



(12) 发明专利

(10) 授权公告号 CN 101409673 B

(45) 授权公告日 2013.07.03

(21) 申请号 200810177727.1

(22) 申请日 2008.11.12

(73) 专利权人 北京恒光创新科技股份有限公司
地址 100097 北京市海淀区蓝靛厂东路2号
院金源时代商务中心2号楼A座3C

专利权人 北京恒光通信技术有限公司
北京恒光科技发展有限公司

(72) 发明人 周志雄 汪锐 赵彦博

(74) 专利代理机构 北京三友知识产权代理有限公司 11127

代理人 任默闻

(51) Int. Cl.

H04L 29/06 (2006.01)

H04L 12/02 (2006.01)

权利要求书4页 说明书10页 附图6页

(54) 发明名称

一种网络适配器数据传输方法、网络适配器及系统

(57) 摘要

本发明提供一种网络适配器数据传输方法、网络适配器及系统，该方法包括通过物理接口接收网络数据包；根据物理接口与硬件接收单元的对应关系，将网络数据包分配给对应的硬件接收单元；由硬件接收单元将网络数据包发送给操作系统内的与网络适配器对应的驱动程序，以由驱动程序对网络数据包进行处理。本发明有益效果在于，使物理接口接收的网络数据按照设定条件分配到相应的硬件接收单元，将网络适配器单个物理接口的流量均衡到不同的硬件接收单元及该硬件接收单元对应的网络外设上，提高网络适配器的每个物理接口数据流量的处理能力，也可将多个物理接口的流量统一到单个硬件接收单元及对应的网络外设上，提高对多个物理接口流量的处理效率。

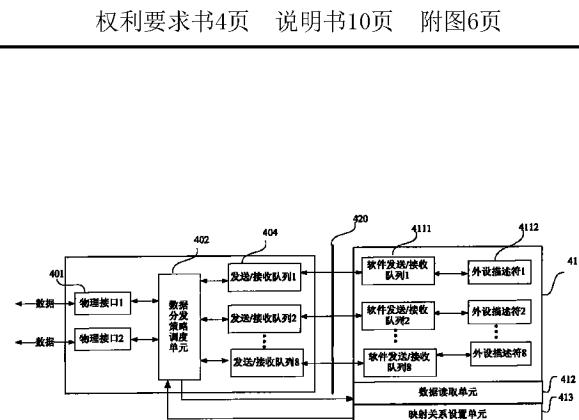
(56) 对比文件

CN 1516840 A, 2004.07.28,

CN 1679281 A, 2005.10.05,

WO 2008098249 A1, 2008.08.14,

审查员 林甡



1. 一种网络适配器数据接收方法,其特征在于,所述的方法包括:

通过物理接口接收网络数据包,所述物理接口为多个;

根据硬件接收单元选择条件,将所述物理接口接收到的网络数据包聚合到多个硬件接收单元中;

由所述多个硬件接收单元将所述网络数据包发送给操作系统内的与所述网络适配器对应的驱动程序,以由所述的驱动程序对所述的网络数据包进行处理;

其中,每一所述硬件接收单元对应绑定在所述操作系统内的一个CPU核上;

当网络数据流量小时,将所述多个物理接口接收到的网络数据包聚合到一个CPU核上集中处理;当网络数据流量大时,将所述多个物理接口接收到的网络数据包均衡到多个CPU核上进行处理。

2. 根据权利要求1所述的方法,其特征在于,根据网络数据包信息和/或操作系统资源信息对所述的硬件接收单元选择条件进行设置。

3. 根据权利要求2所述的方法,其特征在于,所述的网络数据包信息包括:

网络数据包IP地址,网络数据的端口地址,网络数据包数据类型,网络数据包数据容量,网络数据包的发送顺序,网络数据包的到达时间,网络数据包复制数量;

所述的操作系统资源信息包括:CPU负载信息和内存使用信息。

4. 根据权利要求1所述的方法,其特征在于,所述的硬件接收单元选择条件包括:物理接口与硬件接收单元的对应关系;其中

根据所述的网络数据包信息和/或操作系统资源信息对所述的物理接口与硬件接收单元的对应关系进行设置。

5. 根据权利要求4所述的方法,其特征在于,所述的物理接口与硬件接收单元的对应关系包括:

网络数据包IP地址与硬件接收单元标识的对应关系;和/或

网络数据的端口地址与硬件接收单元标识的对应关系;和/或

网络数据包数据类型与硬件接收单元标识的对应关系;和/或

网络数据包数据容量与硬件接收单元标识的对应关系;和/或

硬件接收单元的空满情况与硬件接收单元标识的对应关系;和/或

网络数据包的发送顺序与硬件接收单元标识的对应关系;和/或

网络数据包的到达时间与硬件接收单元标识的对应关系;和/或

网络数据包复制数量与硬件接收单元标识的对应关系;和/或

操作系统资源信息与硬件接收单元标识的对应关系。

6. 一种网络适配器数据发送方法,其特征在于,所述的方法包括:

通过硬件发送单元接收操作系统发送的系统数据包,所述硬件发送单元为多个;

根据物理接口选择条件,将所述硬件发送单元接收到的系统数据包分配给对应的物理接口,所述物理接口为多个;

由所述多个物理接口将所述系统数据包发送给对应的网络设备;

其中,每一所述硬件发送单元对应绑定在所述操作系统内的一个CPU核上;

当网络数据流量小时,将所述多个硬件发送单元接收到的网络数据包聚合到一个CPU核上集中处理;当网络数据流量大时,将所述多个硬件发送单元接收到的网络数据包均衡

到多个 CPU 核上进行处理。

7. 根据权利要求 6 所述的方法, 其特征在于, 根据系统数据包信息和 / 或操作系统资源信息对所述的物理接口选择条件进行设置。

8. 根据权利要求 7 所述的方法, 其特征在于, 所述的物理接口选择条件包括 : 物理接口与硬件发送单元的对应关系 ; 其中

根据所述的系统数据包信息和 / 或操作系统资源信息对所述的物理接口与硬件发送单元的对应关系进行设置。

9. 根据权利要求 6 所述的方法, 其特征在于, 所述的系统数据包信息包括 :

系统数据包 IP 地址, 系统数据包的端口地址, 系统数据包数据类型, 系统数据包数据容量, 系统数据包的发送顺序, 系统数据包的发送时间, 系统数据包复制数量 ;

所述的操作系统资源信息包括 :CPU 负载信息和内存使用信息。

10. 根据权利要求 9 所述的方法, 其特征在于, 所述的物理接口与硬件接收单元的对应关系包括 :

系统数据包 IP 地址与物理接口标识的对应关系 ; 和 / 或

系统数据包的端口地址与物理接口标识的对应关系 ; 和 / 或

系统数据包数据类型与物理接口标识的对应关系 ; 和 / 或

系统数据包数据容量与物理接口标识的对应关系 ; 和 / 或

硬件发送单元的空满情况与物理接口标识的对应关系 ; 和 / 或

系统数据包的发送顺序与物理接口标识的对应关系 ; 和 / 或

系统数据包复制数量与物理接口标识的对应关系 ; 和 / 或

操作系统资源信息与物理接口标识的对应关系。

11. 一种网络适配器, 其特征在于, 所述的网络适配器包括 :

