文章编号:1001 - 9081(2005)01 - 0173 - 03

高速通信网卡中 PCI 接口的研究与实现

雷艳静,苗克坚,康继昌
(西北工业大学 计算机学院,陕西 西安 710072)
(leizilei@163.com)

摘 要:机群系统以其优异的性价比正被应用于越来越多的场合。文中分析了机群系统中高速通信网卡对 PCI 接口的要求,采用紧凑设计思想,将网卡的功能逻辑与 PCI 接口实现在一个 FPGA 芯片中。该 PCI 接口可分别以主模式和从模式进行工作。应用于微机与 SMP 机中,性能良好。

关键词:PCI 总线;机群;信令寻径式网络;DMA
中图分类号: TP393.02 **文献标识码**:A

Research and implementation of PCI interface in high-speed network adapter

LEI Yan-jing, MIAO Ke-jian, KANG Ji-chang
(Institute of Computer Science, Northwestern Polytechnical University, Xi'an Shaanxi 710072, China)

Abstract: Cluster is applied to more and more fields for its high price-performance. Requirements for high speed network adapter in cluster were analyzed in this thesis. The function logic of the network adapter and PCI interface could be implemented in a single FPGA chip. The PCI interface could work in master mode or slave mode respectively. Now the PCI interface has already been applied to PCs and SMP computers, and it is proved to perform well.

Key words: PCI bus; cluster; Token-Routing Net; direct memory access

0 引言

基于 PC 机或工作站的机群系统以其高可用性、高性价比[1]等特点备受用户青睐。机群系统目前正被应用于各种场合,如高性能科学计算、航空电子综合系统和高性能服务器等领域。目前构成机群的通信互联设备大多基于 PCI 总线。常用的 33MHz/32 位的 PCI 总线峰值传输速率可达 132 MB/s,而 66MHz/64 位 PCI 总线的性能更能成倍地提高。PCI 总线以其速度快、扩展性好、使用方便[2,3]等特点,从而成为目前机群通信互联设备的主流连接方式,如千兆以太网、Myrinet、基于 PCI 的 SCI 设备等。

PCI 接口的实现是设计高速网络互联设备的主要部分,有两种常见的实现方式。一种是采用专用的 PCI 接口芯片来实现,如 AMCC 公司的 S5933 和 S5920、PLX 公司的 PCI9050 和 PCI9054 等。这种方式可方便地实现 PCI 接口,但系统的性能不够优化,设计上也缺乏灵活性;另一种方法是采用 FPGA/CPLD 芯片自行设计 PCI 接口逻辑,该方法可根据系统需要有选择性地实现 PCI 的相应功能,设计灵活,有利于系统优化,且具有较高的性价比。

为提高通信效率,笔者在实现基于信令寻径技术[4](专利号[CN1299204A])的高速通信网卡时,采用基于 FPGA 的 PCI 接口实现方法,将网卡的逻辑功能与 PCI 接口实现在一个 FPGA 芯片中,这样有利于优化整个网卡的系统性能。而且,采用这种实现方法,在网卡升级时,只需修改 FPGA 的内部逻辑,而无需更新 PCB 布局布线,故网卡升级非常方便。

1 信令寻径式高速通信网卡的逻辑功能及对 PCI 接口的要求

信令寻径式高速通信网卡的内部结构如图 1 所示。它主要包括发送逻辑、接收逻辑、核心控制逻辑以及 PCI 接口逻辑四部分。其中,发送逻辑用于数据的发送;接收逻辑用于从传输介质接收数据;核心控制逻辑负责对发送逻辑和接收逻辑进行协调与控制。

在信令寻径式高速网络的数据通信过程中,PCI 接口实现了网卡与主机之间的信息交互。首先,主机发送数据的控制命令要通过网卡的 PCI 接口被送入网卡的控制寄存器组中;然后,网卡控制逻辑对控制命令进行解释和处理,若控制命令为发送数据,则核心控制逻辑启动 PCI 接口的 DMA 引擎,以主模式方式从内存中读取有效数据,有效数据通过发送

