

(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国 际 局(43) 国际公布日
2016 年 12 月 1 日 (01.12.2016) WIPO | PCT(10) 国际公布号
WO 2016/188014 A1(51) 国际专利分类号:
G06F 3/06 (2006.01)

ERTY LLC); 中国北京市海淀区清华园清华大学照澜院商业楼 301 室, Beijing 100084 (CN)。

(21) 国际申请号: PCT/CN2015/091773

(81) 指定国 (除另有指明, 要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW。

(22) 国际申请日: 2015 年 10 月 12 日 (12.10.2015)

(84) 指定国 (除另有指明, 要求每一种可提供的地区保护): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG)。

(25) 申请语言: 中文

本国国际公布:

(26) 公布语言: 中文

— 包括国际检索报告(条约第 21 条(3))。

(30) 优先权: 201510278937.X 2015 年 5 月 27 日 (27.05.2015) CN

(71) 申请人: **北京百度网讯科技有限公司 (BEIJING Baidu NETCOM SCIENCE AND TECHNOLOGY CO., LTD.) [CN/CN]**; 中国北京市海淀区上地十街 10 号百度大厦 2 层, Beijing 100085 (CN)。(72) 发明人: **谢广军 (XIE, Guangjun)**; 中国北京市海淀区上地十街 10 号百度大厦 2 层, Beijing 100085 (CN)。 **王柏生 (WANG, Baisheng)**; 中国北京市海淀区上地十街 10 号百度大厦 2 层, Beijing 100085 (CN)。 **李志勇 (LI, Zhiyong)**; 中国北京市海淀区上地十街 10 号百度大厦 2 层, Beijing 100085 (CN)。(74) 代理人: **北京清亦华知识产权代理事务所 (普通合伙) (TSINGYIHUA INTELLECTUAL PROP-**

(54) Title: DATA STORAGE METHOD AND DEVICE

(54) 发明名称: 数据存储方法和装置

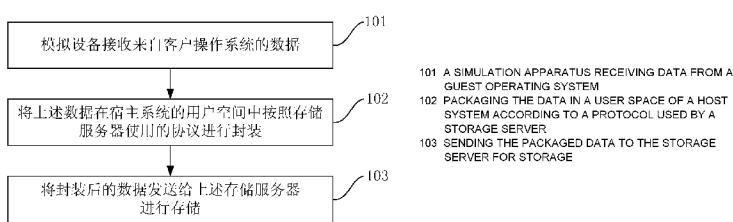


图 1

(57) Abstract: A data storage method and a device. The above-mentioned data storage method comprises: a simulation apparatus receiving data from a guest operating system (101); packaging the data in a user space of a host system according to a protocol used by a storage server (102); and sending the packaged data to the storage server for storage (103). The method can realize that data is stored on a storage server, thereby shortening the I/O path, improving the I/O throughput of a system, reducing the CPU load, and optimizing the system performance.

(57) 摘要: 一种数据存储方法和装置, 上述数据存储方法包括: 模拟设备接收来自客户操作系统的数据 (101); 将所述数据在宿主系统的用户空间中按照存储服务器使用的协议进行封装 (102); 将封装后的数据发送给所述存储服务器进行存储 (103)。该方法可以实现将数据存储在存储服务器上, 缩短了 IO 路径, 提高了系统的 IO 吞吐, 减轻了 CPU 的负载, 优化了系统性能。

数据存储方法和装置

相关申请的交叉引用

5 本申请要求北京百度网讯科技有限公司于 2015 年 5 月 27 日提交的、发明名称为“数
据存储方法和装置”的、中国专利申请号“201510278937.X”的优先权。

技术领域

本发明涉及互联网技术领域，尤其涉及一种数据存储方法和装置。

10

背景技术

对于一个虚拟操作系统来说，为了保证其数据可靠性，需要将数据保存在更可靠的云
存储服务器上。另外，将存储保存在另外一个位置，也是虚拟操作系统迁移的充分条件。

但是输入输出（Input Output；以下简称：IO）对系统的性能是至关重要的，因此，
15 将存储从本地移动到网络上的云存储，需要尽量减少由此带来的性能损失。

现有的技术方案一般都是采取互联网小型计算机接口（Internet Small Computer
System Interface；以下简称：iSCSI）的方式，在本地物理机上创建一个小型计算机接口
（Small Computer System Interface；以下简称：SCSI）设备，将这个设备分配给虚拟操
作系统。SCSI 设备在底层通过 iSCSI 将 SCSI 命令发送给 iSCSI 目标器（iSCSI target），
20 一般云存储服务器都有自己的数据交换协议，所以这其中还要涉及 SCSI 命令和云存储服
务器的协议的转换。iSCSI target 可以部署在本地或者一个固定的服务器上，或者云存储服
务器上。但是上述方案增加了 IO 路径的长度，增加了中央处理器（Central Processing
Unit；以下简称：CPU）的负载，对系统性能的消耗较大。