物理接口, 用于接收网络数据包, 所述物理接口为多个 ;

映射关系存储单元, 用于存储物理接口与硬件接收单元的映射关系 ;

网络数据分配单元, 用于根据物理接口与硬件接收单元的对应关系, 将所述物理接口接收到的网络数据包聚合到多个硬件接收单元中 ;

硬件接收单元, 用于将所述网络数据包发送给操作系统内的与所述网络适配器对应的驱动程序, 以由所述的驱动程序对所述的网络数据包进行处理 ;

其中, 每一所述硬件接收单元对应绑定在所述操作系统内的一个 CPU 核上 ;

当网络数据流量小时, 将所述多个物理接口接收到的网络数据包聚合到一个 CPU 核上集中处理 ; 当网络数据流量大时, 将所述多个物理接口接收到的网络数据包均衡到多个 CPU 核上进行处理。

12. 一种网络适配器, 其特征在于, 所述的网络适配器包括 :

硬件发送单元, 用于接收操作系统发送的系统数据包, 所述硬件发送单元为多个 ;

映射关系存储单元, 用于存储物理接口与硬件发送单元的映射关系,

系统数据分配单元, 用于将所述硬件发送单元接受到的系统数据包分配给对应的物理接口 ;

物理接口, 所述物理接口为多个, 用于将所述系统数据包发送给对应的网络设备 ;

其中, 每一所述硬件发送单元对应绑定在所述操作系统内的一个 CPU 核上 ;

当网络数据流量小时,将所述多个硬件发送单元接收到的网络数据包聚合到一个CPU核上集中处理;当网络数据流量大时,将所述多个硬件发送单元接收到的网络数据包均衡到多个CPU核上进行处理。

13. 一种网络适配器,其特征在于,所述的网络适配器包括:

物理接口,用于接收网络数据包或发送系统数据包;

映射关系存储单元,用于存储物理接口与硬件接收单元和硬件发送单元的映射关系;

网络数据分配单元,用于根据物理接口与硬件接收单元的对应关系将所述网络数据包分配给对应的硬件接收单元;

硬件接收单元,用于将所述网络数据包发送给操作系统内的与所述网络适配器对应的驱动程序,以由所述的驱动程序对所述的网络数据包进行处理;

硬件发送单元,用于接收操作系统发送的系统数据包;

系统数据分配单元,用于将所述系统数据包分配给对应的物理接口;

其中,所述物理接口为多个,所述硬件接收单元和所述硬件发送单元一一对应,为多个;

其中,每一所述硬件发送单元与硬件接收单元对应绑定在所述操作系统内的一个CPU核上;

当网络数据流量小时,将所述多个硬件发送单元或硬件接收单元接收到的网络数据包聚合到一个CPU核上集中处理;当网络数据流量大时,将所述多个硬件发送单元或硬件接收单元接收到的网络数据包均衡到多个CPU核上进行处理。

14. 一种网络适配系统,其特征在于,所述的系统包括:网络适配器和系统主机;

所述的网络适配器包括:物理接口,用于接收网络数据包或发送系统数据包;映射关系存储单元,用于存储物理接口与硬件接收单元的映射关系;网络数据分配单元,用于根据物理接口与硬件接收单元和硬件发送单元的映射关系将所述网络数据包分配给对应的硬件接收单元;硬件接收单元,用于将所述网络数据包发送给操作系统内的与所述网络适配器对应的驱动程序,以由所述的驱动程序对所述的网络数据包进行处理;硬件发送单元,用于接收系统主机发送的系统数据包;系统数据分配单元,用于将所述系统数据包分配给对应的物理接口;

其中,所述物理接口为多个,所述硬件接收单元和所述硬件发送单元一一对应,为多个;

其中,每一所述硬件发送单元与硬件接收单元对应绑定在所述操作系统内的一个CPU核上;

当网络数据流量小时,将所述多个硬件发送单元或硬件接收单元接收到的网络数据包聚合到一个CPU核上集中处理;当网络数据流量大时,将所述多个硬件发送单元或硬件接收单元接收到的网络数据包均衡到多个CPU核上进行处理;

所述的系统主机包括:驱动单元,用于调用驱动程序对所述的网络适配器进行驱动,并由所述驱动程序的外设描述符将所述的网络适配器描述为网络外设;数据读取单元,用于对所述的物理接口与硬件接收单元的映射关系进行读取,供操作系统使用;

映射关系设置单元,用于对所述的物理接口与硬件接收单元的映射关系进行设置。

15. 根据权利要求14所述的系统,其特征在于,所述的网络适配器包括多个物理接口,

所述的驱动程序的外设描述符将所述的网络适配器描述为一个网络外设。

16. 根据权利要求 14 所述的系统，其特征在于，所述的网络适配器包括单个物理接口，所述的驱动程序的外设描述符将所述的网络适配器描述为多个网络外设。

17. 根据权利要求 14 所述的系统，其特征在于，所述的网络适配器包括多物理接口，所述的驱动程序的外设描述符将所述的网络适配器描述为多个网络外设。

18. 根据权利要求 15 或 16 或 17 所述的系统，其特征在于，所述的物理接口的数目与外设描述符的数目不相等。

19. 根据权利要求 17 所述的系统，其特征在于，所述的物理接口的数目与外设描述符的数目相等。

一种网络适配器数据传输方法、网络适配器及系统

技术领域

[0001] 本发明关于计算机网络技术,特别关于计算机网络中的数据传输技术,具体的讲本发明是一种网络适配器数据传输方法、网络适配器及系统。

背景技术

[0002] 网络适配器 (Network Adapter),又称网络接口卡 (Network Interface Card, NIC),俗称网卡,是主机与网络连接的接口设备。

[0003] 从网络的OSI (Open System Interconnection) 七层模型来看,网络适配器实现了物理层与数据链路层。网络适配器中对物理层与数据链路层的一个实现称为一个物理接口。每一个物理接口在硬件中实现为专门的接收 / 发送单元,其与主机操作系统的网络适配器驱动程序配合工作,以单元方式接收或发送数据报文。每一个物理接口在主机操作系统中体现为一个网络外设。

[0004] 网络适配器根据物理接口的数目分为单口网络适配器和多口网络适配器两类:

[0005] 一、单口网络适配器,设有一个物理接口以及一组硬件接收 / 发送单元。

[0006] 二、多口网络适配器,设有两个以上的物理接口和两组以上的硬件接收 / 发送单元,物理接口与硬件接收 / 发送单元的数目相同,且每个物理接口与一组硬件接收 / 发送单元对应。

[0007] 如图 1 所示,为多口网络适配器与主机操作系统的数据传输,多口适配器 100 与主机操作系统 200 之间通过总线连接,主机操作系统 200 为多口适配器 100 的每组硬件接收 / 发送单元提供一个驱动程序,即软件接收 / 发送单元 1 至 M,使得每个物理接口与主机操作系统 200 中一个网络外设对应。

