

1. 一种用于耦合到主机设备的网络输入 / 输出 (I/O) 设备的装置, 包括:
电路;
接收模块, 由所述电路执行以用于接收数据帧, 所述数据帧包括访问在所述网络 I/O 设备中维护的存储子系统的请求, 所述存储子系统包括固态驱动器 (SSD);
缓冲器模块, 由所述电路执行以用于使用耦合到所述网络 I/O 设备的主机设备中的处理器插槽中所包括的处理器电路的缓存中维护的一个或多个缓冲器, 所述一个或多个缓冲器用于交换对于所述请求的控制信息, 所述控制信息与由所述处理器电路执行的协议堆栈交换; 以及
判断模块, 由所述电路执行以用于基于所述交换的控制信息来判断是否从所述存储子系统中读取或向其存储与所述请求相关联的数据。
2. 如权利要求 1 所述的装置, 其特征在于, 包括:
存储器控制器模块, 用于基于由所述判断模块作出的所述判断控制对所述存储子系统的访问。
3. 如权利要求 2 所述的装置, 其特征在于, 包括: 所述存储器控制器模块和至少所述 SSD 被配置为根据包括 PCIe 基础规范修订版 3.0 或 NVMe 规范修订版 1.1 的一个或多个行业标准操作。
4. 如权利要求 1 所述的装置, 其特征在于, 包括: 所述接收模块用于通过与所述主机设备的本地连接从所述主机设备、或通过与所述远程设备的网络连接从远程设备接收所述数据帧, 无论所述数据帧是从所述主机设备还是从所述远程设备接收到的, 所述数据帧都根据同一协议格式来配置。
5. 如权利要求 4 所述的装置, 其特征在于, 包括相同协议格式以包括以太网上的光纤通道 (FCoE)、因特网宽区域 RDMA 协议 (iWARP)、Infiniband 或融合以太网上的 RDMA (RoCE)。
6. 如权利要求 4 所述的装置, 其特征在于, 所述请求的所述控制信息包括与所述数据帧相关联的首部信息, 以便于通过由所述处理器电路执行的所述协议堆栈作出是否接受或拒绝所述请求的判断。
7. 如权利要求 6 所述的装置, 其特征在于, 包括: 所述接收模块用于从所述远程装置接收所述数据帧, 所述请求是请求所述远程装置将与所述数据帧相关联的数据存储到所述存储子系统, 所述数据至少临时被存储到在所述网络 I/O 装置中维护的存储器, 然后, 在所述请求被接受之后, 被导致存储到所述存储子系统。
8. 如权利要求 7 所述的装置, 其特征在于, 包括: 所述接收模块用于从所述远程装置接收所述数据帧, 所述请求是请求所述远程装置从所述存储子系统读取数据。
9. 如权利要求 11 所述的装置, 其特征在于, 包括:
发送模块, 由所述电路执行以用于在所述请求被接受之后通过所述网络连接将所述读取的数据发送到所述远程装置。
10. 如权利要求 8 所述的装置, 其特征在于, 与所述主机装置的所述本地连接被配置为根据包括 PCIe 基础规范修订版 3.0 的一个或多个行业标准操作。
11. 一种在网络输入 / 输出 (I/O) 设备上实现的方法, 包括:
接收数据帧, 所述数据帧包括用于访问在所述网络 I/O 设备中维护的存储子系统的请求, 所述存储子系统包括固态驱动器 (SSD);

使用耦合到所述网络 I/O 设备的主机设备中的处理器插槽中所包括的处理器电路的缓存中维护的一个或多个缓冲器,所述一个或多个缓冲器被配置为用于交换对于所述请求的控制信息,所述控制信息与由所述处理器电路执行的协议堆栈交换;以及

基于所述交换的控制信息,判断是否从所述存储子系统中读取或向其存储与所述请求相关联的数据。

12. 如权利要求 11 所述的方法,其特征在于,包括通过与所述主机设备的本地连接从所述主机设备、或通过与所述远程设备的网络连接从远程设备接收所述数据帧,无论所述数据帧是从所述主机设备还是从所述远程设备接收到的,所述数据帧都根据同一协议格式来配置。

13. 如权利要求 12 所述的方法,其特征在于,所述同一协议格式包括以太网上的光纤通道 (FCoE)、因特网宽区域 RDMA 协议 (iWARP)、Infiniband 或融合以太网上的 RDMA (RoCE)。

14. 如权利要求 12 所述的方法,其特征在于,包括所述从所述主机设备接收到的数据帧以及所述数据帧中包括的所述请求是用于由所述处理器电路执行以访问所述存储子系统的的目的。

15. 如权利要求 14 所述的方法,其特征在于,包括:来自所述应用的请求是请求将与所述数据帧相关联的数据存储到所述存储子系统,在满足所述请求以将所述数据存储到所述存储子系统之前,所述数据至少被临时存储到所述一个或多个缓冲器。

16. 如权利要求 14 所述的方法,其特征在于,包括:来自所述应用的请求是请求从所述存储子系统读取数据,在满足所述请求以从所述存储子系统读取所述数据之后,所述读取的数据至少被临时存储到所述一个或多个缓冲器。

17. 如权利要求 11 所述的方法,其特征在于,所述存储子系统中包括的所述 SSD 具有非易失性存储器,所述非易失性存储器包括 3 维交叉点存储器、闪存、铁电存储器、硅-氧化物-氮化物-氧化物-硅 (SONOS) 存储器、聚合物存储器、纳米线、铁电晶体管随机存取存储器 (FeTRAM 或 FeRAM)、纳米线或电可擦可编程只读存储器 (EEPROM) 中的至少一种。

18. 至少一种机器可读介质,所述至少一种机器可读介质包括多个指令,响应于在设备上被执行,所述指令导致所述设备实现如权利要求 11-17 中任一项所述的方法。

19. 一种设备,包括用于执行根据权利要求 11 到 17 中的任一项所述的方法的装置。

20. 一种在网络输入/输出 (I/O) 设备上实现的方法,包括:

标记主机设备上的处理器插槽中所包括的处理器电路的缓存中维护的一个或多个缓冲器,所述一个或多个缓冲器由耦合到所述主机设备的网络 I/O 设备来标记,以指出使用所述一个或多个缓冲器来集结通过耦合到所述网络 I/O 设备的一个或多个网络连接接收到的或传输的数据,所述数据与访问由耦合到所述主机设备的存储器控制器控制的存储设备的请求相关联;

接收数据帧,所述数据帧中包括远程设备访问所述存储设备的请求;以及

基于所述请求是要从所述存储设备读取所述数据还是将所述数据存储到所述存储设备,将与所述请求相关联的数据转发到一个或多个标记的缓冲器或从一个或多个标记的缓冲器接收与所述请求相关联的数据。

21. 如权利要求 20 所述的方法,其特征在于,标记所述一个或多个缓冲器包括:基于从所述远程设备接收到的一个或多个数据帧中所包括的一个或多个请求向所述一个或多个

缓冲器分配标识符,以指出使用所述一个或多个标记的缓冲器来访问所述存储设备。

22. 如权利要求 20 所述的方法,其特征在于,所述存储设备包括固态驱动器 (SSD),所述存储器控制器和所述 SSD 可以被配置为根据包括 PCIe 基础规范修订版 3.0 或 NVMe 规范修订版 1.1 的一个或多个行业标准来操作。

23. 如权利要求 22 所述的方法,其特征在于,所述 SSD 具有非易失性存储器,所述非易失性存储器包括 3 维交叉点存储器、闪存、铁电存储器、硅 - 氧化物 - 氮化物 - 氧化物 - 硅 (SONOS) 存储器、聚合物存储器、纳米线、铁电晶体管随机存取存储器 (FeTRAM 或 FeRAM)、纳米线或电可擦可编程只读存储器 (EEPROM) 中的至少一种。

24. 如权利要求 20 所述的方法,其特征在于,包括:所述数据帧根据协议格式来配置,所述协议格式包括以太网上的光纤通道 (FCoE)、因特网宽区域 RDMA 协议 (iWARP)、Infiniband 或融合以太网上的 RDMA (RoCE)。

用于在网络输入 / 输出设备和存储设备之间移动数据的技术

技术领域

[0001] 此处所描述的示例一般涉及基于接收到的数据帧中所包括的请求来向存储子系统存储数据或从其中读取数据。

[0002] 背景

[0003] 联网和存储越来越多地交织在一起,因为计算设备在高度分布式的操作模式下部署。例如,对于给定网络,存储的数据可以跨许多计算设备或网络节点分布。给定网络可被配置成一种基于云的服务,诸如作为服务的软件(SAAS)或作为服务的基础结构(IAAS)。通常,单个网络节点的存储和网络子系统在体系结构上被设计为单独的子系统。作为位于单独的子系统之中的结果,从网络到达给定网络节点上的存储器的数据或相反从给定网络节点上的存储器到达网络的数据,需要在各子系统之间被路由。常常,数据可以通过系统存储器路由,以在两个子系统之间移动数据。

[0004] 附图简述

[0005] 图 1 示出了第一示例系统。

[0006] 图 2 示出了第二示例系统。

[0007] 图 3 示出了示例过程。

[0008] 图 4 示出了设备的示例框图。

[0009] 图 5 示出了第一逻辑流程的示例。

[0010] 图 6 示出了第二逻辑流程的示例。

[0011] 图 7 示出了存储介质的示例。

[0012] 图 8 示出了网络输入 / 输出设备的示例。

具体实施方式

[0013] 如在本发明中设想的,在具有单独的网络和存储子系统的网络节点中接收到的或从网络节点传输的数据可以通过系统存储器来路由数据。这种类型的路由最初用于网络数据能力(例如,带宽)是这样的,以便需要相对较小量的系统存储器来在各子系统之间移动数据。还需要系统存储器,因为对于某些类型的存储器(包括硬盘驱动器)具有高访问延迟。然而,由于网络数据能力显著提高,因此,需要较大量的系统存储器来维护可接受的吞吐量性能。与使用系统存储器来在各子系统之间路由数据相关联的开销和延迟还可能进一步降低性能。另外,非易失性存储器的进步还允许在存储器子系统中使用。非易失性存储器可以被包括在这样的存储设备中,比如显著缩短了访问延迟的固态驱动器(SSD)。因此,较大的网络数据能力与对于较新类型的存储器的较低的访问延迟相结合可以导致使用系统存储器来在网络和存储子系统之间移动数据,作为用于移动数据的不太希望的并且效率较低的选项。正是针对这些和其它挑战才需要此处所描述的示例。

[0014] 在某些示例中,可以实现用于在网络(NW)输入 / 输出(I/O)设备和具有一个或多个存储设备的存储子系统之间移动数据的技术。对于这些示例,耦合到数据的NW I/O设备