25 发明内容

本发明的目的旨在至少在一定程度上解决相关技术中的技术问题之一。

为此，本发明的第一个目的在于提出一种数据存储方法。该方法可以实现将数据存储
在存储服务器上，缩短了 IO 路径，提高了系统的 IO 吞吐，减轻了 CPU 的负载，优化了系
统性能。

30 本发明的第二个目的在于提出一种数据存储装置。

为了实现上述目的，本发明第一方面实施例的数据存储方法，包括：模拟设备接收来
自客户操作系统的数据；将所述数据在宿主系统的用户空间中按照存储服务器使用的协议

进行封装；将封装后的数据发送给所述存储服务器进行存储。

本发明实施例的数据存储方法，模拟设备接收来自客户操作系统的数据之后，将上述数据在宿主系统的用户空间中按照存储服务器使用的协议进行封装，然后将封装后的数据发送给上述存储服务器进行存储，从而可以实现将数据存储在存储服务器上，缩短了 I/O 路径，提高了系统的吞吐，减轻了 CPU 的负载，优化了系统性能。

为了实现上述目的，本发明第二方面实施例的数据存储装置，设置在模拟设备中，所述数据存储装置包括：接收模块，用于接收来自客户操作系统的数据；封装模块，用于将所述接收模块接收的数据在宿主系统的用户空间中按照存储服务器使用的协议进行封装；发送模块，用于将所述封装模块封装后的数据发送给所述存储服务器进行存储。

本发明实施例的数据存储装置，实现为虚拟机监控器中的一个模拟设备，接收模块接收数据之后，封装模块将上述数据在宿主系统的用户空间中按照存储服务器使用的协议进行封装，然后发送模块将封装后的数据发送给上述存储服务器进行存储，从而可以实现将数据存储在存储服务器上，缩短了 I/O 路径，提高了系统的 I/O 吞吐，减轻了 CPU 的负载，优化了系统性能。

为了实现上述目的，本发明第三方面实施例提供了一种电子设备，包括：一个或者多个处理器；存储器；一个或者多个程序，所述一个或者多个程序存储在所述存储器中，当被所述一个或者多个处理器执行时进行如下操作：模拟设备接收来自客户操作系统的数据；将所述数据在宿主系统的用户空间中按照存储服务器使用的协议进行封装；将封装后的数据发送给所述存储服务器进行存储。

为了实现上述目的，本发明第四方面实施例提供了一种非易失性计算机存储介质，所述计算机存储介质存储有一个或者多个模块，当所述一个或者多个模块被执行时进行如下操作：模拟设备接收来自客户操作系统的数据；将所述数据在宿主系统的用户空间中按照存储服务器使用的协议进行封装；将封装后的数据发送给所述存储服务器进行存储。

本发明附加的方面和优点将在下面的描述中部分给出，部分将从下面的描述中变得明显，或通过本发明的实践了解到。

附图说明

本发明上述的和/或附加的方面和优点从下面结合附图对实施例的描述中将变得明显和容易理解，其中：

图 1 为本发明数据存储方法一个实施例的流程图；

图 2 为本发明数据存储方法的应用场景一个实施例的示意图；

图 3 为本发明数据存储装置一个实施例的结构示意图。

具体实施方式

下面详细描述本发明的实施例，所述实施例的示例在附图中示出，其中自始至终相同或类似的标号表示相同或类似的元件或具有相同或类似功能的元件。下面通过参考附图描述的实施例是示例性的，仅用于解释本发明，而不能理解为对本发明的限制。相反，本发明的实施例包括落入所附加权利要求书的精神和内涵范围内的所有变化、修改和等同物。

图 1 为本发明数据存储方法一个实施例的流程图，如图 1 所示，该数据存储方法可以包括：

步骤 101，模拟设备接收来自客户操作系统的数据。

10 步骤 102，将上述数据在宿主系统的用户空间中按照存储服务器使用的协议进行封装。

步骤 103，将封装后的数据发送给上述存储服务器进行存储。

本实施例中，上述模拟设备是在上述宿主系统上基于虚拟机监控器（Virtual Machine Monitor），例如快速模拟器（Quick Emulator；以下简称：Qemu）实现的，上述模拟设备为外设部件互连标准（Peripheral Component Interconnect；以下简称：PCI）接口的块设备；上述客户操作系统使用块设备驱动来驱动上述模拟设备，上述客户操作系统运行于虚拟机监控器之上。

