Fiber Channel 协议分析

赵文波，黄士坦
（西安微电子技术研究所，陕西 西安 710075）

摘 要：分析了光缆通道（Fiber Channel，FC）与通用的SCSI（小型计算机系统接口）的主要区别，以及光缆通道较SCSI的优点，光缆通道可以实现大容量、高速度、高可靠性的数据传输。分析了光缆通道的拓扑结构和各级层结构。

关键词：光缆通道；拓扑结构；层结构
中图分类号：TN919　　**文献标识码**：A　　**文章编号**：1673－629X(2006)12－0035－04

The Analysis of Protocol of Fiber Channel

ZHAO Wen-bo, HUANG Shi-tan
(Xi'an Institute of Microelectronics Technology, Xi'an 710075, China)

Abstract: In this paper, analyze the main difference between fiber channel and universal SCSI (small computer system interface), and the merit of fiber channel comparing with SCSI. Data transmission can be big capacity, high speed and high reliability by fiber channel. Also particularly analyze the topology and layer structure of fiber channel.

Key words: fiber channel; topology structure; layer structure

0 前 言

光缆通道是美国国家标准协会（ANSI）制定的一种串行数据接口协议，它是高性能的混合接口，是提供了必要的连通性、距离和多路传输协议网络特性的网络接口。FC支持通用协议，包括FDDI（光纤分布式数据接口）、PI（高效并行接口）、IPI（智能外围接口）、SCSI（小型计算机系统接口）、IP（因特网协议）、ATM（异步传输模式）等多种高级协议，可实现大容量、高速度、高可靠性和高效的信息传输，为实现计算机外部设备的高性能接口提供了实用和可扩展的数据交换标准[1]。

1 通道与网络

数据通信的两种基本方式是通道和网络。通道提供了一个直接的或点对点的连接。通道能高速低消耗地传输数据。相反，网络是分布式节点（象工作站、文件服务器或其他设备）的集合，网络通过节点各自的协议来支持节点的互连通讯。一个网络有相对高的消耗，因为它是软件支持的，所以网络传输慢于通道传输。但是网络比通道能操纵更多的任务。因为它们在一个不可预料的连接环境中操作，而通道仅在一些预先确定地址的设备之间操作。光缆通道结合两种通信方式组成新的I/O接口，以满足通道用户和网络用户的需要[2]。

2 光缆通道的特点

光缆通道的最大特性是将网络和设备的通信协议与传输物理介质隔离开，多种协议可在同一个物理连接上同时传送，高性能存储体和宽带网络使用单一I/O接口互连。同时，它还支持热插拔，可使系统的成本和复杂程度大大降低。它采用仲裁环式拓扑结构，一个用户级仲裁环可容纳4～12个用户稳定运行。如通过Switch（交换机）扩充至交换仲裁复用结构，则可将用户数扩大很多，它使用全双工串行通信原理传输数据。FC的最大数据传输速度为100M byte/s，双环可达200M byte/s，使用同轴线的传输距离为30m，使用单模光纤的传输距离可达10km以上。它与SCSI的比较见表1。

表1 SCSI与Fiber Channel的比较

	SCSI Wide (LVD)	Fibre Channel
带宽	40～80 Mbyte/s	100 Mbyte/s
连通性	15 devices	126 devices
附件	Ribbon cable, jumpers, power	SCA backplane: no jumpers, switches or power connections
距离	1.5 meters total length SE (single ended), 12 meters total length (LVD)	30 meters device to device (copper), 10 kilometers device to device (optical)
冗余性	Parity and running disparity	CRC protected frames

收稿日期：2006－02－21
作者简介：赵文波（1982－），男，山西大同人，硕士研究生，研究方向为高速串行总线接口技术；黄士坦，研究员，研究方向为空间计算机体系结构。

3 Fiber Channel 的拓扑结构

在光缆通道中,连接设备的开关叫做 Fabric。这种连接是通过两根单向的光纤以彼此相反的方向发送到与它们相连的发送器和接收器。每条光纤一端被分配给发送器的接口,另一端被分配给接收器的接口。当一个 Fabric 在一个结构中时,光纤就被分配给一个节点端口(N-Port)和一个 Fabric 端口(F-Port)。