的电路可以能够执行各种模块以促进数据的移动。各种模块可以包括接收包括访问在网络 I/O 设备中维护的存储子系统的请求的数据帧的接收模块。存储子系统可以包括诸如固态硬盘 (SSD) 之类的存储设备。各种模块还可以包括使用主机设备中的处理器插槽中所包括的处理器电路的缓存中维护的一个或多个缓冲器的缓冲器模块。一个或多个缓冲器可以被配置为交换对于请求的控制信息。控制信息可以与由处理器电路执行的协议堆栈交换。各种模块还可以包括基于交换的控制信息来判断是否从存储子系统中读取或向其存储与请求相关联的数据的判断模块。

[0015] 在某些其他示例中,技术还可以包括标记主机设备上的处理器插槽中所包括的处理器电路的缓存中维护的一个或多个缓冲器。对于这些其他示例,一个或多个缓冲器可以由耦合到主机设备的 NW I/O 设备来标记,以指出使用一个或多个缓冲器来集结 (staging) 通过耦合到 NW I/O 设备的一个或多个网络连接接收到的或传输的数据。数据可以与访问由耦合到主机设备的存储器控制器控制的存储设备的请求相关联。此外,还对于这些其他示例,可以接收数据帧,其中包括远程设备访问存储设备的请求。然后,可以基于请求是否将从存储设备读取数据或请求将数据存储到存储设备,将与请求相关联的数据转发到一个或多个标记的缓冲器,或从它们接收。

[0016] 图 1 示出了第一示例系统。如图 1 所示,第一示例系统包括系统 100,其中具有分别通过通信链路 135 和 145 耦合到 NW I/O 设备 130 和存储器控制器 140 的主机设备 105。也如图 1 所示,主机设备 105 能够通过通信信道 140,通过 NW I/O 设备 130 耦合到网络 150。在某些示例中,远程设备 170 可以通过耦合到网络 150 的通信信道 160 并通过耦合到 NW I/O 设备 130 的通信信道 140,建立一个或多个网络连接。对于这些示例,可以使用远程设备 170 和 NW I/O 设备 130 之间的一个或多个网络连接来接收数据帧,其中包括访问由耦合到主机设备 105 的存储器控制器 140 控制的存储设备 (例如,存储设备 142 之中的) 的请求。

[0017] 根据某些示例,术语“主机计算机”、“主机设备”、“主机”、“网络节点”、和“节点”可以可互换地使用,并可以无限制地表示,例如一个或多个终端站、移动因特网设备、智能电话、媒体设备、输入/输出 (I/O) 设备、平板电脑、电器、中间站、网络接口、客户端、和/或其一些部分。也在某些示例中,术语“远程设备”或“远程网络节点”可以可互换地使用,并可以无限制地表示,例如,主机设备远程可访问的 (例如,通过网络连接) 计算设备。

[0018] 根据某些示例,诸如网络 150 之类的“网络”可以包括至少部分地允许、促进和/或使两个或更多实体可通信地耦合在一起的任何机制、手段、特征和/或其一部分。也在某些示例中,如果第一实体能够向第二实体传输和/或从第二实体接收一个或多个命令和/或数据 (例如,数据帧中所包括的),则第一实体可以“可通信地耦合”到第二实体。

[0019] 在某些示例中,如图 1 所示,主机设备 105 包括处理器插槽 110。也如图 1 所示,处理器插槽 110 包括处理器电路 112 和末级缓存 114。处理器电路 112 可以包括一个或多个处理元件 112-1 到 112-n (其中,“n”表示 >1 的任何整数) 和缓冲器管理器 116。处理元件 112-1 到 112-n 可以表示处理电路 112 中所包括的多个处理核或引擎,缓冲器管理器 116 可以是主机设备 105 的操作系统的处理元件 112-1 到 112-n 执行的实例化中的设备驱动程序的一部分。处理元件 112-1 到 112-n 能够使用包括缓冲器 115-1 到 115-m 的末级缓存 114 (其中,“m”表示 >2 的任何整数) 以至少临时集结或存储数据。根据某些示例,末级缓存 114 的缓冲器 115-1 到 115-m 可以被分别地指定给处理器电路 112 的给定处理元件,

或可以在各处理器元件之间共享,并可以包括各种类型的存储器,诸如具有低访问延迟和相对较小的数据能力的易失性存储器(例如,静态随机存取存储器(同步RAM))。在任一种情况下,缓冲器管理器 116 都可以表示由处理元件 112-1 到 112-n 执行的逻辑和 / 或特征,处理元件 112-1 到 112-n 能够管理对缓冲器 115-1 到 115-m 的访问并促进 NW I/O 设备 130 和由存储器控制器 140 控制的存储设备 142 之间的数据的可能的移动。NW I/O 设备 130 和存储设备 142 之间的数据的移动的促进可以包括缓冲器管理器 116 维护至少一些映射信息以将与移动相关联的命令映射到存储设备 142 中所包括的一个或多个存储设备。

[0020] 根据某些示例, NW I/O 设备 130, 如图 1 所示, 包括电路 132 和存储器 134。对于这些示例, 各种模块中包含的逻辑和 / 或特征可以由电路 132 执行, 以使用末级缓存 114 的缓冲器 115-1 到 115-m 来促进从存储设备 142 中读取的数据的移动或写入 / 存储到存储设备 142。数据移动可以响应于从远程设备 170 接收到的数据帧, 数据帧可以包括通过读取请求或写入 / 存储请求, 访问存储设备 142 的请求。NW I/O 设备 130 中的存储器 134 还可以被配置为存储由各种模块实现的固件或软件或至少临时维护与从远程设备 170 接收到的数据帧相关联的数据。

[0021] 在某些示例中, NW I/O 设备 130 中的逻辑和 / 或特征可以能够标记缓冲器 115-1 到 115-m 中的一个或多个, 以指出使用一个或多个缓冲器, 以便集结或至少临时存储通过 NW I/O 设备 130 和远程设备 170 路由的一个或多个网络连接, 通过网络 150, 接收或传输的数据。对于这些示例, 并非使用主机设备 105 中的系统存储器(未示出), 可以用于末级缓存 114 中的缓冲器 115-1 到 115-m 来在缓冲器管理器 116 的协助下快速地在存储设备 142 和 NW I/O 设备 130 之间移动数据。对缓冲器 115-1 到 115-m 的标记可以包括 NW I/O 设备 130 的逻辑和 / 或特征向缓冲器中的一个或多个指定标识符以指出使用一个或多个标记的缓冲器来访问存储设备 142。例如, 可以标记缓冲器 115-1, 并可以将此标记的缓冲器的给定标识符传递到缓冲器管理器 116。随后, 当在由远程设备 170 所发送的并由 NW I/O 设备 130 接收到的数据帧中接收到访问存储设备 142 的请求时, 给定标识符可以使缓冲器管理器 116 能快速地标识用来基于请求来集结可能移动的数据的特定缓冲器。缓冲器管理器 116 还能够审查请求以判断该请求是否是读取或写访问, 并相应地集结数据。

[0022] 根据某些示例, 可以在 NW I/O 设备 130 中从远程设备 170 之中的远程设备接收数据帧。对于这些示例, 数据帧可以包括远程设备访问存储设备 142 的请求。然后, NW I/O 设备 130 中的逻辑和 / 或特征可以基于请求是否是从存储设备 142 读取数据或将数据存储到存储设备 142 中, 将与请求相关联的数据转发到一个或多个标记的缓冲器, 或从它们接收。例如, 如果请求是从存储设备 142 读取数据, 那么, 可以将数据集结在标记的缓冲器中(例如, 缓冲器 115-1), 然后, 由 NW I/O 设备 130 的逻辑和 / 或特征从此标记的缓冲器接收, 然后, 传输到发出请求的远程设备。可另选地, 如果请求是将数据存储到存储设备 142, 那么, 数据可以被包括在从远程设备接收到的数据帧中。然后, 可以由 NW I/O 设备 130 的逻辑和 / 或特征将此数据转发到标记的缓冲器, 以便集结数据, 最终存储在存储设备 142 中。在某些示例中, 缓冲器管理器 116, 基于该请求的控制信息, 可以重新格式化请求, 以便存储器控制器 140 可以响应并相应地满足该请求。

[0023] 在某些示例中, 将 NW I/O 设备 130 和存储器控制器 140 耦合到主机设备 105 的通信链路 135 和 145 能够根据与 I/O 互连相关联的一个或多个行业标准操作, 包括, 但不仅

限于,201011月发布的外围组件互连(PCI)高速基础规范,修订版3.0(“PCI Express”或“PCIe”)。

[0024] 根据某些示例,存储设备142可以包括一个或多个设备,可以分别向其中存储数据和/或从其中取回数据。对于这些示例,存储设备142还可以包括用于数据的存储的非易失性存储器。例如,存储设备142可包括但不限于一个或多个非易失性电机械、磁、光和/或半导体存储设备。这些设备可以包括硬盘驱动器(HDD)或固态驱动器(SSD)。SSD可以具有非易失性存储器,诸如3维交叉点存储器、闪存、铁电存储器、硅氧化物氮化物氧化物硅(SONOS)存储器、聚合物存储器、纳米线、铁电晶体管随机存取存储器(FeTRAM或FeRAM)、纳米线或电可擦可编程只读存储器(EEPROM)。

[0025] 根据某些示例,存储器控制器140、存储设备142和通信链路145能够根据PCIe规范操作,还根据与使用符合PCIe的通信链路和/或协议来访问非易失性存储器类型的存储器(诸如SSD)相关联的其他行业标准操作。这些其他行业标准可以包括,但不仅限于,2012年10月发布的非易失性存储器Express(NVMe)规范,修订版1.1。存储器控制器140、存储设备142和通信链路145还能够根据其他行业标准操作,它们可以包括,但不仅限于,2001年7月发布的串行ATA(SATA)规范,修订版3.1、2010年12月发布的串行外接SCSI(SAS)规范,修订版2.1,和/或2004年4月发布的因特网SCSI(iSCSI),征求意见资料3720。

[0026] 在某些示例中,通信信道140可以包括一个或多个通信链路,通过它们,NW I/O设备130可以耦合到网络150,并与通过通信信道160耦合到网络150的远程设备建立一个或多个网络连接,通信信道160还可以具有一个或多个通信链路。通信信道140或160中所包括的这些通信链路可以包括各种类型的有线、无线或光通信介质。对于这些示例,通信链路可以根据任何版本的一个或多个适用的通信或网络标准操作。例如,通信链路可以根据由电气电子工程师学会(IEEE)发布的对于有线或无线网络的一个或多个标准或规范操作。这些标准和规范可以包括,但不仅限于,2012年3月发布的用于信息技术的IEEE802.11-2012标准——系统之间的电信和信息交换——局域网和城域网——特定要求部分11:WLAN介质访问控制器(MAC)和物理层(PHY)规范(IEEE802.11-2012Standard for Information technology-Telecommunications and information exchange between systems-Local and metropolitan area networks-Specific requirements Part11:WLAN Media Access Controller(MAC)and Physical Layer(PHY)Specifications),和/或此标准对于无线介质的更高版本(“IEEE802.11”)或IEEE802.3-2008,200812月发布的对于有线介质的带有冲突检测的载波侦听多址访问(CSMA/CD)接入方法和物理层规范(Carrier sense Multiple access with Collision Detection(CSMA/CD)Access Method and Physical Layer Specifications)(下文简称为“IEEE802.3”)。