[0008] 如图 1 所示,当多口网络适配器 100 的物理接口 1 收到网络设备发送网络数据包时,物理接口 1 通过硬件接收 / 发送单元 1 和软件接收 / 发送单元 1,将网络数据包发送给映射的网络外设 1;当主机操作系统 200 需要将网络外设 1 的数据发送给网络设备时,主机操作系统 200 的网络外设 1 通过软件接收 / 发送单元 1,将系统数据包发送给硬件接收 / 发送单元 1,再由硬件接收 / 发送单元 1 将系统数据包通过物理接口 1 发送至网络设备。

[0009] 多口网络适配器缺点在于,物理接口与对应的网络外设之间,只能通过一组硬件接收 / 发送单元进行数据交互,无法提高每个物理接口的数据流量的处理能力。

发明内容

[0010] 为解决现有技术的问题,本发明提供一种网络适配器数据传输方法、网络适配器及网络适配器系统,用于使每个物理接口网络数据包分配给不同的硬件接收单元,提高网络适配器的每个物理接口数据流量的处理能力。

[0011] 本发明的目的之一是提供一种网络适配器数据传输方法,该方法包括:通过物理接口接收网络数据包;根据硬件接收单元选择条件将网络数据包分配给对应的硬件接收单元;由硬件接收单元将网络数据包发送给操作系统内的与网络适配器对应的驱动程序,以

由驱动程序对网络数据包进行处理。

[0012] 本发明的目的之一是提供了一种网络适配器数据发送方法,该方法包括:通过硬件发送单元接收操作系统发送的系统数据;根据物理接口选择条件将系统数据分配给对应的物理接口;由物理接口将系统数据发送给对应的网络设备。

[0013] 本发明的目的之一是提供一种网络适配器,该网络适配器包括:物理接口,用于接收网络数据包;映射关系存储单元,用于存储物理接口与硬件接收单元的映射关系;网络数据分配单元,用于根据物理接口与硬件接收单元的对应关系将网络数据包分配给对应的硬件接收单元;硬件接收单元,用于将网络数据包发送给操作系统内的与所述网络适配器对应的驱动程序,以由所述的驱动程序对所述的网络数据包进行处理。

[0014] 本发明的目的之一是提供一种网络适配器,该网络适配器包括:硬件发送单元,用于接收操作系统发送的系统数据包;映射关系存储单元,用于存储物理接口与硬件发送单元的对应关系,系统数据分配单元,用于将系统数据包分配给对应的物理接口;物理接口,用于将系统数据包发送给对应的网络设备。

[0015] 本发明的目的之一是提供一种网络适配器,该网络适配器包括:物理接口,用于接收网络数据包或发送系统数据包;映射关系存储单元,用于存储物理接口与硬件接收单元和硬件发送单元的映射关系;网络数据分配单元,用于根据物理接口与硬件接收单元的对应关系将网络数据包分配给对应的硬件接收单元;硬件接收单元,用于将网络数据包发送给操作系统内的与所述网络适配器对应的驱动程序,以由所述的驱动程序对所述的网络数据包进行处理;硬件发送单元,用于接收操作系统发送的系统数据包;系统数据分配单元,用于将系统数据包分配给对应的物理接口。

[0016] 本发明的目的之一是提供一种网络适配系统,该方法系统包括:网络适配器和系统主机;网络适配器包括:物理接口,用于接收网络数据包或发送系统数据包;映射关系存储单元,用于存储物理接口与硬件接收单元和硬件发送单元的映射关系;网络数据分配单元,用于根据物理接口与硬件接收单元的对应关系将网络数据包分配给对应的硬件接收单元;硬件接收单元,用于将网络数据包发送给操作系统内的与所述网络适配器对应的驱动程序,以由所述的驱动程序对所述的网络数据包进行处理;硬件发送单元,用于接收系统主机发送的系统数据包;系统数据分配单元,用于将系统数据包分配给对应的物理接口;

[0017] 系统主机包括:驱动单元,用于调用驱动程序对网络适配器进行驱动,并由驱动程序的外设描述符将网络适配器描述为网络外设;数据读取单元,用于对物理接口与硬件接收单元的映射关系进行读取,供操作系统使用;映射关系设置单元,用于对物理接口与硬件接收单元的映射关系进行设置。

[0018] 本发明实施例的有益效果在于,使物理接口上的网络数据按照设定条件分配到相应的硬件接收单元,将网络适配器单个物理接口的流量均衡到多个硬件接收单元及对应的网络外设上,提高网络适配器每个物理接口数据流量的处理能力,也可将多个物理接口的流量统一到单个硬件接收单元及对应的网络外设上,提高对多个物理接口流量的处理效率。同时,网络适配器直接表现为一个或多个网络外设,可直接与操作系统中的协议栈进行数据的交换,使网络适配器的使用仍具有通用性。

附图说明

[0019] 此处所说明的附图用来提供对本发明的进一步理解,构成本申请的一部分,并不构成对本发明的限定。在附图中:

- [0020] 图 1 是多口网络适配器与主机操作系统的数据传输的示意图;
- [0021] 图 2A 是本发明实施例 1 的聚合网络适配器的功能结构示意图;
- [0022] 图 2B 是本发明实施例 1 的聚合网络适配器接收网络数据包流程图;
- [0023] 图 3A 是本发明实施例 2 的聚合网络适配器的功能结构示意图;
- [0024] 图 3B 是本发明实施例 2 的聚合网络适配器发送系统数据包流程图;
- [0025] 图 4A 是本发明实施例 3 的聚合网络适配器的功能结构示意图;
- [0026] 图 4B 是本发明实施例 3 的主机系统的功能结构示意图;
- [0027] 图 5 是本发明实施例 3 的聚合网络适配器与主机操作系统的数据传输示意图;
- [0028] 图 6 是本发明实施例 3 的物理接口与硬件接收 / 发送队列的映射关系示意图;
- [0029] 图 7 是本发明实施例 3 的硬件接收 / 发送队列的数据分配策略;
- [0030] 图 8 是本发明实施例 3 的物理接口的数据分发策略。
- [0031] 图 9 是本发明实施例 4 的硬件接收 / 发送队列的数据分配策略;
- [0032] 图 10 是本发明实施例 4 的物理接口的数据分发策略。

具体实施方式

[0033] 为使本发明的目的、技术方案和优点更加清楚明白,下面结合实施方式和附图,对本发明做进一步详细说明。在此,本发明的示意性实施方式及其说明用于解释本发明,但并不作为对本发明的限定。

实施例 1

[0035] 如图 2A 所示,为本发明实施例 1 的网络适配器功能结构,其中网络适配器 200 包括:物理接口 201,用于接收网络数据包;映射关系存储单元 202,用于存储物理接口与硬件接收单元的映射关系;网络数据分配单元 203,用于根据硬件接收单元选择条件及物理接口与硬件接收单元的对应关系将网络数据包分配给对应的硬件接收单元;硬件接收单元 204,用于将网络数据包发送给操作系统内的与所述网络适配器对应的驱动程序,以由所述的驱动程序对所述的网络数据包进行处理。硬件接收单元 204 可以通过多种方式实现,例如硬件接收队列。

[0036] 如图 2B 所示,为本发明实施方式的网络适配器接收网络数据的流程;其中包括:步骤 S211,设置硬件接收单元选择条件;