因为光缆通道系统依靠的是可以互相登陆的端口和 Fabric,所以它与 Fabric 是否是一个电路开关,是否是一个激活的网络集线器或者一个环无关。根据系统的性能要求或包装选项,拓扑结构可以被选择。FC 拓扑结构包括点对点、交叉点开关或者决断环(见图 1)。

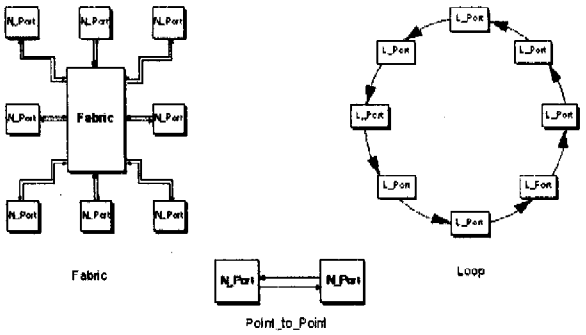

图 1 光缆通道拓扑结构

FC 可以以各种速度进行工作(133 Mbit/s,266 Mbit/s,530 Mbit/s,1 Gbit/s),并在三种光电介质上工作。传送距离由与它相关的速度和介质决定。用长波激光光源的单一模式的光纤介质能提供最高的性能(以 1 Gbit/s 传输最远距离达到了 10km)。

4 Fiber Channel 的层结构

FC 由 5 个功能不同的层次构成[3](见图 2)。

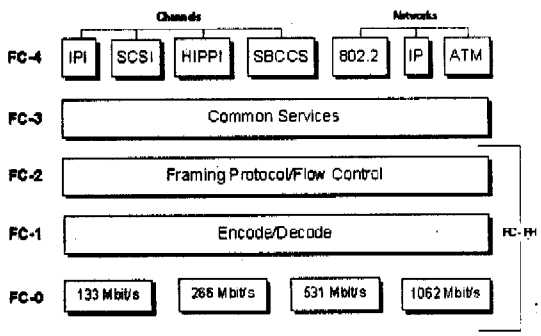

图 2 Fibre Channel 结构

4.1 FC-0 层

最底层(FC-0)在系统中定义了物理连接,包括光纤、连接器等的物理特性、传输速率和光电参数,图 3 说明了光缆通道的连接示意图。

图中,T 为 Optical Transmitter,R 为 Optical Receiver。

系统的位误码率(BER)在支持的介质和速度下低于 10^{-12}。终端对终端的通信路径可以由不同的连接技术完成,已达到最优性能和经济高效。

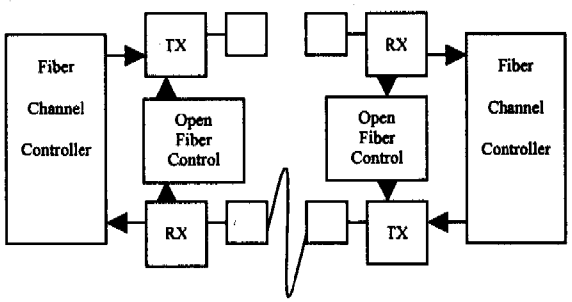

图 3 FC 连接

FC-0 指定了一个安全的系统——开放式的光纤控制系统(OFC)——状态字激光数据连接。因为光的能级远远超过了激光安全标准的限制,如果一个开放的光纤条件发生在连接中,被光纤连接的接收器端口探测它并且给它的激光以最低的占空比提供脉冲,以满足安全的要求。接收器的另一个端口(光纤的另一端)探测这个脉冲信号并以一个低的占空比驱动发送器。当这个开放的光纤路径被重新建立并两个端口接受脉冲信号时,两个握手过程之后,连接在短时间内被自动建立。