本发明图 1 所示实施例提供的数据存储方法可以应用在图 2 所示的架构中，图 2 为本发明数据存储方法的应用场景一个实施例的示意图。

参见图 2，本发明基于虚拟机监控器 Qemu，设计并实现了一个模拟设备。为了避免修改客户操作系统，减少对客户操作系统的影响，该模拟设备被模拟为一个块设备，客户操作系统中无需实现任何对该模拟设备的驱动，直接使用块设备驱动即可；为了更高效的在客户操作系统和宿主系统之间进行通信，模拟设备采用虚拟输入输出（VirtIO）标准实现数据通信；为了避免存储服务器升级等带来的影响，模拟设备与存储服务器进行了解耦，本发明在宿主系统上设计实现了一个存储服务器的代理服务器。

25 为了使客户操作系统不经修改就可以使用本发明图 1 所示实施例提供的方法，这一点尤其是对于那些用户自定义的客户操作系统非常有意义，本发明将模拟设备注册为 PCI 接口的 VirtIO 块设备，客户操作系统发现是 PCI 总线上的 VirtIO 块设备后，将使用虚拟输入输出块设备（virtio-blk）驱动来驱动该模拟设备。

30 本发明图 1 所示实施例中，上述模拟设备与上述客户操作系统之间通过 VirtIO 的方式共享内存，模拟设备与客户操作系统之间通过传递上述数据所在内存的物理地址实现上述数据的访问，从而实现了数据的零复制。

本发明图 1 所示实施例中，将封装后的数据发送给存储服务器进行存储可以为：模拟

设备将封装后的数据通过网络通信协议栈发送给上述存储服务器进行存储。

本发明中，模拟设备接收到数据之后，不再经过宿主系统的内核 I/O 栈，而是在宿主系统的用户空间中按照存储服务器的协议进行封装后，通过网络通信协议栈直接发送给远端的存储服务器。另外，本实施例对 Qemu 中的 I/O 路径进行了优化，缩短了 I/O 路径。

5 具体地，模拟设备将封装后的数据通过网络通信协议栈发送给上述存储服务器进行存储可以为：模拟设备将上述封装后的数据发送给存储服务器的代理服务器，由上述代理服务器将封装后的数据通过网络通信协议栈发送给存储服务器进行存储。其中，上述网络通信协议栈可以为传输控制协议（Transmission Control Protocol；以下简称：TCP）/因特网协议（Internet Protocol；以下简称：IP）协议栈，当然本发明实施例并不仅限于此，
10 网络通信协议栈还可以为其他的网络通信协议栈，只要可以实现模拟设备与存储服务器之间的网络通信即可。

在生产环境上，频繁升级 Qemu 的代价和风险都是比较大的，因此，本发明在模拟设备和远端存储服务器之间实现了一个存储服务器的代理服务器，将业务逻辑分离到代理服务器中来。而代理服务器与 Qemu 中的模拟设备之间只是进行简单的数据通信，不受具体通信
15 协议细节变更的影响，无论是协议升级，或者是修复缺陷，Qemu 均无需变更。

为了提高 I/O 吞吐，本发明将控制平面和数据通道分开实现，使用一个独立的专用的线程负责数据通道，从而避免控制命令占用。本发明图 1 所示实施例中，上述代理服务器采用可扩展的架构，上述模拟设备中的每个模拟设备对应一个进程，这样，对于同一个宿主机上有多个客户操作系统或者多个模拟设备的情形，可以充分利用宿主系统的多颗 CPU，
20 提高并行性，提高整体的 I/O 吞吐。

因为代理服务器可能需要根据存储协议的升级进行必要的升级，或者因为缺陷可能发生意外的重启。但是无论哪种情况，如果不加处理，那么反映到客户操作系统中的表现就是 I/O 错误。为了提高可靠性，本发明设计实现了容错机制，模拟设备实时感知代理服务器的状态，无论代理服务器是因为升级还是因为意外，在代理服务器重启之后，模拟设备将
25 重传尚未完成的 I/O 请求；也就是说，在上述代理服务器重启之后，模拟设备重传在上述代理服务器重启之前未完成传输的数据。

因为对于后端存储服务器，不同的块可能存储在不同的机器上，因此，为了提高总体的并行性，本发明为客户操作系统和模拟设备设计了一种 I/O 调度算法，除了合并同一个块中的相邻请求外，也可以向存储服务器发送位于不同块的请求，这样后端可以调度到不同
30 存储服务器，提高整体的并发，从而提高整体的 I/O 吞吐。这个优化因为涉及到客户操作系统的改动，是可选的，如果不改动客户操作系统，不会影响本发明图 1 所示实施例提供的数据存储方法其他部分的实现。