[0037] 步骤 S212,通过网络接口从网络设备接收网络数据;

[0038] 步骤 S213,根据硬件接收单元选择条件,将物理接口接收的网络数据分配给满足条件的硬件接收单元进行发送。

[0039] 在本实施例中,可以根据网络数据包信息,例如:网络数据包 IP 地址、网络数据的端口地址、网络数据包数据类型、网络数据包数据容量(网络数据包的字节数目)、网络数据包的发送顺序、网络数据包的到达时间等,设置硬件接收单元选择条件;还可以根据网络适配器硬件资源情况,例如硬件接收单元空满情况,设置硬件接收单元选择条件;以及,根据操作系统资源信息,例如:CPU 负载信息和内存使用信息设置硬件接收单元选择条件。

[0040] 本实施例中,硬件接收单元选择条件可以是网络数据包 IP 地址与硬件接收单元

标识的对应关系。网络数据分配单元根据网络数据包 IP 地址和硬件接收单元选择条件,选择硬件接收单元,将网络数据包分配给对应的硬件接收单元。

[0041] 本实施例中,硬件接收单元选择条件可以是网络数据的端口地址与硬件接收单元标识的对应关系;网络数据分配单元根据网络数据包端口地址和硬件接收单元选择条件,将网络数据包分配给对应的硬件接收单元。

[0042] 本实施例中,硬件接收单元选择条件可以是网络数据包字节数目(数据容量)与硬件接收单元标识的对应关系;网络数据分配单元根据网络数据包字节数目和硬件接收单元选择条件,选择硬件接收单元,将网络数据包分配给对应的硬件接收单元。

[0043] 本实施例中,硬件接收单元选择条件可以是网络数据包数据类型与硬件接收单元标识的对应关系;网络数据分配单元根据网络数据包数据类型和硬件接收单元选择条件,选择硬件接收单元,将网络数据包分配给对应的硬件接收单元。

[0044] 本实施例中,硬件接收单元选择条件可以是硬件接收单元的空满情况与硬件接收单元标识的对应关系;网络数据分配单元根据硬件接收单元的空满情况和硬件接收单元选择条件,选择硬件接收单元,将网络数据包分配给对应的硬件接收单元。

[0045] 本实施例中,硬件接收单元选择条件可以是网络数据包的发送顺序与硬件接收单元标识的对应关系;网络数据分配单元根据网络数据包的发送顺序和硬件接收单元选择条件,选择硬件接收单元,将网络数据包分配给对应的硬件接收单元。

[0046] 本实施例中,硬件接收单元选择条件可以是网络数据包的到达时间与硬件接收单元标识的对应关系;网络数据分配单元根据网络数据包的到达时间和硬件接收单元选择条件,选择硬件接收单元,将网络数据包分配给对应的硬件接收单元。

[0047] 本实施例中,硬件接收单元选择条件可以是网络数据包复制数量与硬件接收单元标识的对应关系;网络数据分配单元根据硬件接收单元选择条件,将网络数据包复制为多份,将复制的网络数据包分配给相应的硬件接收单元。

[0048] 实施例 2

[0049] 如图 3A 所示,为本发明实施例 2 的网络适配器功能结构,其中网络适配器 300 包括:硬件发送单元 301,用于接收操作系统发送的系统数据包;映射关系存储单元 302,用于存储物理接口与硬件发送单元的对应关系,系统数据分配单元 303,用于根据存储物理接口与硬件发送单元的对应关系,将系统数据包分配给对应的物理接口;物理接口 304,用于将系统数据包发送给对应的网络设备。硬件发送单元 301 可以通过多种方式实现,例如硬件发送队列。

[0050] 如图 3B 所示,为实施例 2 的网络适配器发送系统数据包的流程:

[0051] 步骤 S311,设置物理接口选择条件;

[0052] 步骤 S312,通过硬件发送单元接收系统数据包;

[0053] 步骤 S313,根据物理接口选择条件将硬件发送单元接收的系统数据包分配给满足条件的物理接口进行发送。

[0054] 在本实施例中,可以根据系统数据包信息,例如:系统数据包 IP 地址、网络数据的端口地址、系统数据包数据类型、系统数据包数据容量(系统数据包的字节数目)、系统数据包的发送顺序、系统数据包的发送时间等,设置物理接口选择条件;还可以根据网络适配器硬件资源情况,例如硬件发送单元空满情况,设置物理接口选择条件;以及,根据操作系

统资源信息,例如:CPU 负载信息和内存使用信息设置物理接口选择条件。

[0055] 本实施例中,物理接口选择条件可以是系统数据包 IP 地址与物理接口标识的对应关系。网络数据分配单元根据系统数据包 IP 地址和物理接口选择条件,选择物理接口,将系统数据包分配给对应的物理接口。

[0056] 本实施例中,物理接口选择条件可以是网络数据的端口地址与物理接口标识的对应关系;网络数据分配单元根据系统数据包端口地址和物理接口选择条件,将系统数据包分配给对应的物理接口。

[0057] 本实施例中,物理接口选择条件可以是系统数据包字节数目(数据容量)与物理接口标识的对应关系;网络数据分配单元根据系统数据包字节数目和物理接口选择条件,选择硬件接收单元,将系统数据包分配给对应的物理接口。

[0058] 本实施例中,物理接口选择条件可以是系统数据包数据类型与物理接口标识的对应关系;网络数据分配单元根据系统数据包数据类型和物理接口选择条件,选择物理接口,将系统数据包分配给对应的物理接口。

[0059] 本实施例中,物理接口选择条件可以是硬件发送单元的空满情况与物理接口标识的对应关系;网络数据分配单元根据硬件发送单元的空满情况和物理接口选择条件,选择物理接口,将系统数据包分配给对应的物理接口。

[0060] 本实施例中,物理接口选择条件可以是系统数据包的发送顺序与物理接口标识的对应关系;网络数据分配单元根据系统数据包的发送顺序和物理接口选择条件,选择物理接口,将系统数据包分配给对应的物理接口。

[0061] 本实施例中,物理接口选择条件可以是系统数据包的发送时间与物理接口标识的对应关系;网络数据分配单元根据系统数据包的发送时间和物理接口选择条件,选择物理接口,将系统数据包分配给对应的物理接口。

[0062] 本实施例中,物理接口选择条件可以是系统数据包复制数量与物理接口标识的对应关系;网络数据分配单元根据物理接口选择条件,将系统数据包复制为多份,将复制的系统数据包分配给相应的物理接口。

[0063] 实施例 3

[0064] 如图 4A 所示,为本发明实施例的聚合网络适配器 400 的功能结构,其中聚合网络适配器包括:两个物理接口 401,用于接收网络设备发送的网络数据包以及将系统数据包发送给网络设备;八个硬件接收 / 发送队列 402,用于将网络数据包发送给操作系统内的与所述网络适配器对应的驱动程序,以由所述的驱动程序对所述的网络数据包进行处理以及用于接收由系统主机的操作系统发送的系统数据包;数据分发策略调度单元 403,用于存储物理接口 401 和硬件接收 / 发送队列 402 的映射关系,硬件接收 / 发送单元选择条件和物理接口选择条件,并将网络数据包分配给对应的硬件接收 / 发送单元 402 或将系统数据包分配给对应的物理接口 401。