4.2 FC-1 层

FC-1 层定义了传输协议包括串行的编解码原理、特殊字符和错误控制。光纤上传输的信息被编码,一次将 8 位编码为 10 位的传输字符。传输编码必须是直流平衡来满足接受单元的电气要求。传输字符必须确保短的运行长度和在串行比特流时保证足够的发送,以便时钟恢复。

(1)FC-1 层的字符转化。

一个非编码的信息字节被编为 8 个信息位:A,B,C,D,E,F,G,H 和控制变量 Z。这个信息被 FC-1 编码为一个 10 位的传输字符中的位 a,b,c,d,e,i,f,g,h,j。控制变量可以是代表数据字符,为 D(D 型)或者是代表特殊字符,为 K(K 型)。每一个有效的传输字符被用以下的规范定义:Zxx.y,Z 位是非译码的 FC-1 信息字节的控制变量,xx 是由 E,D,C,B,A 位组成的二进制数的十进制值,y 是由 H,G 位组成的二进制数的十进制值,H,G 是非编码的 FC-1 信息字节。例如,由十六进制"BC"特殊码(K 型)组成的 FC-1 的传输字符的名字是 K28.5。接受的信息以每次 10 位的速度进行恢复,那些被用作数据(D 型)的传输字符被解码为 256 个 8 位的组合中的一个。一些剩下的传输字符(K 型)被看作是特殊字符,被用作协议的管理。被接收器探测到的既不是 D 型也不是 K 型的码被标记为孤立错误码。

(2)编码规则。

每个数据字节或者特殊字符有两个(不一定不同)传输码。数据字节和特殊字符被分别编码为哪种码,取决于最初的运行不一致(RD)。RD 是一个二进制参数,它计算一个传输字符中子块(头 6 位和最后 4 位)中的 1 和 0 的平衡。一个新的 RD 可从在发送器和接收器之间传送的字符得到。如探测的字符有相反的 RD,发送器应该被发送(取决于先前比特流的 RD),接收器反应为一个孤立的

不一致条件。一个传输字由 4 个相邻的传输字符组成。

4.3 FC－2 层

信号协议层(FC－2)是光缆通道传输的中枢。在端口之间转移的数据帧格式,控制 3 个服务级的不同的机制和控制数据转移时序的方法都被 FC－2 层定义。为了有利于数据传送,以下的块被标准定义:指令组、帧、时序、交换、协议。

(1)指令组。

指令组是包含数据和特殊字符的 4 字节的传输字。指令组使获得位和字同步可行,位和字同步建立了字边界队列。指令组总是开始于一个特殊字符 K28.5。指令组被信号协议定义了 3 种主要类型。

a.帧分隔符(帧开始(SOF)和帧结束(EOF)的设置)是指令组,它可立即领先或紧跟帧内容。为了 Fabric 和节点的时序控制,多个 SOF 和 EOF 分隔符可以被定义。

b.两个初始信号:空闲和接收器准备(R_RDY)是被标准指定的指令组,有特殊的意义。空闲是一个被发送的初始信号,用来告诉一个工作的端口设备准备传输和接受帧。R_RDY 初始信号用来说明接口缓冲器是可以接收更多的帧。

c.初始时序是一个不断被发送和重复的指令组,用来说明端口之间的特定状态。当一个初始时序被接收和识别,经响应后,一个相关的初始时序或者空闲被发送。一个初始时序的识别要求 3 个相同的指令组实例的连续探测。这种被标准支持的初始时序有:脱机的(OLS),非运行的(NOS),连接复位的(LR)和连接复位反应的(LRR)。

(2)帧。

一个基本的 FC 连接块是帧。帧包含了被发送的信息、源和目的端口的地址和连接控制的信息。帧被广义地分为数据帧和连接控制帧。数据帧被分为应答和连接反应帧(忙和拒绝)。Fabric 的初始功能是,接收来自源端口的帧和发送帧到目的端口。FC－2 层的作用就是将数据变为帧格式,并重新整合帧。