[0027] 在某些示例中,NW I/O设备130和远程设备170可以根据一个或多个协议,通过网络150来交换数据帧,协议可以符合和/或兼容于各种类型的远程直接存储器访问(RDMA)协议,诸如因特网宽区域RDMA协议(iWARP)、Infiniband(IB)(无限宽带)协议、以太网协议、传输控制协议/网际协议(TCP/IP)协议和/或融合以太网上的RDMA(RDMA over converged Ethernet)(RoCE)。例如,iWARP协议可以符合和/或兼容于Recio等人所著的“RDMA协议规范(An RDMA Protocol Specification)”(因特网规范草案(Internet Draft Specification),因特网工程任务组(Internet Engineering Task Force(IETF)),

2002年10月21日)。另外,例如,TCP/IP协议可以符合和/或兼容于在1981年9月发布的因特网工程任务组(IETF)征求意见资料(RFC)791和793中所描述的协议。此外,IB协议还可以符合和/或兼容于2012年11月发布的Infiniband™体系结构规范,2卷版本1.3。另外,例如,RoCE协议可以符合和/或兼容于2010年4月发布的无线宽带架构规范补充Vol. 1, Rel. 1. 2. 1, 附录A16:“融合以太网上的RDMA(RoCE)”(Supplement to Infiniband Architecture Specification, Vol. 1, Rel. 1. 2. 1, Annex A16: “RDMA over Converged Ethernet(RoCE)”)。NW I/O设备130和远程设备170还可以通过网络150,根据一个或多个协议来交换数据帧,协议可以封装被称为以太网上的光纤通道(FCoE)的以太网上的光纤通道帧。FCoE可以兼容于在2009年6月发布的认可标准委员会(Accredited Standards Committee) INCITS T11技术委员会的美国国家标准起草的光纤信道中枢-5(FC-BB-5)(Fibre Channel Backbone-5(FC-BB-5))标准,修订版2.0中所描述的协议。许多不同的,额外的,和/或其他协议可以用于这样的数据帧的交换,而不会偏离这些示例(例如,如上所述的,相关的,和/或其他协议的以前的和/或以后开发的版本)。

[0028] 图2示出了第二示例系统。如图2所示,第二示例包括系统200。根据某些示例,系统200包括通过通信链路235耦合到NW I/O设备230的主机设备205。在某些示例中,远程设备270可以通过耦合到网络250的通信信道260并通过耦合到NW I/O设备230的通信信道240,建立一个或多个网络连接。对于这些示例,可以使用远程设备270和NW I/O设备230之间的一个或多个网络连接来接收数据帧,其中包括访问NW I/O设备230中维护的和/或与其联合的存储子系统(例如,存储子系统236)的请求。

[0029] 根据某些示例,NW I/O设备230可以集成网络和存储子系统。作为集成网络和存储子系统的结果,访问存储子系统236的请求可以通过通信链路235来自主机设备205,或通过经由通信信道240路由的一个或多个网络连接,来自远程设备270。对于这些示例,并非在末级缓存的一个或多个缓冲器(例如,缓冲器215-1到215-m)上集结对于远程设备270作出的请求的数据,可以使用一个或多个缓冲器来交换对于请求的控制信息(例如,数据帧首部信息)。对于发自主机设备205(例如,发自应用224)的请求,数据仍可以在一个或多个缓冲器上集结。不管请求的源是什么,数据的移动不包括使用系统存储器(例如,存储器220中所包括的)来在请求者和请求者正在寻求访问的存储子系统之间移动数据。

[0030] 在某些示例中,如图2所示,主机设备205包括处理器插槽210,其中包括如对于图1的主机设备105所提及的类似的组件或特征。例如,处理器插槽210包括具有处理元件212-1到212-n的处理器电路212和缓冲器管理器216。处理器插槽210还可以包括具有缓冲器215-1到215-n的末级缓存214。

[0031] 根据某些示例,主机设备205可以包括通过存储器通道215耦合到处理器插槽210的存储器220。对于这些示例,协议堆栈222和应用224可以维护在存储器220中,并能够由处理器电路212执行。协议堆栈222可以包括被配置为处理与数据帧相关联的控制信息的协议处理软件,数据帧包括访问诸如存储子系统236之类的存储子系统的请求,存储子系统236可以包括位于NW I/O设备230中的SSD237。在某些示例中,数据帧可以发自主机设备205的逻辑和/或特征,诸如来自应用224,或可以发自远程设备270。数据帧可以包括控制信息,通过控制信息,协议堆栈222可以允许主机设备205接受或拒绝访问存储子系统236的请求。

[0032] 根据某些示例,各种模块中包含的逻辑和/或特征可以由NW I/O设备230中的电路232执行。这些模块中的一个或多个可以使用一个或多个缓冲器215-1到215-m来与协议堆栈222交换控制信息。例如,可以在I/O设备230中从远程设备270接收数据帧,数据帧具有数据帧的首部形式的控制信息,并还可以数据有效负载中的数据。对于此示例,可以将首部中所包括的控制信息发送到一个或多个缓冲器215-1到215-m。然后,缓冲器管理器216可以促进该控制信息在协议堆栈222和一个或多个模块之间的交换。然后,可以基于交换的信息,作出关于数据帧中所包括的请求是从存储子系统236读取数据还是将数据存储在存储子系统236的判断。

[0033] 在某些示例中,共同的或同一协议格式可以用于通过通信链路235从主机设备205的逻辑和/或特征或通过经由通信信道240路由的网络连接从远程设备270接收到的数据帧。同一个协议格式可以包括诸如FCoE、iWARP、Infiniband或RoCE之类的协议。对于这些示例,由NW I/O设备230中的电路232执行的各种模块可以接收访问存储子系统236的请求,请求使用同一协议格式被封装,无论来自诸如通信链路235之类的本地连接,还是通过经由通信信道240路由的到网络250的网络连接。

[0034] 在某些示例中,类似于系统100的通信链路135和145,系统200的通信链路235能够根据与I/O互连相关联的包括PCIe的一个或多个行业标准或规范操作。此外,要由NW I/O控制器230中的232执行的模块中的一个或多个和存储子系统236的至少SSD237能够根据NVMe规范操作。此外,通信信道240或260还可以包括各种类型的有线,无线或光通信介质。对于这些示例,通信介质以及网络250可以根据任何版本的一个或多个适用的通信或网络标准操作,包括,但不限于,IEEE802.11或IEEE802.3,仅举几个例子。

[0035] 图3示出了示例过程300。在某些示例中,如1或2所示的系统100或200的元件可以被用来示出涉及图3中所描绘的过程300的流程图的示例操作。所描述的示例操作不仅限于图1或2所描述的系统100或200上的实现。

[0036] 从开始移动到判断框310(数据帧来自主机或远程设备吗?),包括访问存储子系统236的请求的数据帧可能是从主机设备205或从远程设备270接收到的。根据某些示例,访问存储子系统236的请求可以发自应用224,如此,数据帧被认为是来自主机设备205的。对于这些示例,过程移到判断框320。否则,如果数据帧是从远程设备270接收到的,过程移到框350。

[0037] 从判断框310移动到框320(主机创建数据帧的控制信息),协议堆栈222可以创建包括应用224访问存储子系统236的请求的数据帧的控制信息。在某些示例中,主机设备205的元件可以首先验证/认证该请求,然后,一旦请求被验证,例如,应用224具有访问存储子系统236的适当的访问权限,创建控制信息。对于这些示例,控制信息可以与通常用于包括访问存储子系统236的请求的数据帧的协议格式相关联。例如,共同协议格式可以与FCoE、iWARP、Infiniband或RoCE相关联。控制信息,例如,可以是要被路由到NW I/O设备230的数据帧的首部的一部分。

[0038] 从框320前进到框330(交换控制信息),控制信息可以是用于来自应用224的数据帧的,该数据帧可以在协议堆栈222以及NW I/O设备230的逻辑和/或特征之间交换。根据某些示例,一个或多个缓冲器215-1到215-m可以用于交换控制信息。例如,控制信息可以包括应用224是否请求对存储子系统236的读取访问或存储器访问的指示,还可以包

括关于什么数据要被读取或要存储多少数据的进一步的细节。NW I/O 设备 230 还可以提供指出存储子系统 236 是否具有满足访问请求的能力的信息。

[0039] 从框 330 前进到框 340 (主机根据需要附加数据), 协议堆栈 222 可以将与请求相关联的数据附加或添加到数据帧。在某些示例中, 如果请求是对存储子系统 236 的读取访问, 则可不附加数据。可另选地, 如果请求是存储器访问, 则可以将要被存储到存储子系统 236 的数据附加到数据帧。

[0040] 从判断框 310 移动到框 350 (NW I/O 设备分割控制信息和数据), NW I/O 设备 230 中的逻辑和 / 或特征可以分割从远程设备 270 接收到的控制信息与数据帧。根据某些示例, 控制信息可以被包括在数据帧的首部部分中。对于这些示例, 控制信息可以与用于访问存储子系统 236 采用共同协议格式。例如, 如上文所提及的, 共同协议格式可以与 FCoE、iWARP、Infiniband 或 RoCE 相关联。

[0041] 从框 350 前进到框 360 (交换控制信息), NW I/O 设备 230 中的逻辑和 / 或特征可以在主机设备 205 中与协议堆栈 222 交换分裂出的控制信息。在某些示例中, 如上文所提及的, 一个或多个缓冲器 215-1 到 215-m 可以用于交换控制信息。例如, 控制信息可以包括远程设备 270 是否请求对存储子系统 236 的读取访问或存储器访问的指示, 还可以包括关于什么数据要被读取或要存储多少数据的进一步的细节。NW I/O 设备 230 还可以提供指出存储子系统 236 是否具有满足访问请求的能力的信息。

[0042] 从框 360 前进到框 370 (临时存储与该请求相关联的数据), NW I/O 设备 230 中的逻辑和 / 或特征可以存储与请求相关联的数据。在某些示例中, 如果从远程设备 270 接收到的数据帧包括要被存储到存储子系统 236 中的数据, 则可将数据至少临时存储在存储器 236 中。对于这些示例, 基于在框 360 交换的信息, 可以只是临时存储数据, 以等待对请求的响应。

[0043] 从框 340 或框 370 移动 (对于请求的主机响应到达), 协议堆栈 222 可以发送对在 NW I/O 设备 230 中接收到的数据帧中所包括的访问存储子系统 236 的请求的响应。根据某些示例, 响应可以基于在框 330 或框 360 中交换的信息。