[0065] 如图 4B 所示,为本发明实施例的系统主机 410 的功能结构,系统主机 410 包括:驱动单元 411,包括用于调用驱动程序对所述的网络适配器进行驱动,该驱动程序包括软件接收 / 发送单元 4111 以及外设描述符 4112;外设描述符 4112 可以将网络适配器的多个物理接口描述为一个网络外设,或者将网络适配器的单个物理接口描述为多个网络外设,使物理接口的数目与外设描述符的数目不相等,也不相关。数据读取单元 412,用于从网络适配

器 400 的数据分发策略调度单元 403 中, 读取物理接口与硬件接收单元的映射关系, 供系统主机的操作系统使用。映射关系设置单元 413, 用于根据数据读取单元 412 读取物理接口与硬件接收单元的映射关系, 对数据分发策略调度单元 403 中物理接口与硬件接收单元的映射关系进行设置。映射关系设置单元 413 可对设置的物理接口与硬件接收单元的映射关系进行存储。主机 410 的操作系统根据读取的物理接口与硬件接收单元的映射关系, 调用操作系统驱动程序对聚合网络适配器进行驱动并将改聚合网络适配器描述为网络外设。主机 410 可以是 PC 或服务器。

[0066] 如图 5 所示, 为本实施例的聚合网络适配器 400 与服务器的主机操作系统 410 间的数据传输, 网络适配器 410 通过总线 420, 连接系统主机 410。聚合网络适配器 200 设置有两个千兆的物理接口 1 和物理接口 2 以及八个硬件接收 / 发送队列 1 至硬件接收 / 发送队列 8。如图 6 所示, 物理接口 1 与硬件接收 / 发送队列 1 至 4 映射, 物理接口 2 与硬件接收 / 发送队列 5 至 8 映射。

[0067] 网络适配器 200 的物理接口 1 通过交换机与以下 4 个内网网段连接:

[0068] 网段 1 :192.168.0.0/24;

[0069] 网段 2 :192.168.1.0/24;

[0070] 网段 3 :192.168.2.0/24;

[0071] 网段 4 :192.168.3.0/24;

[0072] 网络适配器 200 的物理接口 2 通过交换机与以下 4 个外网网段网段连接:

[0073] 网段 5 :10.0.0.0/24;

[0074] 网段 6 :10.0.1.0/24;

[0075] 网段 7 :10.0.2.0/24;

[0076] 网段 8 :10.0.3.0/24;

[0077] 当物理接口 1 接收源地址为 192.168.0.1 和 192.168.1.4 的数据包时, 聚合网络适配器 400 的数据验证单元 (图 4A 未示) 首先验证数据包是否正确, 若验证结果为数据包正确, 则数据分发策略调度单元 403 根据图 7 中数据包源 IP 地址与数据接收 / 发送队列的对应关系, 将源 IP 地址为 192.168.0.1 的数据包分发给硬件接收 / 发送队列 1, 将源地址为 192.168.1.4 的数据包分发给硬件接收 / 发送队列 2。

[0078] 硬件接收 / 发送队列 1 和硬件接收 / 发送队列 2 通过 DMA (Direct MemoryAccess, 直接内存访问) 方式, 主动向系统主机 410 传送数据包。当硬件接收 / 发送队列 1 完成数据包传送后, 硬件接收 / 发送队列 1 向服务器的 CPU (图 4B 未示) 发生中断, 系统主机 410 的操作系统中与硬件接收 / 发送队列 1 对应的软件数据软件接收 / 发送队列 1 (驱动程序) 的中断服务程序, 对软件接收 / 发送队列 1 的中断进行响应, 并控制取得此数据包。

[0079] 然后, 驱动程序将数据包提交给上层协议栈, 然后提交净荷给用户程序, 用户程序对数据处理后进行响应, 将需要回应的数据提交给协议栈, 协议栈附加包头等信息后构成一个完整数据包传递给驱动程序, 驱动程序将数据包的目的地址 10.0.3.1 告知聚合网络适配器的硬件接收 / 发送队列 7, 硬件接收 / 发送队列 7 通过 DMA 方式获取此数据包。数据分发策略调度单元 403 根据图 8 所示数据包目的 IP 地址与物理接口的对应关系, 将数据包分发至物理接口 2。物理接口 2 将数据包发送至相应的网络设备 (图 4A 未示)。

[0080] 同样, 当硬件接收 / 发送队列 2 完成数据包传送后, 硬件接收 / 发送队列 2 向服务

器的 CPU(图 4A 未示)发生中断,系统主机 410 的操作系统中与硬件接收 / 发送队列 2 对应的驱动程序的中断服务程序,对硬件接收 / 发送队列 2 的中断进行响应,并控制取得此数据包。

[0081] 硬件接收 / 发送队列 2 的驱动程序将数据包提交给上层协议栈,然后提交净荷给用户程序,用户程序对数据处理后进行响应,将需要回应的数据提交给协议栈,协议栈附加包头等信息后构成一个完整数据包传递给驱动程序,驱动程序将数据包的目的地址 192.168.1.4 告知聚合适配器的硬件接收 / 发送队列 2,硬件接收 / 发送队列 2 通过 DMA 方式获取此数据包。数据分发策略调度单元 203 图 8 所示数据包目的 IP 地址与物理接口的对应关系,将其分发至物理接口 1。物理接口 1 将数据包发送至相应的网络设备(图 4A 未示)。

[0082] 若本实施例中服务器具有八个 CPU 核,则可将硬件接收 / 发送队列 1—8 在主机操作系统 220 中对应的每个软件硬件接收 / 发送队列以及外设的中断和内存,分别绑定到服务器的每个 CPU 上。譬如:将硬件接收 / 发送队列 1 在主机操作系统中对应的软件接收 / 发送队列 1 和外设 1 绑定在 CPU0 上,同理,将软件接收 / 发送队列 2—8 以及外设 2—8 的中断和内存,分别绑定在 CPU1—CPU7 上,使每个 CPU 处理一个硬件接收 / 发送队列中的数据,从而充分利用多 CPU 系统的处理能力。在网络适配器 200 具有双千兆物理接口的流量下,服务器的每个 CPU 仅需处理 250M 左右的流量即可。

[0083] 本发明中提供的网络适配器可将单个物理接口的数据聚合到多个硬件接收发送队列中,也可将多个物理接口的数据分发到多个硬件接收发送队列。为方便起见,下面将称之为聚合网络适配器。

[0084] 硬件接收队列选择条件与物理接口选择条件的形式化定义

[0085] 设聚合网络适配器中包含 M 个物理接口、N 个硬件接收 / 发送队列,各个物理接口以 p_i 表示,其中 i 为 1 到 M 之间的整数,所有物理接口的集合为 P, 2^P 为 P 的所有子集的集合,各个硬件接收 / 发送队列以 q_j 表示,其中 j 为 1 到 N 之间的整数,所有硬件接收 / 发送队列的集合为 Q, 2^Q 为 Q 的所有子集的集合,网络适配器中的硬件接收队列选择条件为从集合 P 到 2^Q 的函数 f :

