国防科技大学学报
JOURNAL OF NATIONAL UNIVERSITY OF DEFENSE TECHNOLOGY
第18卷第4期 1996年12月 Vol. 18 No. 4

高性能的智能FDDI网络控制器设计技术*

苏金树

(国防科技大学计算机学院 长沙 410073)

摘 要 在分析传统的网络功能实现结构及通信开销的基础上，论述了高性能网络控制器的设计，并提出提高网络性能的硬件途径和面向FDDI特点的FDDI控制器的设计技术。

关键词 FDDI，计算机网络，网卡

分类号 TP393

The Technology of Designing the High Performance FDDI Network Controller

Su Jinshu

(Institute of Computer, NUDT, Changsha, 410073)

Abstract Based on the analysis of traditional implementation of network function and its overhead. some consideraton of high performance network controller is given. The hardware approaches of promoting network performance are described. And finally, FDDI features-orientied network controller design technology is also presented.

Key words FDDI, computer network, network interface controller

1 传统的网络功能实现一般结构及通信开销

(1)传统网络协议处理

网络功能一般通过网络硬件控制器接口(简称网络控制器)和主机网络协议处理软件实现。网络硬件连接在某种I/O BUS上，实现数据向介质上发送及从介质上接收，而协议软件处理应用程序的请求，并管理硬件控制器，处理中断等。

对于UNIX系统的应用程序发送数据，SOCKET层将数据从用户空间拷贝到系统

* 1996年4月17日收稿

缓冲区，并启动网络传输协议。如果用户要求实现可靠的流协议，一般采用 TCP/IP。协议处理包括打包、错误处理、端端的流控、路由及拥挤控制。为了效率更高，但功能要求不完整的应用一般采用 UDP/IP 协议。协议处理结束时，产生了输向数据链路层的多个数据包。对于接收方，动作是类似的，次序与发送方相反，数据链路层将数据放到接收缓冲区，并启动协议处理，协议处理完成后将数据拷贝到用户空间。

错误处理通常由可靠传输协议处理，大多数协议通常采用端端检查和的方式验证数据完整性，并通过超时机制检测丢失的数据包。

在 TCP/IP 处理中，SOCKET 将应用程序用户空间的数据拷贝到系统缓冲区，传输协议读入数据并计算检查和，最后数据链路层将数据拷贝到网络适配器。总计起来，数据六次经过存贮器总线。有时候，CPU 还需将零散的数据拷贝到单个设备缓冲区，又增加了总线的传输。

从主机方向来看，网络适配器类似一个带有若干控制寄存器的缓冲区。写入缓冲的报文由网络控制器传输，进来的数据由网络控制器拷贝到主存中，最简单的情况是将数据按 FIFO 方式组织缓冲。缓冲区的大小取决于网络传输的速度，增加缓冲区可以减少控制器与主机的交互，从而提高性能。

(2)通信开销

关于网络的通信开销，许多研究者对不同的体系结构、不同的协议作过严格的数学分析或模拟，得出许多有益的结论，在此不作过多的分析说明，只引述结论。纵观各种观点可知，开销不存在唯一的主要因素，而是由多个因素综合而成。要实现高效的网络功能，需要考虑各种相关的因素，而不仅是网络协议处理。

网络的流量往往受限于发送和接收数据的主机软件开销的限制。减少这些开销可以改进应用程序的吞吐量，降低延迟以及提高受通信开销影响的应用程序速度。网络开销的主要影响因素来自三个方面：①通信协议，②应用程序设计接口，③网络接口硬件体系结构。

高性能网络控制器的设计就是要分析上述因素如何影响通信性能，在哪些条件下，网络控制器的硬件能够减少开销、改善性能。

就通信协议本身而言，报文接收与发送的操作主要是：①传输协议处理；②上下文切换(线程或进程切换与中断处理)；③数据链路协议处理；④缓冲管理。

2 高性能网络的设计考虑

(1)高带宽的处理

网络通信的处理给主机带来沉重的压力，直接结果是应用—应用通信的峰值只能支持每秒几十兆位，即使网络的带宽是 100Mbps 或更高(如 ATM、HIPPI 等)。

随着网络加快，类似数据拷贝及检查和之类的字节操作及检查和开销成为主要开销，原因主要有两个：

①随着网络速度加快，采用的数据包越大，每个包的开销也增大。

②存贮系统通常是计算机系统的最临界资源，逐字节操作大量占用存贮资源。而且网络操作不象数值计算操作，数据局部性非常低、所以传统优化存贮系统的 CACHE 等

技术通常对网络处理无法奏效。

上面已经分析了数据在传统网络界面上的传送次数。如果存储带宽没有远远超过网络带宽，这些传输就会限制网络吞吐量。对于当前工作站的存储带宽(约每秒几百万个字节)，主存的带宽已经成为 100Mbps 以上网络的主要瓶颈。