[0044] 从框 375 前进到判断框 385 (接受请求吗?) NW I/O 设备 230 中的逻辑和 / 或特征可以判断是否接受访问存储子系统 236 的请求。在某些示例中, 可以基于指出存储子系统 236 具有服务于该请求的能力的交换的控制信息 (例如, 具有可用空间, 具有请求的数据或请求者具有适当的访问权限)。如果请求被接受, 则过程移到框 390。否则, 过程移到判断框 395。

[0045] 从判断框 385 移动到框 390 (导致请求被满足), 导致接收到的数据帧中所包括的请求被满足, 并且过程结束。根据某些示例, NW I/O 设备 230 中的逻辑和 / 或特征可以导致与存储子系统 236 内的存储设备 (例如, 与 SSD 237 一起) 一起的控制器 (未示出) 满足访问请求。例如, 在请求来自于应用 224 存储数据的情况下, NW I/O 设备中的逻辑和 / 或特征可以使用一个或多个缓冲器 215-1 到 215-m 接收或拉取由应用 224 集结的或临时存储到这些缓冲器中的数据。例如, 在请求是来自远程设备 270 的存储请求的情况下, 逻辑和 / 或特征可以从用于临时存储从远程设备接收到的数据帧中所包括的数据的存储器 237 拉取数据。此外, 例如, 在请求是来自应用 224 从存储子系统 236 读取数据的读取请求的情况下, 可以将读取的数据推或集结到一个或多个缓冲器 215-1 到 215-m。还例如, 在请求是来

自远程设备 270 的读取请求的情况下,读取的数据可以被包括在数据帧中,并通过在通信信道 240 和 260 上路由的网络连接并通过网络 250,发送到远程设备 270。

[0046] 从判断框 385 移动到框 395(丢弃数据 / 数据帧),如果访问存储子系统 236 的请求被拒绝,则可以丢弃数据和 / 或数据帧,且过程结束。在某些示例中,可以丢弃从远程设备 270 接收到的数据帧中的临时存储到存储器 236 中的任何数据。在其他示例中,可以丢弃可能从应用 224 集结到缓冲器 215-1 到 215-m 中的任何一个的任何数据。

[0047] 图 4 示出了设备 400 的示例框图。尽管图 4 中示出的设备 400 具有处于特定拓扑中的有限数量的元素,但可以明白,设备 400 可包括给定实现所需的处于替换拓扑中的更多或更少元素。

[0048] 设备 400 可以由耦合到主机设备(例如,NW I/O 设备 130/230 的电路 132/232)的 NW I/O 设备中维护的电路 420 支持。电路 420 可以被配置为执行一个或多个软件或固件实现的组件或模块 422-a。值得注意的是,如此处所使用的“a”和“b”和“c”以及类似的符号旨在作为表示任何正整数的变量。如此,例如,如果实现设置 $a = 6$ 的值,那么,模块 422-a 的完整的一组软件或固件可以包括模块 422-1、422-2、422-3、422-4、422-5 或 422-6。呈现的示例在此上下文不受限制,所使用的不同的变量可以表示相同或不同的整数值。

[0049] 根据某些示例,电路 420 可以包括处理器或处理器电路。处理器或处理器电路可以是各种市场上可买到的处理器中的任一种,包括但不限于,AMD® Athlon®、Duron® 和 Opteron® 处理器;ARM® 应用程序、嵌入式和安全处理器;IBM® 和 Motorola® DragonBall® 以及 PowerPC® 处理器;IBM 和 Sony® Cell 处理器;Intel® Atom®、Celeron®、Core(2) Duo®、Core i3、Core i5、Core i7、Itanium®、Pentium®、Xeon®、Xeon Phi® 和 XScale® 处理器;以及类似的处理器。根据某些示例,电路 420 也可以是专用集成电路(ASIC),至少某些模块 422-a 可以实现为 ASIC 的硬件元件。

[0050] 根据某些示例,设备 400 可以包括供电路 420 执行的标记模块 422-1。标记模块 422-1 能够标记耦合到包括设备 400 的 NW I/O 设备的主机设备上的处理器插槽中所包括的处理器电路的缓存中维护的一个或多个缓冲器。对一个或多个缓冲器的标记可以包括发送标识信息 405,标识信息 405 可以包括由标记模块 422-1 指定给一个或多个缓冲器的一个或多个标识符。标识符可以将某些缓冲器标记为用于集结通过耦合到 NW I/O 设备的一个或多个网络连接接收到的或传输的数据。数据可以与访问由耦合到主机设备的存储器控制器控制的存储设备的请求相关联。例如,存储设备 142 由存储器控制器 140 控制,如上文对于图 1 所示出的系统 100 所提及的。标记模块 422-1 可以能够至少临时存储缓冲器信息 424-a(例如,在诸如查找表(LUT)之类的数据结构中)。缓冲器信息 424-a 可以包括指定给标记的缓冲器的标识符。

[0051] 在某些示例中,设备 400 也可以包括供电路 420 执行的接收模块 422-2。接收模块 422-2 能够接收远程设备的数据帧 410,如对于图 1 所示出的系统 100 所描述的,或从远程设备或由主机设备处的处理器电路执行的应用,如对于如图 2 所示的系统 200 所描述的。接收模块 422-2 能够至少临时存储协议信息 425-b(例如,在 LUT 中)以促进接收数据帧 410。协议信息 425-b 可以允许接收模块 422-2 至少能够解码数据帧 410 的至少首部部分,首部

部分可以是与通过 NW I/O 设备访问存储设备或子系统相关联的各种协议格式。这些各种协议格式可以包括,但不仅限于,FCoE、iWARP、Infiniband 或 RoCE 协议格式。

[0052] 在某些示例中,设备 400 也可以包括供电路 420 执行的缓冲器模块 422-3。缓冲器模块 422-3 能够使用主机设备的处理器插槽中所包括的处理器电路的缓存中维护的一个或多个缓冲器。可以使用缓冲器模块 422-3 来与由主机设备处的处理器电路执行的协议堆栈交换控制信息(信息)415 中所包括的控制信息。交换的控制信息可以是针对控制信息 415,例如,可以包括从数据帧 410 分开的首部信息,如果远程设备发送了数据帧 410,并还可以包括与通过其作出了访问请求的存储子系统和 / 或存储设备相关联的访问信息。在其他示例中,如果数据帧 410 来自于主机设备(例如,来自应用),则控制信息 415 还可以包括与通过其作出了访问请求的存储子系统和 / 或存储设备相关联的访问信息。在其他示例中,控制信息 415 还可以包括请求是读取还是存储请求的指示。对于所有上述示例,控制信息 415 中所包括的控制信息可以使用一个或多个缓冲器(例如,基于标记的标识符)在缓冲器模块 422-3 和协议堆栈之间传递。最终,至少部分地基于控制信息 415 中所包括的交换的控制信息,可以接收请求状态 430 以指出协议堆栈和 / 或主机设备的其他元件是否接受或拒绝访问存储子系统和 / 或存储设备的请求。

[0053] 根据某些示例,设备 400 也可以包括供电路 420 执行的判断模块 422-4。判断模块 422-4 能够判断是否从存储子系统和 / 或存储设备中读取或向其存储与请求相关联的数据。对于这些示例,判断模块 422-4 可以至少临时存储控制信息 426-c,该控制信息 426-c 包括在缓冲器模块 422-3 和协议堆栈之间交换的控制信息 415 中的控制信息。

[0054] 在某些示例中,设备 400 也可以包括供电路 420 执行的存储器控制器模块 422-5。存储器控制器模块 422-5 能够根据判断模块 422-4 作出的判断,和 / 或还基于请求状态 430 是否指出请求的被接受,来控制对存储子系统和 / 或存储设备的访问。

[0055] 根据某些示例,控制信息 415 中所包括的交换的控制信息可以指出读取,请求状态 430 指出请求的被接受。对于这些示例,存储器控制器模块 422-5 可以导致数据被从存储子系统和 / 或存储设备中读取,作为数据读取 435。在某些其他示例中,交换的控制信息可以指出存储,请求状态 430 指出请求的被接受。对于这些其他示例,存储器控制器模块 422-5 可以导致要存储的数据 440 中所包括的数据存储到存储子系统和 / 或存储设备。可以从在 NW I/O 设备中维护的存储器中取回(临时从数据帧 410 中所包括的数据存储的)或可以从一个或多个缓冲器(当从主机设备中的请求者接收到时,通过其集结或临时存储数据(例如,应用——还从数据帧 410 中所包括的数据存储))拉要存储的数据 440 中所包括的此数据。

[0056] 根据某些示例,设备 600 也可以包括供电路 420 执行的发送模块 422-6。发送模块 422-6 能够将数据读取 435 中所包括的数据发送到发送了数据帧 410 的请求者。根据某些示例,发送模块 422-6 能够使用协议信息 425-b 来发送数据帧中的数据读取 435 中所包括的数据,数据帧可以是这样的协议格式,可以包括,但不仅限于,FCoE、iWARP、Infiniband 或 RoCE 协议格式。数据帧可以通过网络连接被发送到远程设备请求者,或通过 NW I/O 设备耦合到主机设备的本地连接被发送到主机设备中的请求者。

[0057] 此处所包括的是一组代表用于执行所公开的体系结构的新颖方面的示例方法的逻辑流程图。尽管为了简洁起见,作为一系列动作示出和描述了一个或多个方法,但是,所

属领域的技术人员将了解和明白,方法不受动作的顺序的限制。某些动作可以按与此处所示出和描述的不同的顺序进行和 / 或与其他动作并行地进行。例如,本领域的技术人员将明白并理解,方法可被替换地表示为一系列相互相关联的状态或事件,诸如以状态图的形式。此外,并非方法中所示出的所有动作都是新颖实现所必需的。

[0058] 逻辑流程可以以软件、固件,和 / 或硬件来实现。在软件和固件实施例中,逻辑流程可通过计算机执行的指令来实现,该计算机执行的指令被存储在例如光学存储、磁存储或半导体存储之类的至少一种非瞬态计算机可读介质或机器可读介质上。在此上下文中,实施例不受限制。

[0059] 图 5 示出了第一逻辑流程的示例。如图 5 所示,第一逻辑流程包括逻辑流程 500。逻辑流程 500 可代表诸如设备 400 之类的本文所述的一个或多个逻辑、特征或设备所执行的操作中的一些或全部。更具体而言,逻辑流程 500 可以通过标记模块 422-1、接收模块 422-2、缓冲器模块 422-3、判断模块 422-4、存储器控制器模块 422-5 或发送模块 422-6 中的一个或多个来实现。

[0060] 根据某些示例,逻辑流程 500 在框 502 可以在耦合到主机设备的 NW I/O 设备上接收数据帧,其中包括访问在 NW I/O 设备中维护的存储子系统的请求,存储子系统包括 SSD。例如,数据帧 410 可以由接收模块 422-2 接收,并可以包括访问可以包括设备 400 的 NW I/O 设备中维护的存储子系统的请求。