上述数据存储方法中，模拟设备接收来自客户操作系统的数据之后，将上述数据在宿主系统的用户空间中按照存储服务器使用的协议进行封装，然后将封装后的数据发送给上述存储服务器进行存储，从而可以实现将数据存储在存储服务器上，缩短了 I/O 路径，提高了系统的 I/O 吞吐，减轻了 CPU 的负载，优化了系统性能。

5 图 3 为本发明数据存储装置一个实施例的结构示意图，本实施例中的数据存储装置实现为虚拟机监控器中的一个模拟设备，可以实现本发明图 1 所示实施例的流程。如图 3 所示，上述数据存储装置可以包括：接收模块 31、封装模块 32 和发送模块 33；

其中，接收模块 31，用于接收来自客户操作系统的数据；

10 封装模块 32，用于将接收模块 31 接收的数据在宿主系统的用户空间中按照存储服务器使用的协议进行封装；

发送模块 33，用于将封装模块 32 封装后的数据发送给上述存储服务器进行存储。

本实施例中，上述模拟设备是在宿主系统上基于虚拟机监控器，例如：Qemu 实现的，上述模拟设备为 PCI 接口的块设备；上述客户操作系统使用块设备驱动来驱动上述模拟设备，上述客户操作系统运行于上述虚拟机监控器之上。

15 参见图 2，本发明基于虚拟机监控器 Qemu，设计并实现了一个模拟设备。为了避免客户操作系统的改动，减少对客户操作系统的影响，该模拟设备被模拟为一个块设备，客户操作系统中无需实现任何对该模拟设备的驱动，直接使用块设备驱动即可；为了更高效的在客户操作系统和宿主系统之间进行通信，模拟设备采用 VirtIO 标准实现数据通信；为了避免存储服务器升级等带来的影响，模拟设备与存储服务器进行了解耦，本发明在宿主系统上设计实现了一个存储服务器的代理服务器。

20 为了使客户操作系统不经修改就可以使用本实施例提供的装置，这一点尤其是对于那些用户自定义的客户操作系统非常有意义，本发明将模拟设备注册为 PCI 接口的 VirtIO 块设备，客户操作系统发现是 PCI 总线上的 VirtIO 块设备后，将使用 virtio-blk 驱动来驱动该模拟设备。

25 本实施例中，上述模拟设备与客户操作系统之间通过 VirtIO 的方式共享内存，模拟设备与客户操作系统之间通过传递上述数据所在内存的物理地址实现上述数据的访问，从而实现了数据的零复制。

上述数据存储装置中，发送模块 33，具体用于将封装模块 32 封装后的数据通过网络通信协议栈发送给上述存储服务器进行存储。本实施例中，接收模块 31 接收到数据之后，
30 不再经过宿主系统的内核 I/O 栈，而是由封装模块 32 在宿主系统的用户空间中按照存储服务器的协议进行封装后，发送模块 33 通过网络通信协议栈直接发送给远端的存储服务器。另外，本实施例对 Qemu 中的 I/O 路径进行了优化，缩短了 I/O 路径。

更具体地，发送模块 33，用于将封装模块 32 封装后的数据发送给上述存储服务器的代理服务器，由代理服务器将上述封装后的数据通过网络通信协议栈发送给存储服务器进行存储。其中，上述网络通信协议栈可以为 TCP/IP 协议栈，当然本发明实施例并不仅限于此，网络通信协议栈还可以为其他的网络通信协议栈，只要可以实现模拟设备与存储服务器之间的网络通信即可。

在生产环境上，频繁升级 Qemu 的代价和风险都是比较大的，因此，本发明在模拟设备和远端存储服务器之间实现了一个存储服务器的代理服务器，将业务逻辑分离到代理服务器中来。而代理服务器与 Qemu 中的模拟设备之间只是进行简单的数据通信，不受具体通信协议细节变更的影响，无论是协议升级，或者是修复缺陷，Qemu 均无需变更。

为了提高 I/O 吞吐，本发明将控制平面和数据通道分开实现，使用一个独立的专用的线程负责数据通道，从而避免控制命令占用。本实施例中，上述代理服务器采用可扩展的架构，上述模拟设备中的每个模拟设备对应一个进程，这样，对于同一个宿主系统上有多个客户操作系统或者多个模拟设备的情形，可以充分利用宿主系统的多颗 CPU，提高并行性，提高整体的 I/O 吞吐。