[0086] $f:P \rightarrow 2^Q$

[0087] 由于空集 Φ 包含在集合 2^Q 中,因此允许物理接收接口不映射到硬件接收 / 发送队列中任何硬件接收队列中。

[0088] 聚合网络适配器中的物理接口选择条件为从集合 Q 到 2^P 的函数 g :

[0089] $g:Q \rightarrow 2^P$

[0090] 同样由于空集 Φ 包含在集合 2^P 中,因此允许硬件接收 / 发送队列的硬件发送队列不映射到任何物理接口上。

[0091] 根据硬件接收队列选择函数 f,即接收数据时的对应关系,定义接收聚合各名词如下:

[0092] 若硬件接收队列选择函数 f 满足如下条件,则称之为接收正向聚合:

[0093] $\{f \mid |f(p_i)| \geq 1 \text{ 且 } \cap f(p_i) = \Phi, 1 \leq i \leq M\}$

[0094] 即对任意一个物理接口的接收,映射后的集合中元素的个数超过 1 个,且各个物理接口映射的硬件接收队列是不同的。

- [0095] 若硬件接收队列选择函数 f 满足如下条件, 则称之为接收反向聚合:
- [0096] $\{f \mid |f(p_i)| = 1, 1 \leq i \leq M\}$
- [0097] 即对任意一个物理接口的接收, 只能映射到一个硬件接收队列中, 但多个物理接口可以映射到相同的硬件接收队列。
- [0098] 若硬件接收队列选择函数 f 满足如下条件, 则称之为接收平行聚合:
- [0099] $\{f \mid |f(p_i)| = 1 \text{ 且 } \cap f(p_i) = \emptyset, 1 \leq i \leq M\}$
- [0100] 接收平行聚合使接收正向聚合与接收反向聚合的交集, 即一个物理接口对应一个硬件接收队列。
- [0101] 接收正向聚合与接收反向聚合的并集称为接收单一聚合为, 接收混合聚合为接收单一聚合的补集, 即多个物理接口同时对应多个硬件接收队列。
- [0102] 根据物理接口选择关系 g , 即发送数据里的对应关系, 定义发送聚合名词如下:
- [0103] 若物理接口选择函数 g 满足如下条件, 则称之为发送正向聚合:
- [0104] $\{g \mid |g(q_j)| = 1, 1 \leq j \leq N\}$
- [0105] 即对任意一个发送队列, 只能映射到一个物理发送接口中, 但多个发送队列可以映射到相同的物理发送接口。
- [0106] 若物理接口选择函数 g 满足如下条件, 则称之为发送反向聚合:
- [0107] $\{g \mid |g(q_i)| \geq 1 \text{ 且 } \cap g(q_i) = \emptyset, 1 \leq i \leq N\}$
- [0108] 即对任意一个硬件发送队列, 可映射到一个或多个物理接口, 但各个硬件发送队列映射的物理接口不同。
- [0109] 若物理接口选择函数 g 满足如下条件, 则称之为发送平行聚合:
- [0110] $\{g \mid |g(q_j)| = 1 \text{ 且 } \cap g(p_j) = \emptyset, 1 \leq j \leq N\}$
- [0111] 即为接收正向聚合与接收反向聚合的交集。
- [0112] 发送单一聚合为发送正向聚合与发送反向聚合的并集, 发送混合聚合为发送单一聚合的补集。
- [0113] 根据上述硬件接收队列选择函数 f 与物理接口选择函数 g 的关系, 若硬件接收队列选择函数 f 与物理接口选择函数 g 所表示的对应关系是互逆的, 称为互逆聚合。
- [0114] 互逆聚合定义如下:
- [0115] $\{(f, g) \mid q_j \in f(p_i) \Leftrightarrow p_i \in g(q_j)\}, 1 \leq i \leq M, 1 \leq j \leq N;$
- [0116] 综合考虑上述硬件接收队列选择函数 f 与物理接口选择函数 g , 对聚合网络适配器中的聚合做如下定义:若硬件接收队列选择函数 f 为接收正向聚合, 且发物理接口选择函数 g 为接收正向聚合, 则合称为正向聚合。若 f 为接收反向聚合, 且 g 为发送正向聚合, 则称为反向聚合。若 f 为接收平行聚合, 且 g 为发送平行聚合, 则称为平行聚合。若 f 为接收单一聚合, 且 g 为发送单一聚合, 则称为单一聚合。若 f 为接收混合聚合, 且 g 为发送混合聚合, 则称为混合聚合。
- [0117] 既是正向聚合又是互逆聚合, 称为互逆正向聚合。既是反向聚合又是互逆聚合, 称为互逆反向聚合。既是平行聚合又是互逆聚合, 称为互逆平行聚合。既是单一聚合又是互逆聚合, 称为互逆单一聚合。既是混合聚合又是互逆聚合, 称为互逆混合聚合。
- [0118] 由于本实施例中, 物理接口数 $M = 2$; 硬件接收 / 发送队列数目 $N = 8$; 则物理接口与硬件接收 / 发送队列的聚合对应函数为: 互逆正向聚合方式。接收对应函数 f 为: $f(1)$

$= \{1, 2, 3, 4\}$, $f(2) = \{5, 6, 7, 8\}$ 。发送对应函数 g 为 : $g(1) = \{1\}$, $g(2) = \{1\}$, $g(3) = \{1\}$, $g(4) = \{1\}$, $g(5) = \{2\}$, $g(6) = \{2\}$, $g(7) = \{2\}$, $g(8) = \{2\}$ 。

[0119] 本实施例中,图 7 和图 8 中所示的数据分发策略是按照数据包的源地址和目的地址设置的,使源地址和目的地址相同的数据包由聚合网络适配器中的同一硬件接收 / 发送队列进行处理。在设置聚合网络适配器的数据分发调度数据分发策略时,在本实施例上述例举的物理接口与硬件接收 / 发送队列的映射关系中,接收正向聚合与发送反向聚合中每个物理接口与多个硬件接收 / 发送队列对应,可以按照一对多的形式,将源地址相同的数据包按照一定的规则平均地分发到多个硬件接收 / 发送队列。硬件接收 / 发送队列的数据包将数据包复制为多份,往每个硬件接收 / 发送队列的数据包都送一份。用户可根据不同的需求,采取不同的方式和规则,设置数据分发策略。

[0120] 本实施例的聚合网络适配器中,硬件接收 / 发送队列 1 与其驱动程序间的数据交互,以及硬件接收 / 发送队列 2 与其驱动程序间的数据交互方式,采用了相同的数据交互方式 DMA。但是,本领域技术人员根据本实施例的内容,对各硬件接收 / 发送队列与其驱动程序间的数据交互方式进行变化,硬件接收 / 发送队列可采用不同方式和数据结构与系统主机 410 的操作系统交换数据,如采用基于描述符环的方式交互数据,或采用基于流的方式交互数据或其他交换方式,使得各硬件接收 / 发送队列与系统主机 410 的操作系统交互的方式可以相同也可以不同,使得聚合网络适配器 200 中硬件接收 / 发送队列的硬件实现可以相同也可以不同。