[0061] 在某些示例中,逻辑流程 500 在框 504 可以使用主机设备中的处理器插槽中所包括的处理器电路的缓存中维护的一个或多个缓冲器,一个或多个缓冲器被配置为交换对于请求的控制信息,该控制信息与由处理器电路执行的协议堆栈交换。例如,缓冲器模块 422-3 可以使用一个或多个缓冲器来与协议堆栈交换控制信息 415 中所包括的控制信息。

[0062] 根据某些示例,逻辑流程 500 在框 506 可以基于交换的控制信息,判断是否从存储子系统中读取或向其存储与请求相关联的数据。例如,判断模块 422-4 可以使用控制信息 415 中所包括的控制信息来作出关于访问类型的判断。如上文所提及的,设备 400 的附加的模块可以促进对存储子系统的访问,并导致数据被存储到存储子系统或从中被读取。

[0063] 图 6 示出了第二逻辑流程的示例。如图 6 所示,第二逻辑流程包括逻辑流程 600。逻辑流程 600 可代表诸如设备 400 之类的本文所述的一个或多个逻辑、特征或设备所执行的操作中的一些或全部。更具体而言,逻辑流程 600 可以通过标记模块 522-1、接收模块 522-2、缓冲器模块 522-3、判断模块 522-4、存储器控制器模块 522-5 或发送模块 522-6 中的一个或多个来实现。

[0064] 根据某些示例,逻辑流程 600 在框 602 还可以包括标记主机设备上的处理器插槽中所包括的处理器电路的缓存中维护的一个或多个缓冲器。对于这些其他示例,一个或多个缓冲器可以由耦合到主机设备的包括设备 500 的 NW I/O 设备来标记,以指出使用一个或多个缓冲器来集结通过耦合到 NW I/O 设备的一个或多个网络连接接收到的或传输的数据,该数据与访问由耦合到主机设备的存储器控制器控制的存储设备的请求相关联。例如,标记模块 522-1 可以提供标记标识信息 505 以标记用于集结接收到的数据的一个或多个缓冲器。

[0065] 在某些示例中,逻辑流程 600 在框 604 可以接收数据帧,该数据帧包括远程设备访问存储设备的请求。例如,接收模块 522-2 可以接收包括远程设备的请求的数据帧 510。

[0066] 根据某些示例, 逻辑流程 600 在框 606 可以基于请求是否要从存储设备读取数据或将数据存储到存储设备, 将与请求相关联的数据转发到一个或多个标记缓冲器或从一个或多个标记缓冲器接收该数据。例如, 判断模块 522-4 可以判断数据帧 510 中所包括的访问请求是否是读取或存储到存储设备。作为读取的判断的结果, 存储器控制器模块 422-5 可以导致数据读取 435 中所包括的数据被从存储设备中读取, 此数据可以被集结到一个或多个标记的缓冲器。然后, 缓冲器模块 422-3 可以从一个或多个标记的缓冲器拉取数据读取 435 中所包括的数据, 并将数据提供到发送模块 422-6。然后, 发送模块 422-6 可以将数据读取 435 中所包括的数据发送到远程设备。可另选地, 作为存储的判断的结果, 缓冲器模块 422-3 可以将要存储的数据 440 中所包括的数据集结到标记的一个或多个缓冲器。对于此替代方案, 集结的数据可以从数据帧 410 中获取。然后, 存储器控制器模块 422-5 可以导致此集结的要存储的数据 440 中所包括的数据存储到存储设备。

[0067] 图 7 示出了计算机可读介质 700 的示例。计算机可读介质 700 可以包括制品。在某些示例中, 计算机可读介质 700 可以包括任何非瞬态的计算机可读介质或机器可读的介质, 诸如光学存储、磁存储或半导体存储。计算机可读介质 700 可以存储各种类型的计算机可执行指令, 诸如实现逻辑流程 500 或逻辑流程 600 的指令。计算机可读的或机器可读的存储介质的示例可包括能够存储电子数据的任何有形的介质, 包括易失性存储器或非易失性存储器, 可移动或不可移动存储器, 可擦除或不可擦存储器, 可写入或可重写的存储器等等。计算机可执行指令的示例可包括任何合适类型的代码, 如源代码、已编译的代码、已解释的代码、可执行代码、静态代码、动态代码、面向对象的代码、可视代码等等。在此上下文中, 示例不受限制。

[0068] 图 8 示出了示例 NW I/O 设备 800。在某些示例中, 如图 8 所示, NW I/O 设备 800 可以处理组件 840、其他平台组件或通信接口 860。根据某些示例, 如上文所提及的, NW I/O 设备 800 可以在耦合到主机或客户端设备的 NWI/O 设备中实现。在某些示例中, NW I/O 设备 800 可以或可能不包括存储子系统, 存储子系统进一步包括诸如 SSD 之类的存储设备。

[0069] 根据某些示例, 处理组件 840 可以执行设备 400 和 / 或计算机可读介质 700 的处理操作或逻辑。处理组件 840 可包括各硬件要素、软件要素或两者的组合。硬件元件的示例可包括器件、逻辑器件、组件、处理器、微处理器、电路、处理器电路、电路元件 (例如, 晶体管、电阻器、电容器、感应器等等)、集成电路、专用集成电路 (ASIC)、可编程逻辑器件 (PLD)、数字信号处理器 (DSP)、现场可编程门阵列 (FPGA)、存储器单元、逻辑门、寄存器、半导体器件、芯片、微芯片、芯片组等等。软件元件的示例可以包括软件组件、程序、应用、计算机程序、应用程序、设备驱动程序、系统程序、软件开发程序、机器程序、操作系统软件、中间件、固件、软件模块、例程、子例程、函数、方法、过程、软件接口、应用程序编程接口 (API)、指令集、计算代码、计算机代码、代码段、计算机代码段、字、值、符号, 或其任何组合。判断一个示例是否使用硬件元件或软件元件来示例可以根据任意数量的因素而不同, 如所希望的计算速率、功率级别、耐热性、处理周期预算、输入数据速率、输出数据速率、存储器资源、数据总线速度, 及其他设计或性能约束, 如根据需要给定示例的需要。

[0070] 在某些示例中, 其他 I/O 组件 850 可以包括, 但不仅限于, 一个或多个较高速度的存储器单元形式的各种类型的存储介质, 诸如 ROM、RAM、DRAM、DDRAM、SDRAM、SRAM、PROM、EPROM、EEPROM、闪存或适用于存储信息的任何其他类型的存储介质。

[0071] 在某些示例中,通信接口 860 可以包括支持通信接口的逻辑和 / 或特征。对于这些示例,通信接口 860 可以包括根据各种通信协议或标准来操作以通过直接或网络通信链路进行通信的一种或多种通信接口。直接通信可以通过使用在一个或多个行业标准(包括后代和变体)(诸如那些与 PCIe 规范、NVMe 规范、RDMA 协议规范相关联的(例如, iWARP、RoCE、Infiniband、IEEE802.3 或 802.11 规范、RFC791、RFC793 或 FCoE 标准)中所描述的通信协议或标准来进行。

[0072] NW I/O 设备 800 的组件和特征可以使用不连续线路、专用集成电路 (ASIC)、逻辑门和 / 或单芯片体系结构的任何组合来实现。进一步地,在适当的情况下,NW I/O 设备 800 的特征可以使用微控制器、可编程逻辑阵列和 / 或微处理器或前面的各项的任何组合来实现。值得注意的是,硬件、固件和 / 或软件元件可以共同地或分别地称为“逻辑”或“电路”。

[0073] 应该理解,图 8 的框图所示出的示例性 NW I/O 设备 800 可以表示许多潜在的实现的一个功能上描述性的示例。相应地,附图中所描绘的块功能的分割、省略或包括不会推断,用于实现这些功能的硬件组件、电路、软件和 / 或元件将一定被分割,省略,或包括在各实施例中。

[0074] 应该理解,图 8 的框图所示出的示例性 NW I/O 设备 800 可以表示许多潜在的实现的一个功能上描述性的示例。相应地,附图中所描绘的块功能的分割、省略或包括不会推断,用于实现这些功能的硬件组件、电路、软件和 / 或元件将一定被分割,省略,或包括在各实施例中。

[0075] 至少一个示例的一个或多个方面可以通过存储在至少一种机器可读介质上的代表性的指令来实现,指令表示处理器内的各种逻辑,指令在由机器、计算设备或系统读取时使机器、计算设备或系统制造执行此处所描述的技术的逻辑。被称为“IP 核”的这样的表示可以存储在有形的机器可读介质中,并提供给各种客户或生产设施,以加载到实际制造逻辑或处理器的制造机器中。

[0076] 各示例可以使用硬件元件、软件元件或两者的组合来实现。在某些示例中,硬件元件可包括器件、逻辑器件、组件、处理器、微处理器、电路、处理器组件、电路元件(例如,晶体管、电阻器、电容器、感应器等等)、集成电路、专用集成电路 (ASIC)、可编程逻辑器件 (PLD)、数字信号处理器 (DSP)、现场可编程门阵列 (FPGA)、存储器单元、逻辑门、寄存器、半导体器件、芯片、微芯片、芯片组等等。在某些示例中,软件元件可以包括软件组件、程序、应用程序、计算机程序、系统程序、机器程序、操作系统软件、中间件、固件、软件模块、例程、子例程、函数、方法、过程、软件接口、应用程序接口 (API)、指令集、计算代码、计算机代码、代码段、计算机代码段、字、值、符号,或其任何组合。判断一个示例是否使用硬件元件或软件元件来实现可以根据任意数量的因素而不同,如所希望的计算速率、功率级别、耐热性、处理周期预算、输入数据速率、输出数据速率、存储器资源、数据总线速度,及其他设计或性能约束,如根据需要给定实现的需要。

[0077] 某些示例可以包括制品或至少一种计算机可读介质。计算机可读介质可以包括存储逻辑的非瞬态的存储介质。在某些示例中,非瞬态的存储介质可包括一种或多种类型的能够存储电子数据的计算机可读存储介质,包括易失性存储器或非易失性存储器,可移动或不可移动存储器,可擦除或不可擦存储器,可写入或可重写的存储器等等。在某些示例中,逻辑可包括各种软件元件,如软件组件、程序、应用、计算机程序、应用程序、系统程

序、机器程序、操作系统软件、中间件、固件、软件模块、例程、子例程、函数、方法、过程、软件接口、API、指令集、计算代码、计算机代码、代码段、计算机代码段、字、值、符号,或其任何组合。

