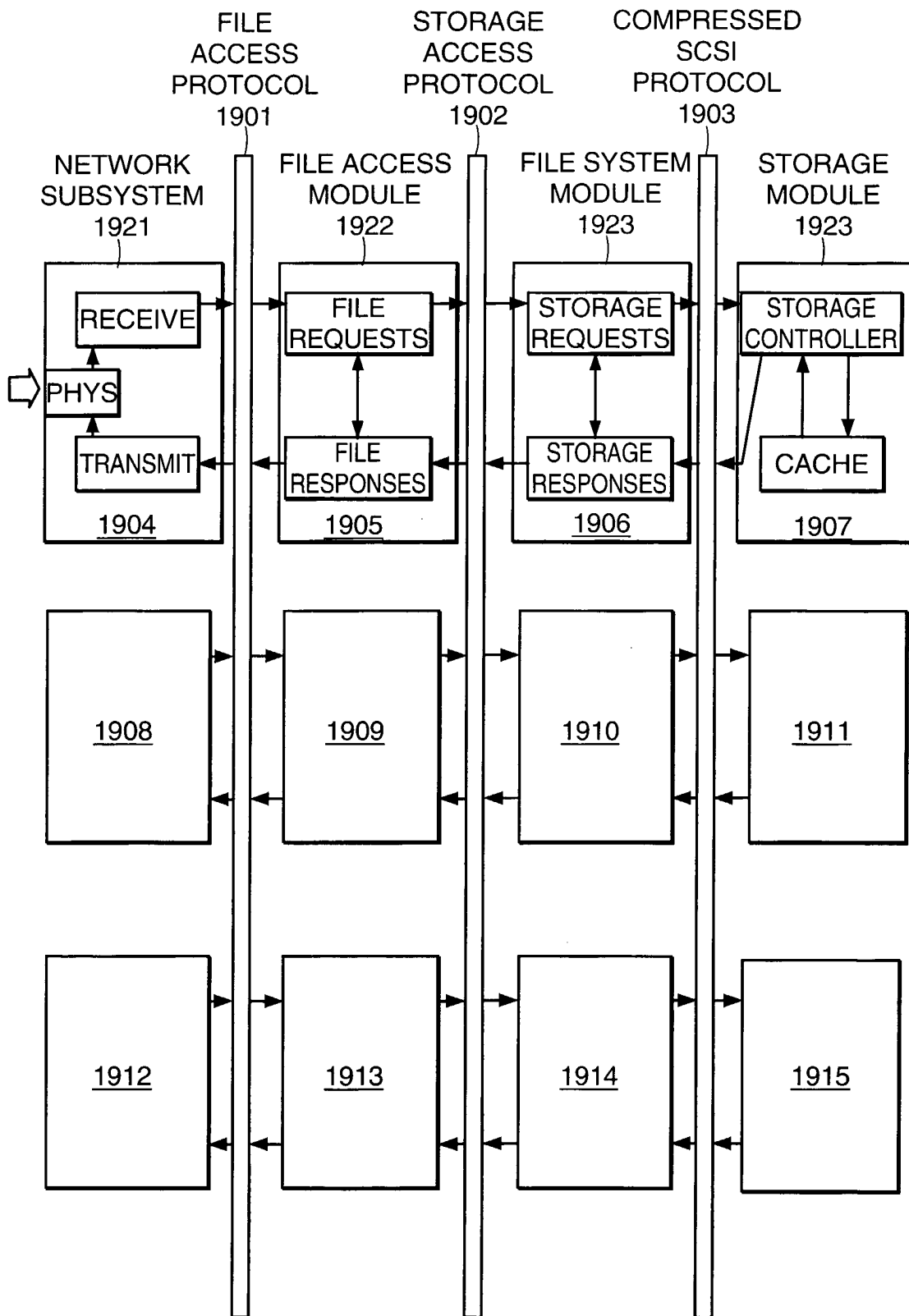


FIG. 19