每个帧开始和结束都有一个帧分隔符(见图 4),帧头跟在 SOF 分隔符之后。帧头被用来控制连接应用,控制设备协议传输,并探测缺陷或者不整齐的帧。一个可选的帧头可以包含更多的连接控制信息。一个最大 2112 个字节的长区域(payload)包含了要从源节点转移到目标节点的信息。4 个字节的循环冗余码校验(CRC)之后是帧结束分隔符。CRC 被用来检测传输错误。

(3)时序。

时序是由一个或多个相关的、由一个节点向另一个节点单向传送的帧设置组成的。每个包含时序的帧被时序计数器唯一地计数。被一个高级的协议层控制的错误恢复经常发生在时序的边沿。

(4)交换。

一个操作交换是由一个或多个不并发的时序组成。交换在两个节点之间可以是单向的也可以是双向的。在单个交换中,在任何时候仅一个时序可以是激活的,但不同交换时,时序可以并发激活。

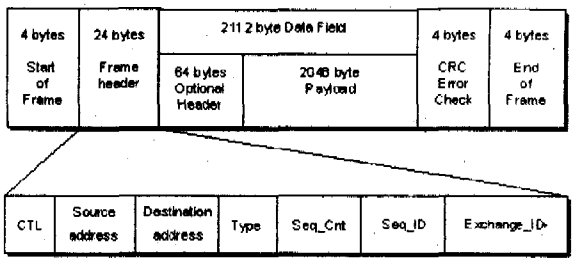

图 4　帧结构

(5)协议。

协议与光缆通道提供的服务有关。虽然协议对于高一层的服务可以是特定的,但是光缆通道提供了自己的协议设置,以管理数据转移的操作环境。以下是被标准指定的协议:

a.初始时序协议是基于初始时序的,并且为了连接错误而设定。

b.Fabric 登录协议:用于 fabric 一个节点的服务参数的互相交换。

c.节点登录协议:在数据转移执行以前,节点与另一个节点互相交换它的服务参数。

d.数据转移协议用光缆通道流控制来描述转移高层协议数据的方法。

e.当节点请求排除来自其它节点的服务参数时,节点注销协议被执行。这可用来在被连接节点上释放资源。

●流控制。

流控制是 FC－2 层的控制进程,用来配合节点间和节点与 Fabric 之间帧的流动,以防止接收器的溢出。流控制是由服务级别而定的。一级帧用于点对点的流控制,三级帧用于缓冲器对缓冲器的流控制,二级帧用于以上两种帧的流控制。

流控制被时序发生器(源)和用 Credit 和 Credit－CNT 的时序接收(目的地)端口管理。Credit 是分配给一个传输端口的缓冲器的数量。Credit－CNT 代表了没有被时序接受器应答的数据帧的数量。

点对点的流控制配合节点间的帧的流动。在这种情况下,时序接受器通过应答帧应答接受的有效数据帧。当接收缓冲器的数量对于引入的帧是不足时,一个"Busy"帧被发生器端口发出,当一个带错误的帧被接收,一个"Reject"帧将被发生器端口发出。时序发生器管理 EE－Credit－CNT。节点登录被用于建立 EE－Credit。

缓冲器对缓冲器的流控制在节点与 F－Port 之间或在点对点拓扑结构中的节点之间作用。每个端口都可以操作 BB－ Credit－CNT。BB－ Credit 在 Fabric 登录时被建立。时序接受端口(目的地)通过发送一个 Receiver－Ready 的初始信号给发送端口来响应它是否有空闲的缓冲器来接受到来的帧。

图 5～图 7 显示了不同服务级流控制的方法。

●服务级。

为了确保不同类型通信的高效传输,FC定义了三级服务。用户可以选择基于它们应用特性的服务级,象信息包的长度和传输延时,并可以通过Fabric登录协议分配服务。

一级是一种提供专用连接的服务,提供了一种相当于专门的物理连接。一旦建立,一级连接就被Fabric保障。这种服务保证了两个节点间最大的带宽,所以这是最好的支持高吞吐量的交易。在一级连接中,帧以它们被发送时的顺序传送到目的节点。图5显示了一级连接的控制操作。