(2)网络延迟的处理

网络性能的另一个重要参数是短报文的延迟。所谓延迟也就是从一个应用传送到另一个应用所花费的时间。对于短报文。每个包每次接收/发送是整个通信开销的主要部分。此时拷贝字节所花费的开销是很少的。对于短报文，拷贝数据比共享缓冲可能更具有优势，因此此时协议栈不同部分共享缓冲，增加缓冲管理开销，可以通过优化每个包的操作减少延迟。

3 提高性能的硬件途径

通过硬件提高网络性能主要分为三种，第一是提供 DMA 操作能力，加速数据传输；其次是加大适配器的缓冲空间，减少拷贝操作；第三是提供对协议处理的硬件支持，这一点涉及到如何合理划分协议处理，在下一小节论述。

(1)DMA 机制

在网卡上设置 DMA 机制可以减少 CPU 的干预，实现数据在主存和控制器缓冲间的数据传送。与程序 I/O(PIO)相比，DMA 具有以下优势：

①DMA 只需一次经过 BUS，而 PIO 需要 2 次；

②大多数高速总线依赖于大量阵发式传输实现高吞吐量。DMA 机制可以很容易地实现高吞吐率，而 PIO 则受限于移动单字数据。

③DMA 可以与 CPU 并行工作，从而可以提高 CPU 利用率，重叠的比率则依赖于存贮器的带宽和 CACHE 的命中率。

虽然 DMA 可以获得较高的峰值，但依然有一些额外的开销：

①采用 DMA 可能会出现由于 CACHE 与存储器不一致而引出错误数据。如果系统采用回写策略，该数据只在 CACHE 中，而 DMA 机制从存贮器中取出它的数据。为此软件在启动 DMA 前，必须保证一致性，也就是将数据写回主存。对于接收，必须置 CACHE 无效，从而保证一致性。刷新及置无效是费时的动作，并明显减少吞吐量。值得庆幸的是，目前已有许多 RISC CPU 提供 CACHE 一致性，以便用于多处理机系统。

②当采用 DMA 时，用户空间含有数据的页面必须锁定在存贮器中。而对于专用系统缓冲或用户与系统共享缓冲时，可发永久锁定，无需每个操作都锁定，从而减少开销。

③采用 DMA 时，数据异步拷贝，由于 CPU 与控制器必须在 DMA 结束后进行一次同步，通常是中断，这也是开销。

DMA 与 PIO 的比较，依赖于 I/O BUS 的特性及主机体系结构与软件。一般而言，长报文适合于 DMA，以得到高吞吐量，而短数据块则更适合于 PIO。

(2)大缓冲机制

硬件的第二种途径是设置大的缓冲区，以便存放多个报文，好处主要是：

①如果传输协议要求将检查和放在头部，那么可以在数据拷贝到控制以后插入检查

和。采用最小一个报文的缓冲区，就可以在拷贝数据的同时计算检查和。

②对于某些网络，控制器必须访问几个报文，以便根据网络条件调度报文传送。最简单的例子是基于开关的网络，报文调度由开关上可用出口驱动。这时，控制器需要访问几个不同目的地址的报文。

4 合理分划，提高整体效应

前面已谈到，虽然协议开销只是网络通信开销的一小部分，但是却经常作为开销的主要根源而加以分析、研究，提出了一系列措施，比较常用的方法是采用硬件技术或适配器技术加 CPU 技术来处理协议。

在设计智能 FDDI 控制器时，认真分析了硬件对协议本身处理的利与弊。对于完全独立的管理协议，如 SNMP 或 FDDI SMT 采用 CPU 方式加以支持，对于完整的传输协议 TCP/IP 进行合理的划分，并加以适当的外部支持。

传统的方法是在主机上完成全部的协议处理。最理想的方法是在网络控制器完成传输协议的处理，折衷的方法是网络控制器提供一些支持，如头部与数据的分离，协议处理由主机完成。