[0078] 根据某些示例,计算机可读介质可以包括存储或维护指令的非瞬态的存储介质,所述指令,当由机器、计算设备或系统执行时,导致机器、计算设备或系统执行根据所描述的示例的方法和/或操作。指令可包括任何合适类型的代码,如源代码、已编译的代码、已解释的代码、可执行代码、静态代码、动态代码等等。指令可以根据预定义的计算机语言、方式或语法来实现,以便指示机器、计算设备或系统来执行某一功能。指令可以使用任何合适的高级别的、低级别的、面向对象的、可视的、已编译的和/或解释性编程语言来实现。

[0079] 可以使用表达“在一个示例中”或“示例”以及它们的派生词来描述某些示例。这些术语意味着,参考示例所描述的特定功能、结构或特征包括在至少一个示例中。在本说明书中的不同位置出现短语“在一个示例中”不一定都是指同一个示例。

[0080] 可以使用表达“耦合”和“连接”以及它们的派生词来描述某些示例。这些术语不一定作为彼此的同义词。例如,使用术语“连接”和/或“耦合”的描述可以表示,两个或更多元件彼此处于直接的物理或电接触的状态。然而,术语“耦合”也可以意味着,两个或更多彼此不直接接触,但是仍彼此合作或进行交互的元件。

[0081] 在某些示例中,耦合到主机设备的网络 I/O 设备的示例设备可以包括电路。示例第一设备还可以包括由电路执行以接收数据帧的接收模块,数据帧包括访问在网络 I/O 设备中维护的存储子系统的请求。存储子系统可以包括 SSD。示例第一设备还可以包括由电路执行以使用主机设备中的处理器插槽中所包括的处理器电路的缓存中维护的一个或多个缓冲器的缓冲器模块。一个或多个缓冲器可以被用来交换对于请求的控制信息。控制信息可以与由处理器电路执行的协议堆栈交换。示例第一设备还可以包括判断模块,该判断模块由电路执行以基于交换控制信息,判断是否从存储子系统数据中读取与请求相关联的数据或将其存储存储子系统。

[0082] 根据某些示例,示例设备还可以包括基于判断模块的判断来控制对存储子系统的访问的存储器控制器模块。

[0083] 在示例设备的某些示例中,存储器控制器模块和至少 SSD 可以被配置为根据包括 PCIe 基础规范修订版 3.0 或 NVMe 规范,修订版 1.1 的一个或多个行业标准操作。

[0084] 根据示例设备的某些示例,接收模块可以通过与主机设备的本地连接,从主机设备,或通过与远程设备的网络连接,从远程设备,接收数据帧,数据帧根据同一协议格式配置,无论是从主机设备还是远程设备接收到的。

[0085] 在示例设备的某些示例中,同一个协议格式可以包括 FCoE、iWARP、Infiniband 或 RoCE。

[0086] 根据示例设备的某些示例,接收模块可以从主机设备接收数据帧,数据帧中所包括的请求是由处理器电路执行的应用访问存储子系统。

[0087] 根据示例设备的某些示例,来自应用的请求可以将与数据帧相关联的数据存储到存储子系统。在满足将数据存储到存储子系统的请求之前,数据至少临时存储到一个或多个缓冲器。

[0088] 在示例设备的某些示例中,来自应用的请求可以从存储子系统读取数据。在满足

从存储子系统读取数据的请求之后,读取的数据至少临时存储到一个或多个缓冲器。

[0089] 根据示例设备的某些示例,对于请求的控制信息可以包括与数据帧相关联的首部信息,以促进由处理器电路执行的协议堆栈作出的关于是否接受或拒绝请求的判断。

[0090] 在示例设备的某些示例中,接收模块可以从远程设备接收数据帧,请求是远程设备将与数据帧相关联的数据存储到存储子系统。数据至少临时存储到在网络 I/O 设备中维护的存储器,然后,在请求被接受之后被导致存储到存储子系统。

[0091] 根据示例设备的某些示例,接收模块可以从远程设备接收数据帧,请求是远程设备从存储子系统读取数据。

[0092] 在某些示例中,示例设备也可以包括由电路执行的发送模块,该发送模块在请求被接受之后通过网络连接将读取的数据发送到远程设备。

[0093] 根据示例设备的某些示例,与主机设备的本地连接被配置为根据包括 PCIe 基础规范修订版 3.0 的一个或多个行业标准操作。

[0094] 在某些示例中,在耦合到主机设备的网络 I/O 设备上的实现的示例第一方法可以包括接收数据帧,其中包括访问网络 I/O 设备中维护的存储子系统的请求。对于这些示例,存储子系统可以包括 SSD。示例第一方法还可以包括使用主机设备处的处理器插槽中所包括的处理器电路的缓存中维护的一个或多个缓冲器。一个或多个缓冲器可以被配置为交换对于请求的控制信息。控制信息可以与由处理器电路执行的协议堆栈交换。示例第一方法还可以包括基于交换的控制信息来判断是否从存储子系统中读取或向其存储与请求相关联的数据。

[0095] 根据示例第一方法的某些示例,可以通过与主机设备的本地连接,从主机设备,或通过远程设备的网络连接,从远程设备,接收数据帧。对于这些示例,数据帧可以根据同一协议格式配置,无论是从主机设备还是远程设备接收到的。

[0096] 在示例第一方法的某些示例中,同一个协议格式可以包括 FCoE、iWARP、Infiniband 或 RoCE。

[0097] 根据示例第一方法的某些示例,可以从主机设备接收数据帧,数据帧中所包括的请求是由处理器电路执行的应用访问存储子系统。

[0098] 根据示例第一方法的某些示例,来自应用的请求可以将与数据帧相关联的数据存储到存储子系统。在满足将数据存储到存储子系统的请求之前,数据至少临时存储到一个或多个缓冲器。

[0099] 根据示例第一方法的某些示例,来自应用的请求可以是存储子系统读取数据。在满足从存储子系统读取数据的请求之后,读取的数据至少临时存储到一个或多个缓冲器。

[0100] 在示例第一方法的某些示例中,包括与数据帧相关联的首部信息的对于请求的控制信息可以是促进由处理器电路执行的协议堆栈作出的关于是否接受或拒绝请求的判断。

[0101] 根据示例第一方法的某些示例,可以从远程设备接收数据帧,请求是远程设备将与数据帧相关联的数据存储到存储子系统。数据至少临时存储到在网络 I/O 设备中维护的存储器,然后,在请求被接受之后存储到存储子系统。

[0102] 在示例第一方法的某些示例中,可以从远程设备接收数据帧,请求是远程设备从存储子系统读取数据。对于这些示例,在请求被接受之后,可以通过网络连接将读取的数据

发送到远程设备。

[0103] 根据示例第一方法某些示例,存储子系统中所包括的 SSD 可以具有非易失性存储器,非易失性存储器包括 3 维交叉点存储器、闪存、铁电存储器、硅-氧化物-氮化物-氧化物-硅 (SONOS) 存储器、聚合物存储器、纳米线、铁电晶体管随机存取存储器 (FeTRAM 或 FeRAM)、纳米线或电可擦可编程只读存储器 (EEPROM) 中的至少一种。

[0104] 在示例第一方法的某些示例中,与主机设备的本地连接被配置为根据包括 PCIe 基础规范修订版 3.0 的一个或多个行业标准操作。

[0105] 在某些示例中,包括多个指令的第一至少一种机器可读的介质,所述指令,响应于在耦合到主机设备的网络 I/O 设备上执行,导致网络 I/O 设备接收包括访问在网络 I/O 设备中维护的存储子系统的请求的数据帧,存储子系统包括 SSD。指令还可以导致网络 I/O 设备使用主机设备处的处理器插槽中所包括的处理器电路的缓存中维护的一个或多个缓冲器。一个或多个缓冲器可以被配置为交换对于请求的控制信息。控制信息可以与由处理器电路执行的协议堆栈交换。指令还可以导致网络 I/O 设备基于交换的控制信息来判断是否从存储子系统中网络 I/O 设备取或向其存储与请求相关联的数据。

[0106] 根据示例第一至少一个机器可读的介质的某些示例,可以通过与主机设备的本地连接,从主机设备,或通过与远程设备的网络连接,从远程设备,接收数据帧。数据帧可以根据同一协议格式配置,无论是从主机设备还是远程设备接收到的。

[0107] 在第一至少一个机器可读的介质的某些示例中,同一个协议格式可以包括 FCoE、iWARP、Infiniband 或 RoCE。

[0108] 根据第一至少一个机器可读的介质的某些示例,可以从主机设备接收数据帧,数据帧中所包括的请求是由处理器电路执行的应用访问存储子系统。

[0109] 根据第一至少一个机器可读的介质的某些示例,来自应用的请求可以将与数据帧相关联的数据存储到存储子系统。对于这些示例,在满足将数据存储到存储子系统的请求之前,数据至少临时存储到一个或多个缓冲器。

[0110] 根据第一至少一个机器可读的介质的某些示例,来自应用的请求可以是存储子系统读取数据。在满足从存储子系统读取数据的请求之后,读取的数据至少临时存储到一个或多个缓冲器。

[0111] 根据第一至少一个机器可读的介质的某些示例,对于请求的控制信息可以包括与数据帧相关联的首部信息,以促进由处理器电路执行的协议堆栈作出的关于是否接受或拒绝请求的判断。

[0112] 根据第一至少一个机器可读的介质的某些示例,可以从远程设备接收数据帧,请求是远程设备将与数据帧相关联的数据存储到存储子系统。指令还可以导致网络 I/O 设备至少临时将数据存储到在网络 I/O 设备中维护的存储器,然后,在请求被接受之后,导致数据存储到存储子系统。

[0113] 在第一至少一个机器可读的介质的某些示例中,可以从远程设备接收数据帧,请求是远程设备从存储子系统读取数据。指令还可以导致网络 I/O 设备在请求被接受之后将读取的数据发送到远程设备。

[0114] 根据第一至少一个机器可读的介质的某些示例,与主机设备的本地连接被配置为根据包括 PCIe 基础规范修订版 3.0 的一个或多个行业标准操作。

[0115] 在某些示例中,在耦合到主机设备的网络 I/O 设备上实现的示例第二方法可以包括标记主机设备中的处理器插槽中所包括的处理器电路的缓存中维护的一个或多个缓冲器。一个或多个缓冲器可以由耦合到主机设备的网络 I/O 设备来标记,以指出使用一个或多个缓冲器来集结通过耦合到网络 I/O 设备的一个或多个网络连接接收到的或传输的数据。数据可以与访问由耦合到主机设备的存储器控制器控制的存储设备的请求相关联。示例第二方法还可以包括接收数据帧,其中包括远程设备访问存储设备的请求。示例第二方法还可以包括基于请求是否将从存储设备读取数据或将数据存储到存储设备,将与请求相关联的数据转发到一个或多个标记的缓冲器或从它们接收。

[0116] 根据示例第二方法的某些示例,标记一个或多个缓冲器可以包括向一个或多个缓冲器指定标识符以指出使用一个或多个标记的缓冲器来基于从远程设备接收到的一个或多个数据帧中所包括的一个或多个请求来访问存储设备。