上述数据存储装置中，发送模块 33，还用于在代理服务器重启之后，重传在代理服务器重启之前未完成传输的数据。因为代理服务器可能需要根据存储协议的升级进行必要的升级，或者因为缺陷可能发生意外的重启。但是无论哪种情况，如果不加处理，那么反映到客户操作系统中的表现就是 I/O 错误。为了提高可靠性，上述数据存储装置设计实现了容错机制，实时感知代理服务器的状态，无论代理服务器是因为升级还是因为意外，在代理服务器重启之后，发送模块 33 将重传尚未完成的 I/O 操作。

上述数据存储装置中，接收模块 31 接收来自客户操作系统的数据之后，封装模块 32 将上述数据在宿主系统的用户空间中按照存储服务器使用的协议进行封装，然后发送模块 33 将封装后的数据发送给上述存储服务器进行存储，从而可以实现将数据存储在存储服务器上，缩短了 I/O 路径，提高了系统的 I/O 吞吐，减轻了 CPU 的负载，优化了系统性能。

本发明实施例还提供一种电子设备，包括：一个或者多个处理器；存储器；一个或者多个程序，所述一个或者多个程序存储在所述存储器中，当被所述一个或者多个处理器执行时进行如下操作：模拟设备接收来自客户操作系统的数据；将所述数据在宿主系统的用户空间中按照存储服务器使用的协议进行封装；将封装后的数据发送给所述存储服务器进行存储。

本发明实施例还提供一种非易失性计算机存储介质，所述计算机存储介质存储有一个或者多个模块，当所述一个或者多个模块被执行时进行如下操作：模拟设备接收来自客户操作系统的数据；将所述数据在宿主系统的用户空间中按照存储服务器使用的协议进

行封装；将封装后的数据发送给所述存储服务器进行存储。

需要说明的是，在本发明的描述中，术语“第一”、“第二”等仅用于描述目的，而不能理解为指示或暗示相对重要性。此外，在本发明的描述中，除非另有说明，“多个”的含义是两个或两个以上。

5 流程图中或在此以其他方式描述的任何过程或方法描述可以被理解为，表示包括一个或更多个用于实现特定逻辑功能或过程的步骤的可执行指令的代码的模块、片段或部分，并且本发明的优选实施方式的范围包括另外的实现，其中可以不按所示出或讨论的顺序，包括根据所涉及的功能按基本同时的方式或按相反的顺序，来执行功能，这应被本发明的实施例所属技术领域的技术人员所理解。

10 应当理解，本发明的各部分可以用硬件、软件、固件或它们的组合来实现。在上述实施方式中，多个步骤或方法可以用存储在存储器中且由合适的指令执行系统执行的软件或固件来实现。例如，如果用硬件来实现，和在另一实施方式中一样，可用本领域公知的下列技术中的任一项或他们的组合来实现：具有用于对数据信号实现逻辑功能的逻辑门电路的离散逻辑电路，具有合适的组合逻辑门电路的专用集成电路，可编程门阵列
15 (Programmable Gate Array；以下简称：PGA)，现场可编程门阵列 (Field Programmable Gate Array；以下简称：FPGA) 等。

本技术领域的普通技术人员可以理解实现上述实施例方法携带的全部或部分步骤是可以通过程序来指令相关的硬件完成，所述的程序可以存储于一种计算机可读存储介质中，该程序在执行时，包括方法实施例的步骤之一或其组合。

20 此外，本发明各个实施例中的各功能模块可以集成在一个处理模块中，也可以是各个模块单独物理存在，也可以两个或两个以上模块集成在一个模块中。上述集成的模块既可以采用硬件的形式实现，也可以采用软件功能模块的形式实现。所述集成的模块如果以软件功能模块的形式实现并作为独立的产品销售或使用时，也可以存储在一个计算机可读取存储介质中。

25 上述提到的存储介质可以是只读存储器，磁盘或光盘等。

在本说明书的描述中，参考术语“一个实施例”、“一些实施例”、“示例”、“具体示例”、或“一些示例”等的描述意指结合该实施例或示例描述的具体特征、结构、材料或者特点包含于本发明的至少一个实施例或示例中。在本说明书中，对上述术语的示意性表述不一定指的是相同的实施例或示例。而且，描述的具体特征、结构、材料或者特点可以在任何30 的一个或多个实施例或示例中以合适的方式结合。