[0121] 本实施例的有益效果在于,根据设置的数据分发策略,将聚合网络适配器应用为单口网络适配器或多口网络适配器,可以将单个物理接口的流量均衡到多个接收模块上,也可以将多个物理接口的流量合并到同一个发送模块上。当物理接口接收的数据流量较小时,可以将聚合网络适配器的硬件接收 / 发送队列的多路数据集合至一个 CPU 集中处理;当物理接口接收的网络数据流量较大,可以将数据分流至多个硬件接收 / 发送队列,在具有多 CPU 处理能力的系统中,可以充分发挥该系统多处理平台对聚合网络适配器物理端口流量的处理能力。

[0122] 实施例 2

[0123] 本实施例的系统与实施例 1 相同,仍如图 4A 所示,本发明实施例的聚合网络适配器 400 设置有两个千兆的物理接口 1 和物理接口 2 以及八个硬件接收 / 发送队列 1 至硬件接收 / 发送队列 8;且如图 6 所示,物理接口 1 与硬件接收 / 发送队列 1 至 4 映射,物理接口 2 与硬件接收 / 发送队列 5 至 8 映射。

[0124] 本实施例的数据分发策略与实施例 1 不同,每个硬件接收 / 发送队列的选择条件、与物理接口选择条件如图 9 与图 10 所示。两个物理接口与各个网段相连接,但映每个硬件接收 / 发送队列与物理接口的映射方式不是与 IP 地址相关,而是与端口号相关。如从物理接口 1 进入的数据包中,所有 21 端口的数据都被分配到硬件发送队列 1,所有 80 端口的数据都被分配到硬件发送队列 2 等。

[0125] 21 端口为 FTP 服务的知名端口号,21 端口的数据都属于 FTP 协议的命令数据,故可在硬件接收 / 发送队列 1 上启动 FTP 的相关处理程序;80 端口为 HTTP 服务的知名端口号,80 端口的数据都属于 HTTP 服务的数据,可在硬件接收 / 发送队列 2 上启动 HTTP 的相关处理程序。通过本实施例可以看到,本发明中的聚合网络适配器根据合理的配置,可将不同

类型的服务分流到不同的硬件接收 / 发送队列上,从而为上层应用提供更大的方便。

[0126] 同时,聚合网络适配器中的硬件接收队列选择条件和物理接收选择条件是动态,这为如FTP等服务提供了额外的方便。在FTP服务的工作方式中,存在被动连接 (passive) 的工作方式,用户与FTP服务间的数据连接也由客户端发起,服务器提供一个动态分配的用于与用户进行数据连接的端口号,该端口号由FTP协议中的Pssv命令指定。对于这种情况,根据对所有经过本网卡上接收和发送的数据报文进行分析,可以获得用户进行被动连接的端口号,并动态将与该端口进行连接传输FTP数据的网络数据报文都分配到硬件接收和发送队列1中。在本实施例中,聚合网络适配器将Pssv命令所指定的端口数据加入到硬件接收 / 发送队列1中。聚合网络适配器本身可完成捕获Pssv命令的过程,并自适应地动态将该端口的数据映射到相应的硬件接收 / 发送队列中;也可由软件直接动态将一个对应关系设置添加到硬件接收队列选择条件和物理接收选择条件中。

[0127] 以上所述的具体实施方式,对本发明的目的、技术方案和有益效果进行了进一步详细说明,所应理解的是,以上所述仅为本发明的具体实施方式而已,并不用于限定本发明的保护范围,凡在本发明的精神和原则之内,所做的任何修改、等同替换、改进等,均应包含在本发明的保护范围之内。