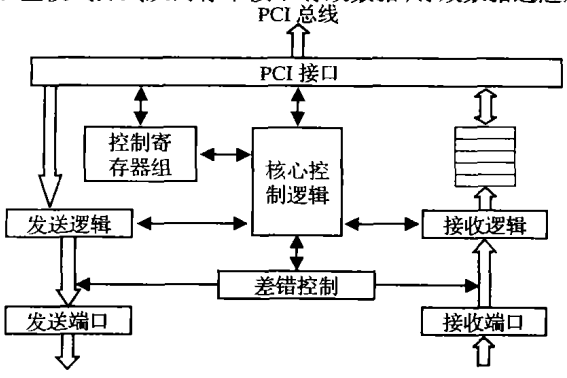

图 1 信令寻径式高速通信网卡的内部结构

收稿日期:2004 - 06 - 05;**修订日期**:2004 - 12 - 02
基金项目:国家 863 计划资助项目(2003AA001018);航空科学基金资助项目(02F53031)

作者简介:雷艳静(1979 -),女,河南洛阳人,博士研究生,主要研究方向:并行/分布式计算、机群计算、机群网络协议; 苗克坚(1962 -),男,辽宁营口人,教授,博士,主要研究方向:并行/分布式计算、高性能计算; 康继昌(1930 -),男,上海市人,教授,博士生导师,主要研究方向:并行/分布式计算、机群计算、高性能计算.

逻辑进行发送;在接收方,网卡检测到传输介质中有数据到来时,核心控制逻辑会启动 PCI 接口的 DMA 引擎,以主模式将收到的有效数据送入主机内存,从而完成数据通信。

从上述分析可见,数据通信过程中,网卡与主机之间要通过 PCI 接口进行大量的信息交互。因此,网卡的 PCI 接口要能够提供如下的服务:

(1) 要能够作为从设备,以从模式方式进行工作。网卡通过从模式工作方式,接收用户的命令,包括发送数据的命令、网卡复位和读取网卡工作状态信息等。同时,在网卡初始化时,系统要对 PCI 设备进行配置,所有对 PCI 设备配置空间的访问都通过从模式方式进行。

(2) 要能够作为主设备,以主模式方式进行工作。网卡通过主模式,实现了主存与网卡之间直接的数据传输,主模式也称为 DMA 方式。在数据通信过程中,网卡通过 DMA 方式从主机内存中读数据或向内存中写数据,主机 CPU 可以进行用户程序的计算。因此,通过 DMA 方式的数据通信,可以实现主机 CPU 计算和 I/O 操作的并行工作,有利于提高系统整体效率。

2 高速通信网卡中的 PCI 接口的实现

PCI 总线是基于握手协议的主/从控制总线。提出数据传输的设备称为主设备(Master),响应数据传输的设备称为目标设备或从设备(Slave),数据传输在主/从设备的握手下完成。PCI 设备的传输过程通常包括以下几个步骤:

(1) 传输启动。主设备通过拉低 FRAME 信号启动一次传输,此时 A/D 线上包含了要访问的地址信息,C/BE 线上给出访问的命令类型。各个从设备要锁存这些信息并进行译码。

(2) 设备响应。若从设备译码发现本次 PCI 访问的地址区间落在自己的地址段内或是针对自己的配置访问,则拉低 DEVSEL 信号以表明自己愿意承担本次传输。若设备不能在规定的时间内(16 个 PCI 时钟周期)进行第一个数据传输,则要发出重试(Retry)信息(拉低 DEVSEL 的同时拉低 STOP 信号)。

(3) 数据传输。若设备满足第一次数据传输的响应时间要求,则根据设备状态控制 TRDY 进行数据传输。在时钟上升沿,若 IRDY 和 TRDY 同时有效,则进行一次传输。否则,由主设备(IRDY 为高)或从设备(TRDY 为高)插入等待周期。

(4) 传输终止。传输完成之后或者需要停止传输时,从设备可发出 STOP 信号来要求终止目前的传输。主设备接到该握手信号后会立即拉高 IRDY 和 FRAME 以指示传输结束。为了避免信号竞争,还要经过一个时钟周期的持续三态信号驱动时间,设备才能彻底停止驱动总线。

依据这个传输过程,可分别进行主模式和从模式状态机的设计。

2.1 PCI 接口从模式的实现

PCI 接口从模式有限状态机如图 2 所示,状态机的具体转变过程如下:

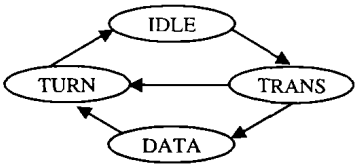

图 2 PCI 接口从模式有限状态机

(1) IDLE:PCI 接口在空闲时处于 IDLE 状态。

(2) IDLE→TRANS:用户通过网卡接口向网卡发送控制命令,使得 PCI 总线的 FRAME 信号有效。当 FRAME 有效并且总线空闲时,PCI 接口的状态机从 IDLE 态转变成 TRANS 状态。