尽管上面已经示出和描述了本发明的实施例，可以理解的是，上述实施例是示例性的，不能理解为对本发明的限制，本领域的普通技术人员在本发明的范围内可以对上述实施例

进行变化、修改、替换和变型。

权利要求书

1、一种数据存储方法，其特征在于，包括：

模拟设备接收来自客户操作系统的数据；

5 将所述数据在宿主系统的用户空间中按照存储服务器使用的协议进行封装；

将封装后的数据发送给所述存储服务器进行存储。

2、根据权利要求 1 所述的方法，其特征在于，所述模拟设备是在所述宿主系统上基于虚拟机监控器实现的，所述模拟设备为外设部件互连标准 PCI 接口的块设备；所述客户操作系统使用块设备驱动来驱动所述模拟设备，所述客户操作系统运行于所述虚拟机监控器之上。

3、根据权利要求 2 所述的方法，其特征在于，所述模拟设备与所述客户操作系统之间通过共享内存的方式传递数据；所述模拟设备与所述客户操作系统之间通过传递所述数据所在内存的物理地址实现所述数据的访问。

4、根据权利要求 1-3 任意一项所述的方法，其特征在于，所述将封装后的数据发送给所述存储服务器进行存储包括：

所述模拟设备将所述封装后的数据通过网络通信协议栈发送给所述存储服务器进行存储。

5、根据权利要求 4 所述的方法，其特征在于，所述模拟设备将所述封装后的数据通过网络通信协议栈发送给所述存储服务器进行存储包括：

20 所述模拟设备将所述封装后的数据发送给所述存储服务器的代理服务器，由所述代理服务器将所述封装后的数据通过网络通信协议栈发送给所述存储服务器进行存储。

6、根据权利要求 5 所述的方法，其特征在于，所述代理服务器采用可扩展的架构，每个模拟设备对应一个进程。

7、根据权利要求 5 或 6 所述的方法，其特征在于，还包括：

25 在所述代理服务器重启之后，所述模拟设备重传在所述代理服务器重启之前未完成传输的数据。

8、一种数据存储装置，其特征在于，设置在模拟设备中，所述数据存储装置包括：

接收模块，用于接收来自客户操作系统的数据；

30 封装模块，用于将所述接收模块接收的数据在宿主系统的用户空间中按照存储服务器使用的协议进行封装；

发送模块，用于将所述封装模块封装后的数据发送给所述存储服务器进行存储。

9、根据权利要求 8 所述的装置，其特征在于，所述模拟设备是在所述宿主系统上基于

虚拟机监控器实现的，所述模拟设备为外设部件互连标准 PCI 接口的块设备；所述客户操作系统使用块设备驱动来驱动所述模拟设备，所述客户操作系统运行于所述虚拟机监控器之上。

10、根据权利要求 9 所述的装置，其特征在于，所述模拟设备与所述客户操作系统之

5 间通过共享内存的方式传递数据；所述模拟设备与所述客户操作系统之间通过传递所述数据所在内存的物理地址实现所述数据的访问。

11、根据权利要求 8-10 任意一项所述的装置，其特征在于，

所述发送模块，具体用于将所述封装模块封装后的数据通过网络通信协议栈发送给所述存储服务器进行存储。

10 12、根据权利要求 11 所述的装置，其特征在于，

所述发送模块，具体用于将所述封装模块封装后的数据发送给所述存储服务器的代理服务器，由所述代理服务器将所述封装后的数据通过网络通信协议栈发送给所述存储服务器进行存储。