[0117] 在示例第二方法的某些示例中,存储设备包括 SSD 和存储器控制器,SSD 可以被配置为根据包括 PCIe 基础规范修订版 3.0 或 NVMe 规范,修订版 1.1 的一个或多个行业标准操作。

[0118] 根据示例第二方法的某些示例,SSD 具有非易失性存储器,所述非易失性存储器包括 3 维交叉点存储器、闪存、铁电存储器、硅-氧化物-氮化物-氧化物-硅 (SONOS) 存储器、聚合物存储器、纳米线、铁电晶体管随机存取存储器 (FeTRAM 或 FeRAM)、纳米线或电可擦可编程只读存储器 (EEPROM) 中的至少一种。

[0119] 在示例第二方法的某些示例中,数据帧可以根据可以包括 FCoE、iWARP、Infiniband 或 RoCE 的协议格式来配置。

[0120] 在某些示例中,包括多个指令的第二至少一个机器可读的介质,所述指令,响应于在耦合到主机设备的网络 I/O 设备上执行,导致网络 I/O 设备标记主机设备上的处理器插槽中所包括的处理器电路的缓存中维护的一个或多个缓冲器。一个或多个缓冲器可以由网络 I/O 设备来标记,以指出使用一个或多个缓冲器来集结通过耦合到网络 I/O 设备的一个或多个网络连接接收到的或传输的数据。数据可以与访问由耦合到主机设备的存储器控制器控制的存储设备的请求相关联。指令还可以导致网络 I/O 设备接收包括远程设备访问存储设备的请求的数据帧。指令还可以导致网络 I/O 设备基于请求是否是从存储设备读取数据或将数据存储到存储设备中,将与请求相关联的数据转发到一个或多个标记的缓冲器,或从它们接收。

[0121] 根据第二至少一个机器可读的介质的某些示例,标记一个或多个缓冲器的指令可以包括导致网络 I/O 设备基于从远程设备接收到的一个或多个数据帧中所包括的一个或多个请求,向一个或多个缓冲器指定标识符以指出使用一个或多个标记的缓冲器来访问存储设备的指令。

[0122] 在第二至少一个机器可读的介质的某些示例中,存储设备可以包括 SSD 和存储器控制器,SSD 可以被配置为根据可以包括 PCIe 基础规范修订版 3.0 或 NVMe 规范,修订版 1.1 的一个或多个行业标准操作。

[0123] 根据第二至少一个机器可读的介质的某些示例,SSD 可以具有非易失性存储器,该非易失性存储器包括 3 维交叉点存储器、闪存、铁电存储器、硅氧化物氮化物氧化物硅 (SONOS) 存储器、聚合物存储器、纳米线、铁电晶体管随机存取存储器 (FeTRAM 或 FeRAM)、纳

米线或电可擦可编程只读存储器 (EEPROM) 中的至少一种。

[0124] 在第二至少一个机器可读的介质的某些示例中,数据帧可以根据可以包括 FCoE、iWARP、Infiniband 或 RoCE 的协议格式来配置。

[0125] 要强调的是,提供可公开内容的摘要以符合 37C. F. R. 第 1.72(b) 段,要求将可使读者快速地弄清本技术公开的特征的摘要。但应理解,它不能被用来解释或限制权利要求的范围或含义。此外,在前面的“具体实施方式”中,可以看出,各种特点可以组合在一个示例中,以便简化说明。本发明的此方法不应被解释为反映带权利要求的各示例需要比每一个权利要求中明确地列举的特点更多的特点的意图。相反,如下面的权利要求所反映的,本发明的主题在于少于所公开的单一示例的所有特点。如此,下面的权利要求被包括到“详细描述”中,每一个权利要求本身也作为单独的示例。在所附权利要求书中,术语“包括”和“其特征在于”被用作相应的术语“包含”和“其中”的普通英语等效词。此外,术语“第一”、“第二”、“第三”等等只用作标记,并不旨在对它们的对象施加数值要求。

[0126] 尽管用结构特征和 / 或方法动作专用的语言描述了本主题,但可以理解,所附权利要求书中定义的主题不必限于上述具体特征或动作。相反,上述具体特征和动作是作为实现权利要求的示例形式公开的。