(3) TRANS:在 TRANS 状态时,PCI 接口对操作命令和地址分别进行锁存和译码。

(4) TRANS→SDATA:如果译码的结果为对本网卡进行操作,则 PCI 接口状态机变为 SDATA 状态,准备进行数据传输。

(5) TRANS→TURN:如果译码的结果不是对本网卡进行操作,则 PCI 接口状态机直接转到 TURN 状态。

(6) SDATA:PCI 接口发出 TRDY、DEVSEL 等响应信号,同时网卡根据译码的结果进行相应的操作。

(7) SDATA→TURN:主设备撤销 FRAME 信号。

(8) TURN→IDLE:PCI 接口完成操作,撤销响应信号,进入 IDLE 状态。

至此,网卡的 PCI 接口以从模式方式完成了用户的操作命令。

2.2 PCI 接口主模式的实现

图 3 给出了 PCI 接口主模式的状态机。主模式包括 DMA 读和 DMA 写两种操作,虽然这两种操作数据传输的方向相反,但操作过程的状态机则基本相同。为描述方便,现仅对 DMA 写的状态机进行说明。

(1) IDLE:PCI 接口不通过 DMA 方式向内存写数据时,处于 IDLE 状态。

(2) IDLE→ADDR:PCI 接口接到核心控制逻辑发送来的 DMA 写请求后,向系统申请 PCI 总线的使用权。PCI 总线仲裁逻辑通过一定的仲裁机制对总线的使用权进行仲裁。若总线仲裁逻辑同意网卡对 PCI 总线使用权的申请,则向网卡 PCI 接口发出 GNT 响应信号。PCI 接口取得了 PCI 总线的使用权之后,可以利用总线通过 DMA 方式向内存写数据。这时,PCI 接口由 IDLE 态转变成 ADDR 态。

(3) ADDR:在该状态下,PCI 接口将锁存 DMA 写的目标内存起始地址和写操作命令,并进行相应的译码。

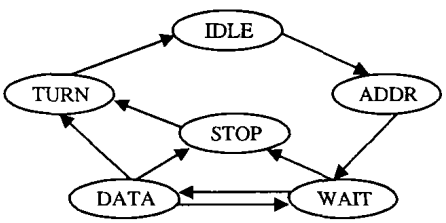

图 3 PCI 接口主模式 DMA 写状态机

(4) WAIT:PCI 接口等待从设备(主机内存)准备好接收数据。从设备准备好后,DMA 写的状态机进入 DATA 态。

(5) DATA:进行 DMA 写操作。若在写的过程中,从设备不能立即完成操作,则 PCI 接口再次进入 WAIT 态。

(6) STOP:在数据传输过程中,若 PCI 总线仲裁器逻辑剥夺了网卡的总线使用权,则网卡不能继续使用 PCI 总线进行 DMA 写操作,网卡的 DMA 写状态机只能从 DATA 态进入 STOP 态;并且,若网卡在 DMA 写的过程中不能立即准备好数据,从而也不能继续进行 DMA 写操作,此时网卡 DMA 写状态机也进入 STOP 态。在 STOP 态,网卡处理释放 PCI 总线使用权的善后操作,然后进入 TURN 态,最后进入 IDLE 态。在 IDLE 态,网卡 DMA 写逻辑重新申请总线的使用权,申请成功后继续进行 DMA 写操作。

(7) TURN:缓冲一拍,继续进行一些善后操作,再进入 IDLE 态。

至此,网卡的 PCI 接口以主模式方式完成了 DMA 写的功能。

3　结语

机群系统以其优异的性能正被应用于越来越多的场合。基于 PCI 接口的网络通信设备成为主流。为提高系统效率,优化性能,笔者在实现信令寻径式高速通信网卡时采用基于 FPGA 的 PCI 接口实现方法,将网卡的逻辑功能与 PCI 接口实现在一个 FPGA 芯片中。本文分析了高速通信网卡对 PCI 接口的要求,实现了 PCI 接口的主/从工作模式。目前,该 PCI 接口逻辑已分别应用于微机与 SMP 机中,使用性能良好。

参考文献:

[1] (美)GREGORY F, FISTER P. In Search of Clusters (Second Edition) [M]. PRENTICE HALL, 1998.