13、根据权利要求 12 所述的装置，其特征在于，所述代理服务器采用可扩展的架构，

15 每个模拟设备对应一个进程。

14、根据权利要求 12 或 13 所述的装置，其特征在于，

所述发送模块，还用于在所述代理服务器重启之后，重传在所述代理服务器重启之前未完成传输的数据。

15、一种电子设备，其特征在于，包括：

20 一个或者多个处理器；

存储器；

一个或者多个程序，所述一个或者多个程序存储在所述存储器中，当被所述一个或者多个处理器执行时：

执行如权利要求 1-7 任一项所述的方法。

25 16、一种非易失性计算机存储介质，其特征在于，所述计算机存储介质存储有一个或者多个模块，当所述一个或者多个模块被执行时：

执行如权利要求 1-7 任一项所述的方法。

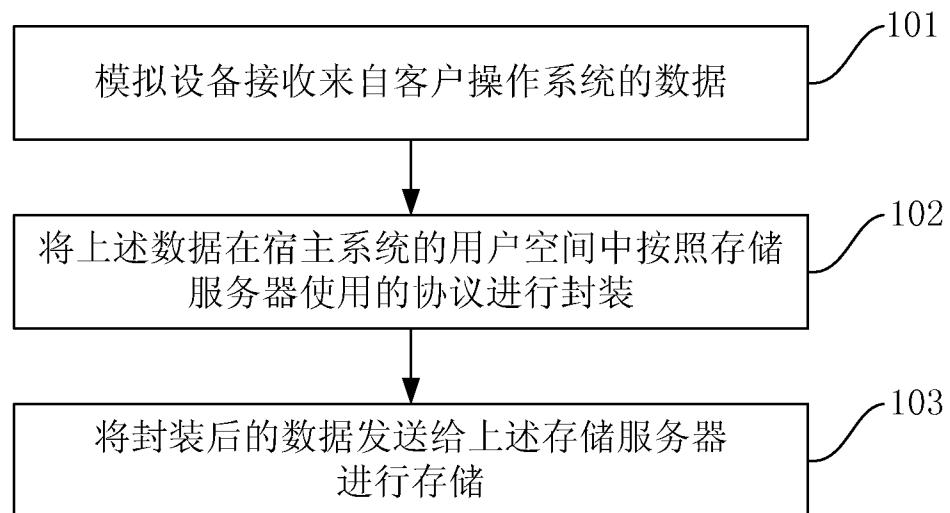


图 1

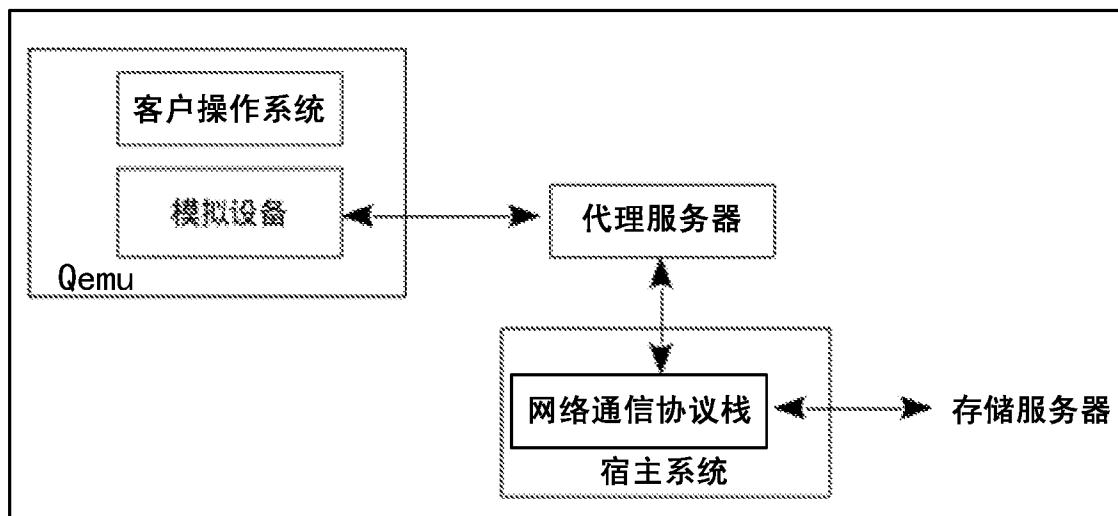


图 2

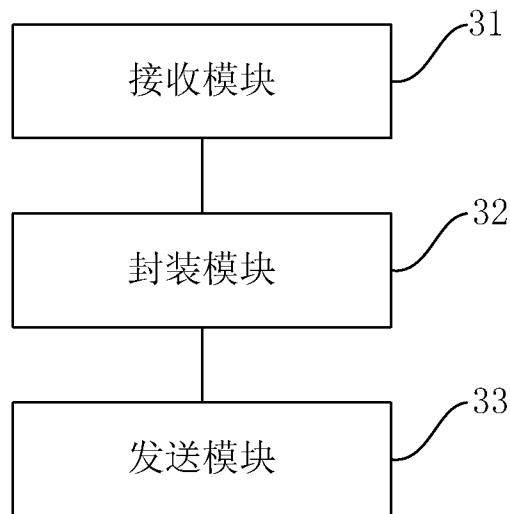


图 3

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2015/091773

A. CLASSIFICATION OF SUBJECT MATTER

G06F 3/06 (2006.01) i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G06F

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

EPODOC, WPI, CNPAT, GOOGLE: drive, storage, host, user space, protocol, package, driver, block device, address, proxy

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 104881254 A (BEIJING BAIDU NETWORK TECHNOLOGY CO., LTD.), 02 September 2015 (02.09.2015), claims 1-14, and description, paragraphs [0006]-[0012], [0018]-[0032] and [0050]-[0052]	1-16
X	CN 104636076 A (CHINA TELECOM CORPORATION LIMITED), 20 May 2015 (20.05.2015), description, paragraphs [0002], [0006]-[0015], [0022]-[0027], [0034]-[0036], [0058]-[0067], [0075] and [0079]	1-16
A	CN 104092759 A (STANDARD CLOUD TECHNOLOGY CO., LTD.), 08 October 2014 (08.10.2014), the whole document	1-16
A	US 2011307886 A1 (ORACLE INTERNATIONAL CORPORATION), 15 December 2011 (15.12.2011), the whole document	1-16

Further documents are listed in the continuation of Box C.