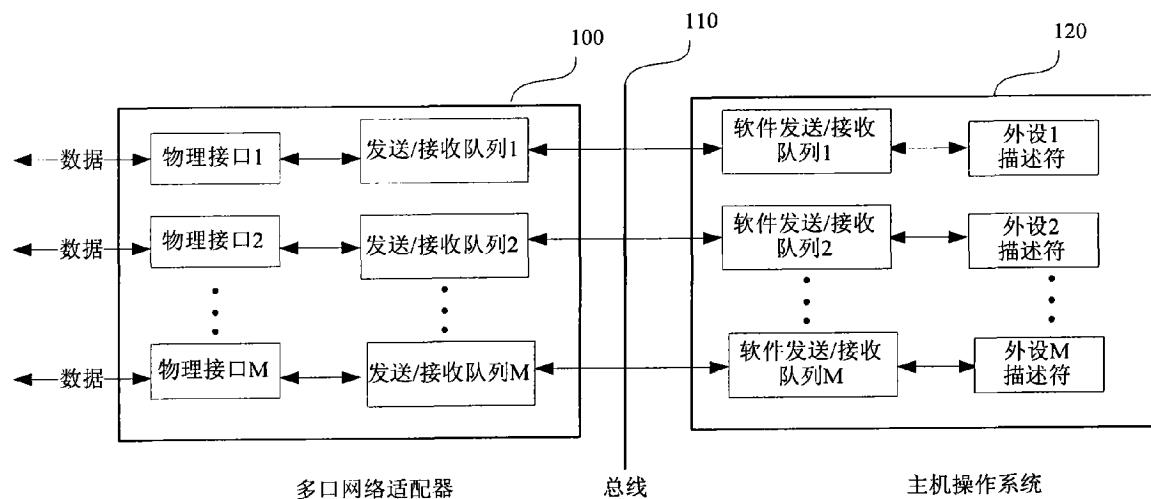


图 1

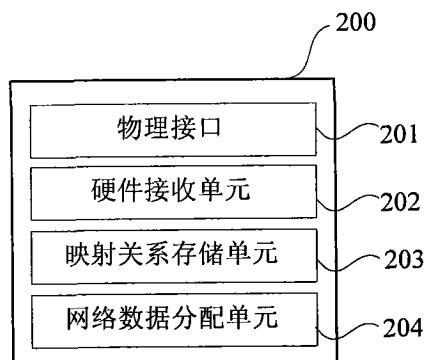


图 2A

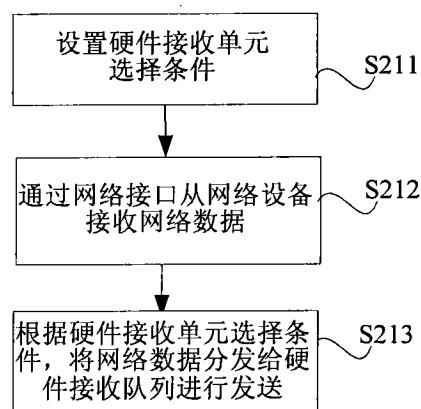


图 2B

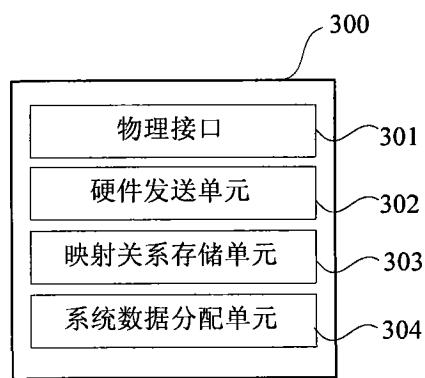


图 3A

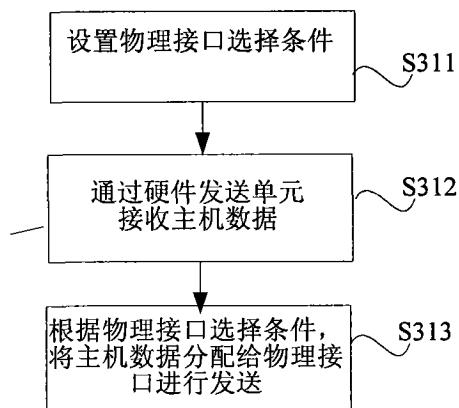


图 3B

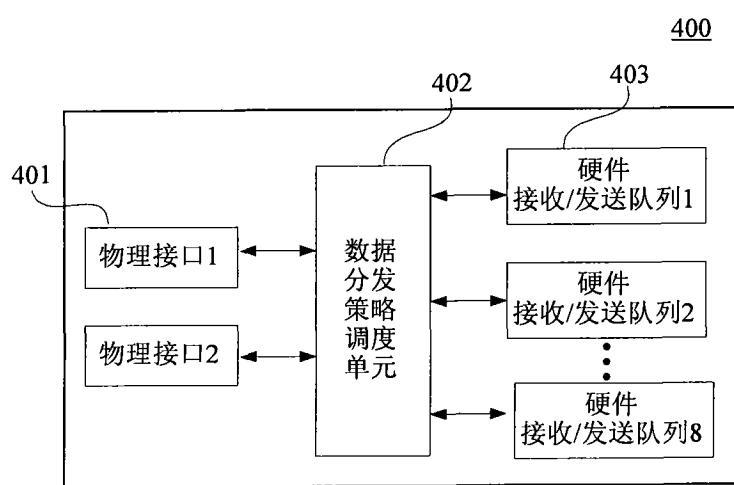


图 4A

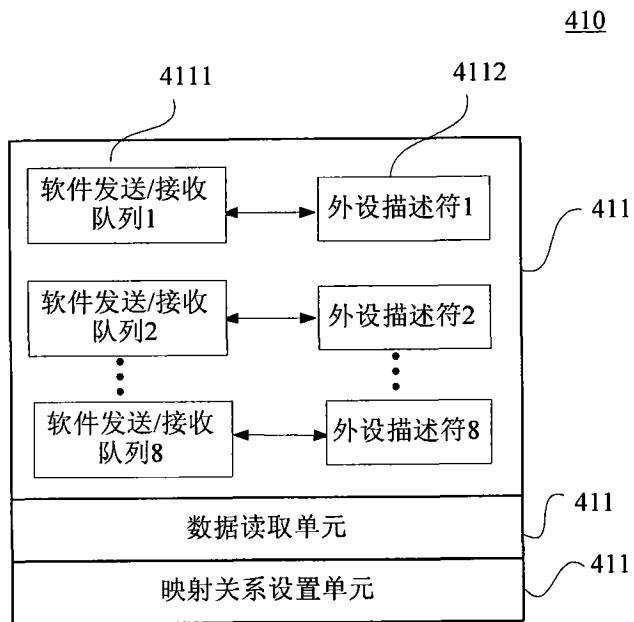


图 4B

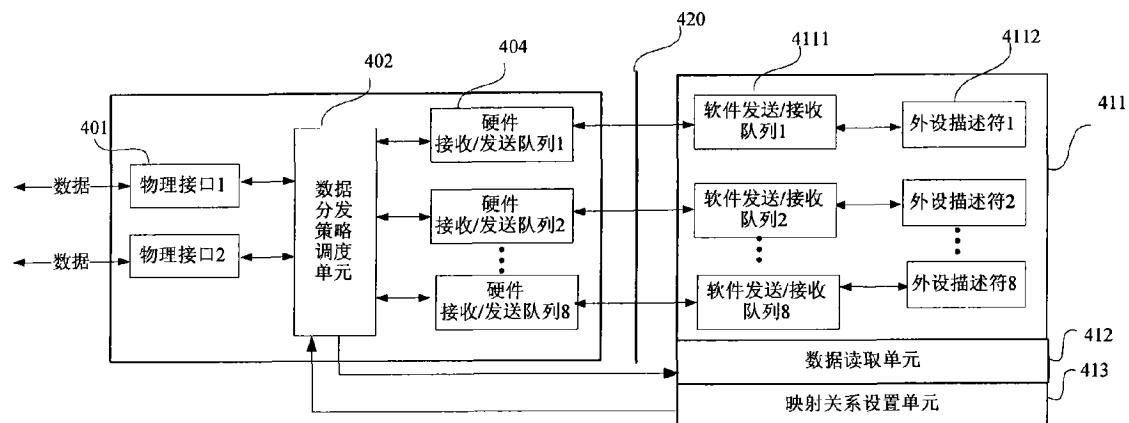


图 5

物理接口号	映射的硬件接收/发送队列号
物理接口 1	硬件接收/发送队列 1
物理接口 1	硬件接收/发送队列 2
物理接口 1	硬件接收/发送队列 3
物理接口 1	硬件接收/发送队列 4
物理接口 2	硬件接收/发送队列 5
物理接口 2	硬件接收/发送队列 6
物理接口 2	硬件接收/发送队列 7
物理接口 2	硬件接收/发送队列 8

图 6

硬件接收/发送队列号	源 IP 地址
硬件接收/发送队列 1	192.168.0.0/24
硬件接收/发送队列 2	192.168.1.0/24
硬件接收/发送队列 3	192.168.2.0/24
硬件接收/发送队列 4	192.168.3.0/24
硬件接收/发送队列 5	10.0.0.0/24
硬件接收/发送队列 6	10.0.1.0/24
硬件接收/发送队列 7	10.0.2.0/24
硬件接收/发送队列 8	10.0.3.0/24

图 7

物理接口号	目的 IP 地址
物理接口 1	192.168.0.0/24
物理接口 1	192.168.1.0/24
物理接口 1	192.168.1.0/24
物理接口 1	192.168.1.0/24
物理接口 2	10.0.0.0/24
物理接口 2	10.0.1.0/24
物理接口 2	10.0.2.0/24
物理接口 2	10.0.3.0/24

图 8

硬件接收/发送队列号	源 IP 地址	源端口地址
硬件接收/发送队列 1	192.0.0.0/8	21
硬件接收/发送队列 2	192.0.0.0/8	80
硬件接收/发送队列 3	192.0.0.0/8	22
硬件接收/发送队列 4	192.0.0.0/8	110
硬件接收/发送队列 5	10.0.0.0/8	21
硬件接收/发送队列 6	10.0.0.0/8	80
硬件接收/发送队列 7	10.0.0.0/8	22
硬件接收/发送队列 8	10.0.0.0/8	110

图 9

物理接口号	目的 IP 地址	目的端口地址
物理接口 1	192.0.0.0/8	21
物理接口 1	192.0.0.0/8	80
物理接口 1	192.0.0.0/8	22
物理接口 1	192.0.0.0/8	110
物理接口 2	10.0.0.0/8	21
物理接口 2	10.0.0.0/8	80
物理接口 2	10.0.0.0/8	22
物理接口 2	10.0.0.0/8	110

图 10