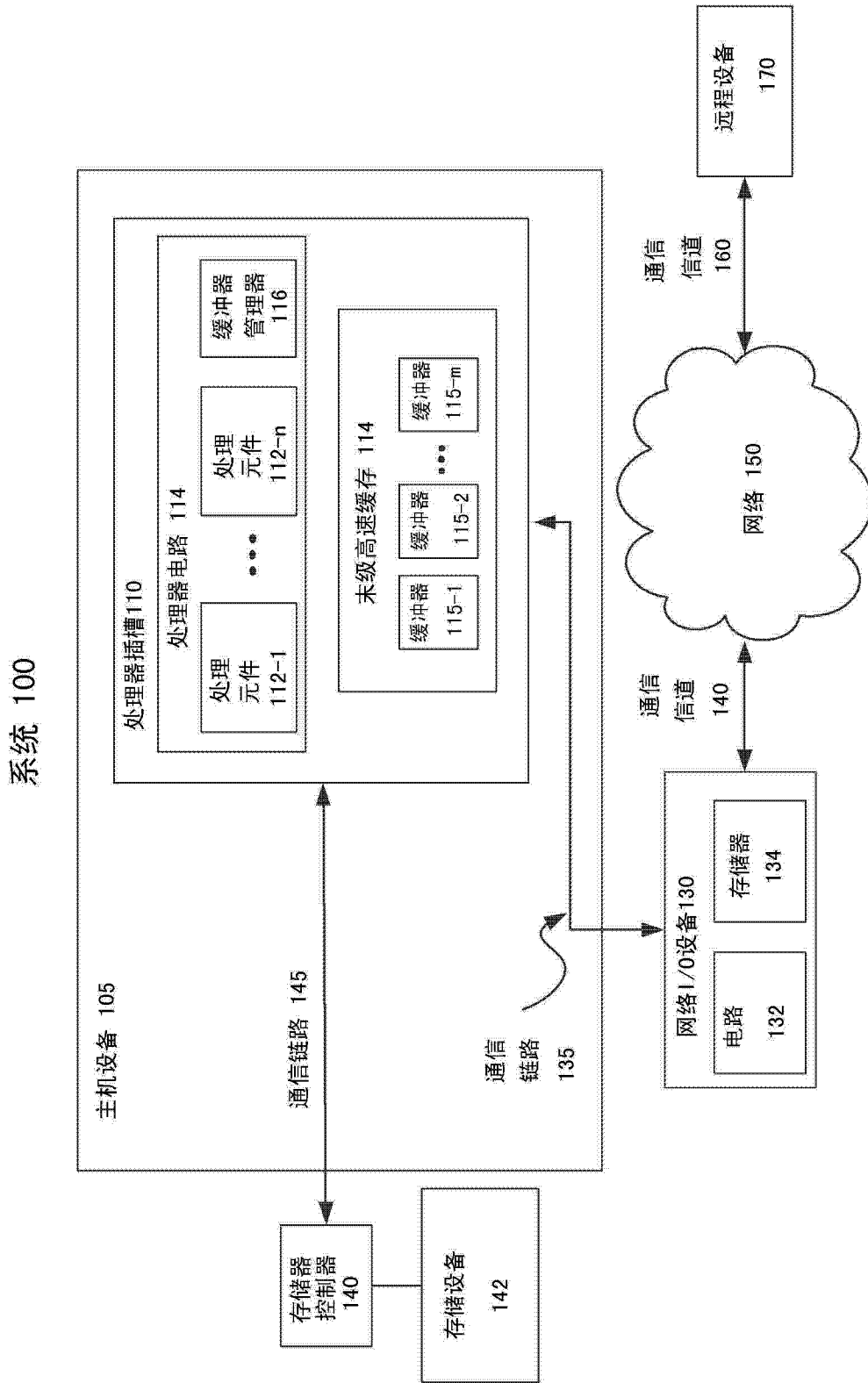


图 1

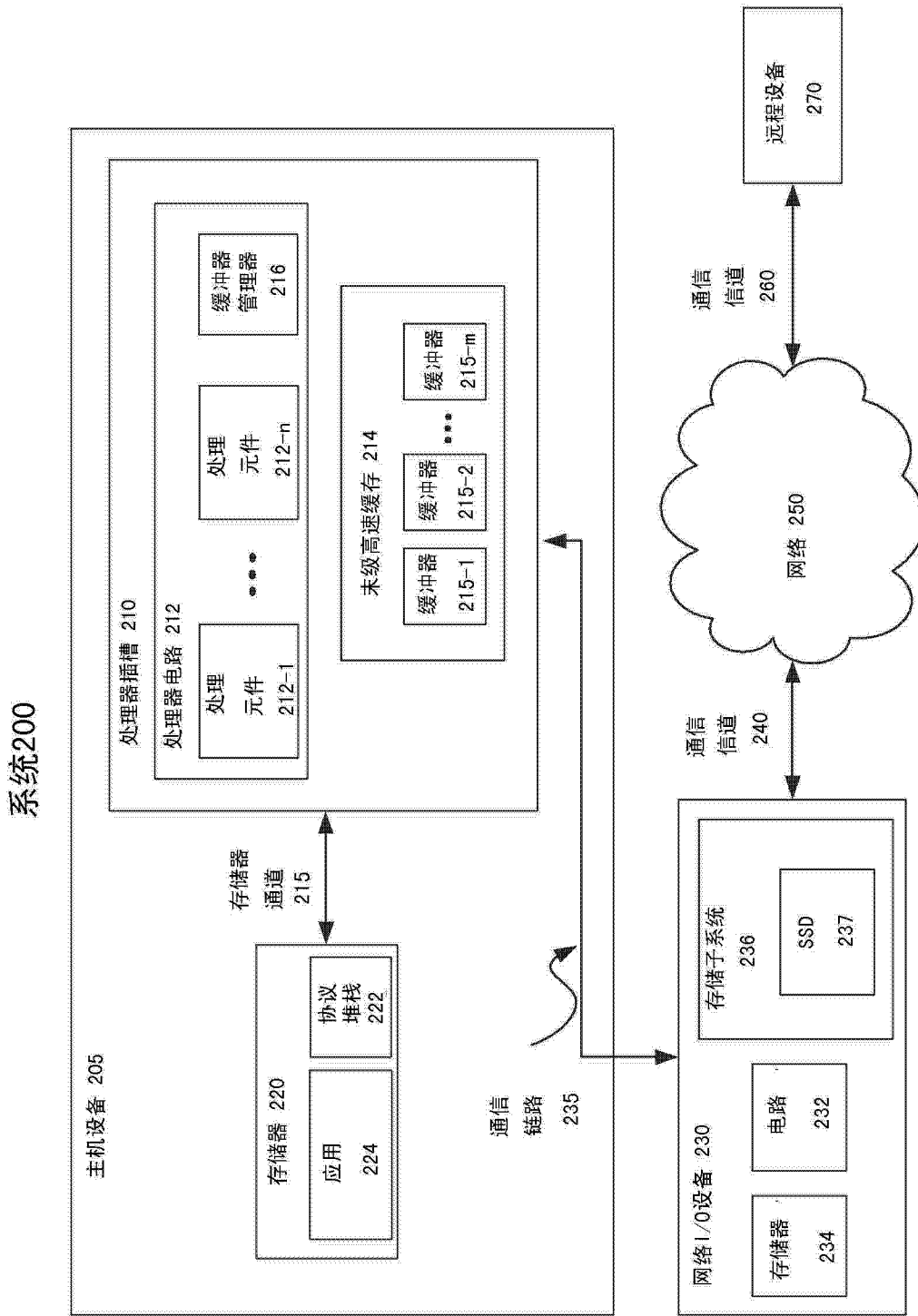


图 2

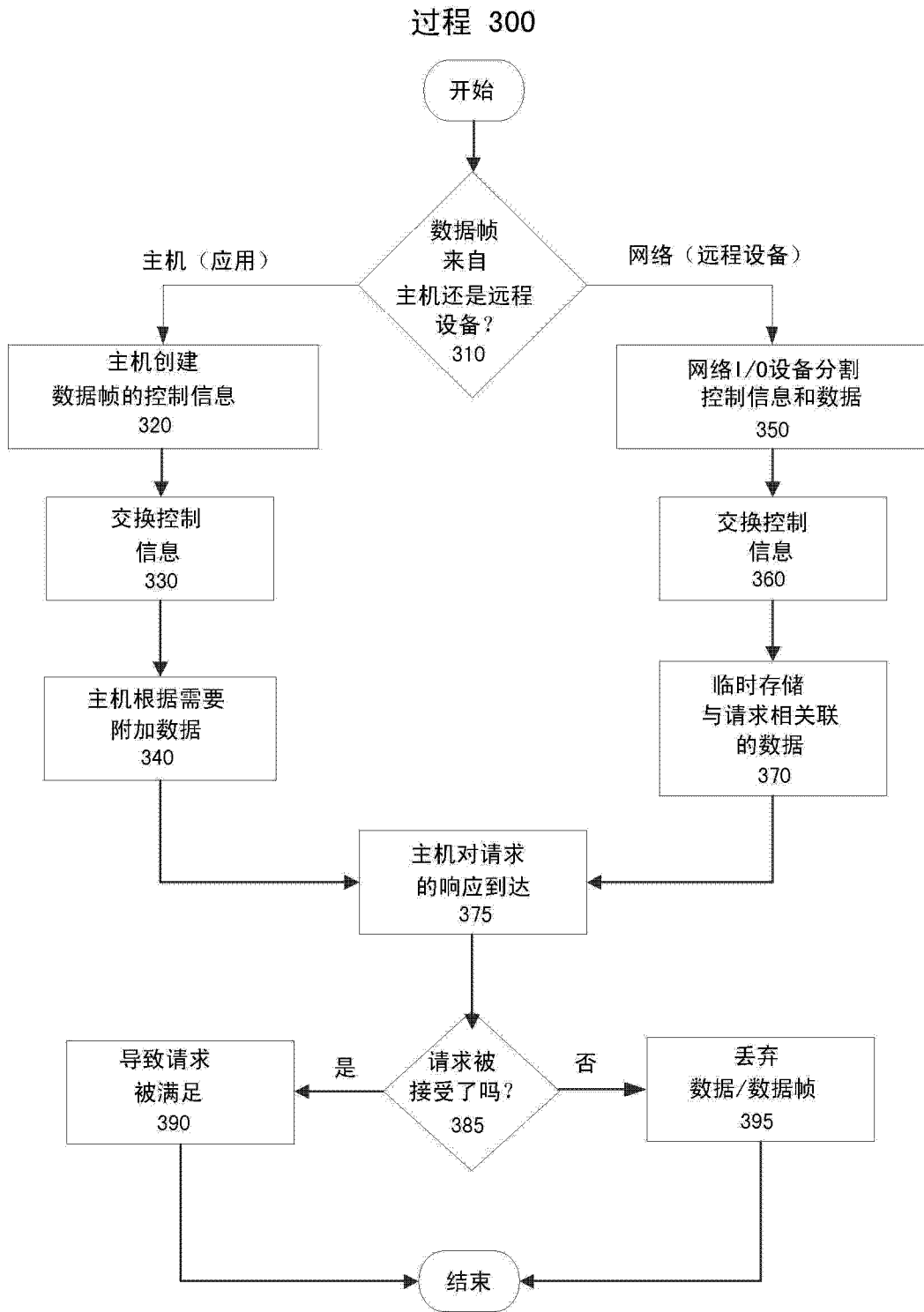


图 3



图 4

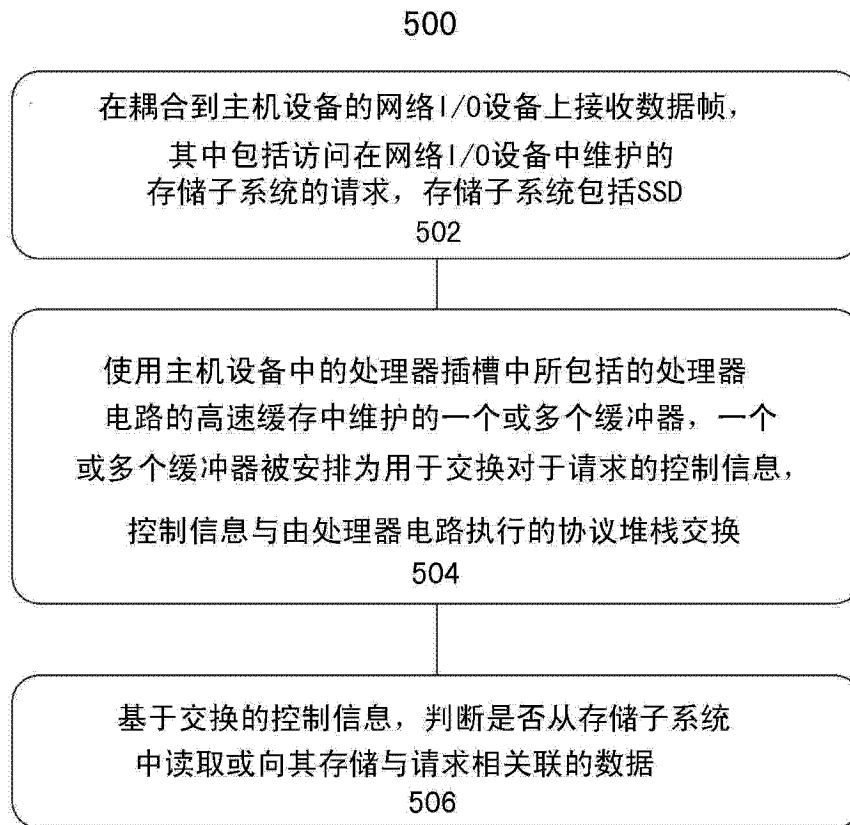


图 5

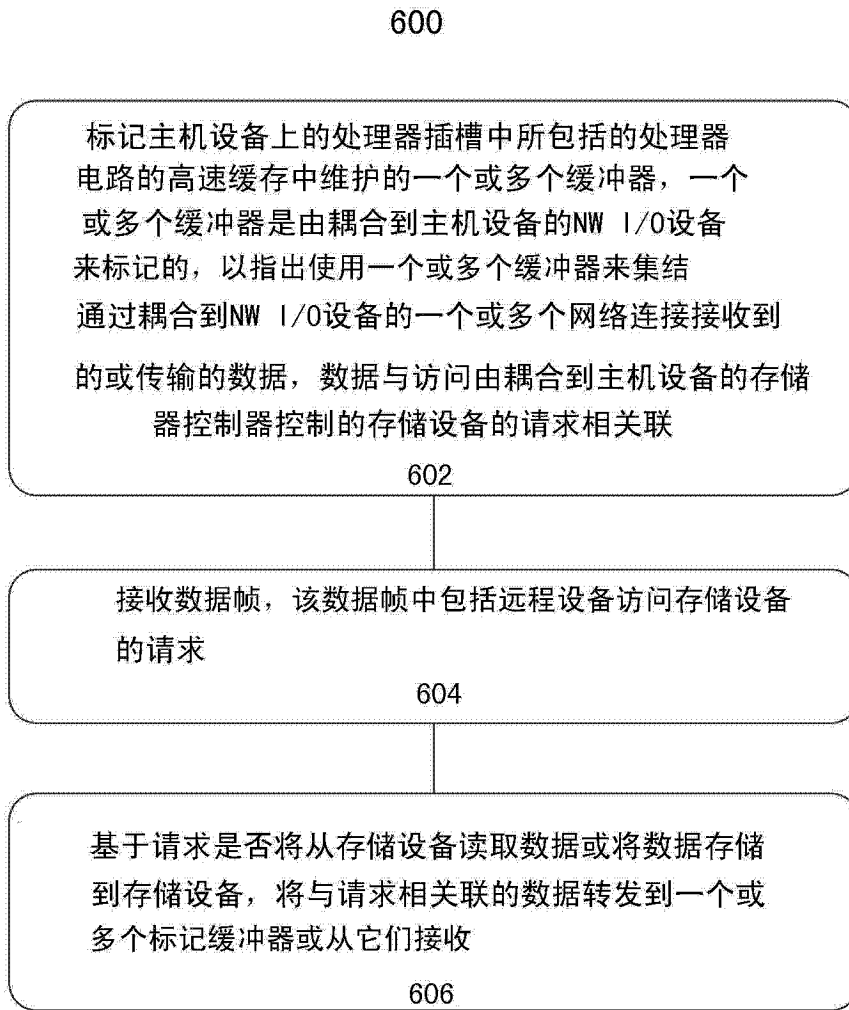


图 6



图 7

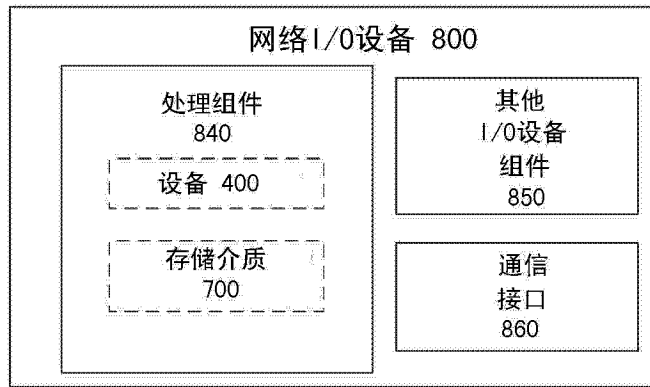


图 8