我们认为，在网络控制器上实现全部的协议处理并不合适，主要的不利因素是：其一，网络控制器复杂性增加；其次，必须支持多种协议，如 TCP/IP，DECNET，速度必须与主机速度相当，与主机的交互变得复杂；最后，有人分析过，除了检查和的计算外，传输协议的处理是很简单的，优化的协议处理可以减少到 200 条指令左右，这些程序移到网络控制器上并无明显的好处。为此，在 FDDI 控制器设计上，只考虑对数据检查和的计算及协议头部的分离的支持。

5 面向 FDDI 的特点

与传统网络相比，FDDI 有许多新的新点，如速度快，有多种传输要求及复杂的 SMT 管理协议等。在智能 FDDI 控制器上，除了考虑上述通用的策略外，还考虑了以下 FDDI 特点。

(1)多通道以适应多种服务

FDDI 支持同步、异步、受限、异步受限、立即服务等类别。对于同步传输，采用所分配的同步带宽传输一个或多个帧，对于异步传输，除了目标巡回旋转时间(NTTR)线索外，还提供可编程的异步奇偶线索。对于受限异步传输，支持受限对话的开始、继续及结束。对于立即传输，支持从 DATA，BEACOM CLAIM 状态发送帧，以及忽略响应所接受的字节流等功能。智能 FDDI 控制器中采用多通道体系结构，提供三个输入通路和两个输出通道以便独立与并发操作。由驱动程序分别配置，以便管理特殊帧的接收和传输(如同步帧)。

(2)32 位地址/数据的高带宽通路

网络控制器提供 32 位宽的同步数据接口，可以连接到标准多主设备的系统总线，局部总线运行频率为 12.5MHz 到 33MHz，采用 BIG 或 LITTLE ENDIAN 字节顺序，局部存储器可以是静态的或动态的。为了最大程度提高性能，采用阵发式传输，每次四或

八个 32bit 字。为了便于使用阵发式传输能力，阵发式传输周期内的三位地址作为非多路信息传输。最高速率为每个时钟 32bit 字，通过插入等待状态，可支持较低的速率。

网络控制器可与 CACHE、非 CACHE、页式、非页式存贮环境一起使用。为了支持这种能力，所有数据结构全部定义在页边界，网络控制器在边界块内完成所有总线事件，以消除对 CACHE 环境的接口。

(3)协议功能处理

网络控制器的协议功能处理主要是信息头与信息分离的支撑，为了支撑高性能协议处理，网络控制器分离(非 MAC/SMT)帧的头与信息部分。从帧控制字段(FC)到用户定义长度拷贝到指定的通道 1，其余的(信息)拷贝到指定的通道 2。这种处理对于从数据中分离协议头部，并存贮在不同的区之中，以防止不必要的拷贝是非常必要的。此外，协议监控应用可以只拷贝每帧的头部。

其次是桥接功能的支持，通过内部/外部分类模式提供桥接及监控应用。所有与外部地址(要求桥接的帧)匹配的帧分放在指定通道 2，匹配内部地址(环机制)的 MAC 及 SMT 帧分放在指定通道 0，所以其它符合内部地址(长或短)的帧分放在指定通道 1。

(4)簇中断方式

网络控制器可按查询或中断驱动方式操作。网络控制器提供在组边界中断(提供信号)。族边界由硬件预先定义，其它由用户在配置通道时定义。中断机制有效地减少了对主机的中断次数，从而降低主机处理的开销。

(5)独立 CPU，支持 SMT

FDDI 网络需要复杂的 SMT 协议处理支持，如果采用主机运行 SMT 协议，将给主机带来沉重的负荷，为此采用独立 CPU 支持 FDDI 和 SMT 管理，以提高整体性能。

6 小结

本文主要从硬件角度讨论了高性能网络控制器的设计，以及面向 FDDI 特点的 FDDI 控制器的设计技术。通过上述方法，提高了网络处理效率约 30%，同时减少了 CPU 的占用率，从而间接提高了应用效率。对于支持多媒体应用的网络协议软件处理技术，将另文论述。

参考文献

1 Metcalfe R. Computer/Network interface design:lessons from arpanet and ethernet. IEEE J. Selected areas comm, 1993. 11(2):173～180

2 Clark D, Jacobson V, Romkey. Analysis of TCP processing overhead. IEEE communications, 1989, 27(6):23～29

3 苏金树．支持多媒体及可视化计算的高速网络技术．计算机网络世界，1994，11