[2] (美) SHANLEY T, ANDERSON D. PCI System Architecture (Fourth Edition) [M]. Addison-Wesley, 2000.2 - 3.

[3] 李贵山, 戚德虎. PCI 总线应用指南[M]. 西安: 西安电子科技大学出版社, 1996.

[4] 苗克坚, 康继昌, 等. 低度并行系统互连网络交换器的研究与设计[J]. 西北工业大学学报, 1999, 17(12): 159 - 163.

[5] 苗克坚, 等. 微机 PCI 总线接口的研究与设计[J]. 航空计算技术, 1998, 28(1).

[6] 田秀珍, 岳海霞. PCI 局部总线接口设计方法[J]. 山西电子技术, 2001, (4).

[7] 张蕴玉, 胡修林, 等. PCI 总线接口设计及其专用芯片应用[J]. 电子工程师, 2002, 28(2).

[8] 张玉. PCI 总线接口及其可编程逻辑设计[J]. 南京邮电学院学报, 1996, 16(4).

[9] 方粮, 尹佳斌, 等. PCI 总线卡设计与实现的几个关键问题[J]. 计算机自动测量与控制, 2000, (2).

(上接第 164 页)

范畴,是处于操作系统软件与用户的应用软件的中间。中间件在操作系统、网络和数据库之上,应用软件的下层,总的作用是为处于自己上层的应用软件提供运行与开发的环境,帮助用户灵活、高效地开发和集成复杂的应用软件。

在众多的中间件中,消息中间件是一种基本的中间件,适用于任何需要进行网络通信的系统,负责建立网络通信的通道,进行数据或文件发送。利用消息中间件进行通信时,通信的双方通过一系列的消息进行通信。当消息在发送方和接收方之间进行传输时,消息的发送者和接收者不一定同时连接,消息中间件将消息放进某个消息队列中,这个消息队列防止消息在传输过程中丢失。然后由消息中间件将这个消息传输到接收方。接收方的消息中间件在接收到完整的消息后,将消息放到某个消息队列中,并通知接收方到该消息队列中去取消息。但即使发送者和接收者之间的通信不是同时发生的,通信过程中也会有请求/应答的保障机制。

因此,在设计 Ad hoc 网络体系结构时,考虑到 Ad hoc 的无线、移动和自组织的特点,由于无线通信不可靠性和无线连接微弱性和高比特错误率,有线网络中的应用不能在无线网络中正常运行,与有线网络服务相比,无线通信具有通信距离短、传输时延长和传统输速率低的缺点。无线通信的以上限制严重削弱了在计算机和信息技术(尤其是网络和分布式计算)在移动作战和移动救灾中的能力。针对无线通信内在弱点,我们借鉴消息中间件的思想,在传输层和应用层之间增加一层中间件,屏蔽操作系统和网络的低层细节,同时增加通信的可靠性、安全性。

5　结语

Ad hoc 网络是一种特殊的对等式移动网络,它具有无中心、自组织、多跳路由等特点。在军事应用和其他特殊场合有着广泛的应用前景。

参考文献:

[1] MACKER J, CORSON S, FENNER B, *et al.* Mobile Ad-hoc Networks (manet) [EB/OL]. http://www.ietf.org/html.charters/manet-charter.html.

[2] IEEE STD 802.11. Wireless LAN Medium access Control(MAC) and physical layer(PHY) specifications[S], 1999.

[3] 3GPP. Opportunity Driven Multiple Access, 3G TR 25.924 V1.0.0 [S].

[4] UMTS. UMTS terrestrial radio access(UTRA): concept evaluation, UMTS:30.06 V3.0.0[S].

[5] XU SG, SAADAWI T. Does the IEEE 802.11 MAC protocol work well in multihop wireless ad hoc network[J]. IEEE Communication Magazine, 2001, (6), 130 - 137.

[6] CHANDRA A, GUMMALLA V, LIMB JO. Effect of turn-around times on the performance of high speed Ad-hoc MAC protocols[A]. Proc. Networking 2000 Conf[C]. Paris, May 2000. 507 - 517.

[7] BROCH J, MALTZ DA, JOHNSON DB, *et al.* A performance comparison of multi-hop wireless ad hoc network routing protocols[A]. Proceeding of the Fourth Annual ACM/IEEE International Conference on Mobile Computing and Networking[C]. Dallas, TX, October 1998. 85 - 97.