See patent family annex.

* Special categories of cited documents:

- “A” document defining the general state of the art which is not considered to be of particular relevance
- “E” earlier application or patent but published on or after the international filing date
- “L” document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- “O” document referring to an oral disclosure, use, exhibition or other means
- “P” document published prior to the international filing date but later than the priority date claimed

“T” later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

“X” document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

“Y” document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

“&” document member of the same patent family

Date of the actual completion of the international search
29 January 2016 (29.01.2015)

Date of mailing of the international search report
24 February 2016 (24.02.2015)

Name and mailing address of the ISA/CN:
State Intellectual Property Office of the P. R. China
No. 6, Xitucheng Road, Jimenqiao
Haidian District, Beijing 100088, China
Facsimile No.: (86-10) 62019451

Authorized officer

MA, Lili

Telephone No.: (86-10) **62413996**

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No.

PCT/CN2015/091773

Patent Documents referred in the Report	Publication Date	Patent Family	Publication Date
CN 104881254 A	02 September 2015	None	
CN 104636076 A	20 May 2015	None	
CN 104092759 A	08 October 2014	None	
US 2011307886 A1	15 December 2011	US 9047136 B2	02 June 2015

国际检索报告

国际申请号

PCT/CN2015/091773

A. 主题的分类

G06F 3/06 (2006. 01) i

按照国际专利分类(IPC)或者同时按照国家分类和IPC两种分类

B. 检索领域

检索的最低限度文献(标明分类系统和分类号)

G06F

包含在检索领域中的除最低限度文献以外的检索文献

在国际检索时查阅的电子数据库(数据库的名称, 和使用的检索词(如使用))

EPODOC, WPI, CNPAT, GOOGLE; 存储, 宿主, 用户空间, 协议, 封装, 驱动, 块设备, 地址, 代理, storage, host, user space, protocol, package, driver, block device, address, proxy

C. 相关文件

类型*	引用文件, 必要时, 指明相关段落	相关的权利要求
PX	CN 104881254 A (北京百度网讯科技有限公司) 2015年 9月 2日 (2015 - 09 - 02) 权利要求1-14, 说明书第[0006]-[0012]、[0018]-[0032]、[0050]-[0052]段	1-16
X	CN 104636076 A (中国电信股份有限公司) 2015年 5月 20日 (2015 - 05 - 20) 说明书第[0002]、[0006]-[0015]、[0022]-[0027]、[0034]-[0036]、[0058]-[0067]、[0075]、[0079]段	1-16
A	CN 104092759 A (南京斯坦德云科技股份有限公司) 2014年 10月 8日 (2014 - 10 - 08) 全文	1-16
A	US 2011307886 A1 (ORACLE INTERNATIONAL CORPORATION) 2011年 12月 15日 (2011 - 12 - 15) 全文	1-16

 其余文件在C栏的续页中列出。 见同族专利附件。

* 引用文件的具体类型:

“A” 认为不特别相关的表示了现有技术一般状态的文件

“T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件

“E” 在国际申请日的当天或之后公布的在先申请或专利

“X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性

“L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件(如具体说明的)

“Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性

“0” 涉及口头公开、使用、展览或其他方式公开的文件

“&” 同族专利的文件

“P” 公布日先于国际申请日但迟于所要求的优先权日的文件

国际检索实际完成的日期

2016年 1月 29日

国际检索报告邮寄日期

2016年 2月 24日

ISA/CN的名称和邮寄地址

中华人民共和国国家知识产权局(ISA/CN)
中国北京市海淀区蓟门桥西土城路6号 100088

受权官员

马丽莉

传真号 (86-10) 62019451

电话号码 (86-10) 62413996

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2015/091773

检索报告引用的专利文件		公布日 (年/月/日)		同族专利		公布日 (年/月/日)	
CN	104881254	A	2015年 9月 2日	无			
CN	104636076	A	2015年 5月 20日	无			
CN	104092759	A	2014年 10月 8日	无			
US	2011307886	A1	2011年 12月 15日	US	9047136 B2	2015年 6月 2日	

表 PCT/ISA/210 (同族专利附件) (2009年7月)