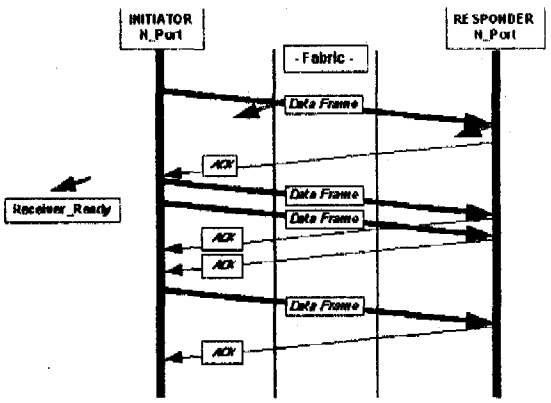

图5 一级流控制

二级是一种帧开关,是允许带宽被多路帧共享的无连接服务,这种多路帧是来自同一通道或不同通道的多个源。Fabric不能保证传送的顺序,帧可以被不按次序地传送。当连接建立时间大于一个短信息的反应时间的时候,这种服务级被使用。一级和二级发送应答帧响应帧的传送。如果由于阻塞传送不能完成,一个Busy帧被回复并且发送器重试。

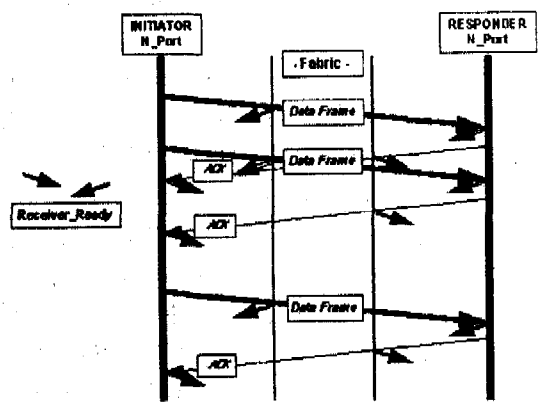

图6 二级流控制

除了帧传送不需要被批准外,三级服务与二级服务相同(流控制仅在缓冲器级被操作,见图7)。这种自带寻址信息的数据包的传送提供了最快的传输,因为它不用通过发送请求。这种服务对于实时广播是有用的,即在时间是关键的场合或这一时刻不被接收的信息是没用的场合。

FC标准还定义了一个可选的服务模式,叫做混合。混合是一级服务的一个选项,在一级连接中,帧保证了一个特殊的带宽量,而二级和三级中帧在通道上是多路的,仅当充足的带宽可以被分享连接时,才生效[2,4]。

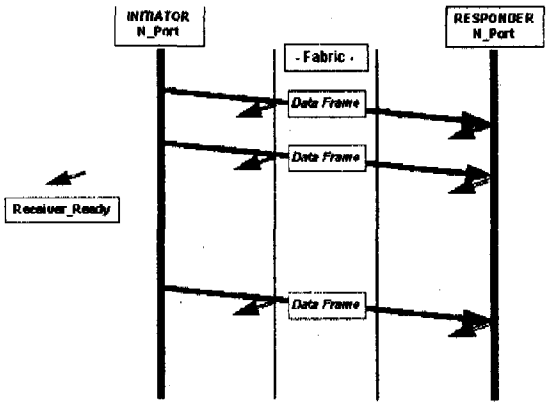

图7 三级流控制

4.4 FC-3层

FC-3层是公共服务层,被用来提供高级特性所要求的公共服务,例如:

(1)条块化-通过多路连接,并行多个节点来传输一个信息单元以增加带宽。

(2)搜索组-多节点来反应同一个地址的能力。通过减少到达一个忙节点的机会来提高效率。

(3)多点广播-发送一个信息到多个目的端口。这包括发送到Fabric上的所有的节点(广播)或仅到一个Fabric上节点的子集。

4.5 FC-4层

FC-4,FC结构中最高的层,定义了能执行光缆通道的应用接口。它用较低的FC层指定了较高层的协议的映射规则。光缆通道相当适合于网络和通道传输信息。而且它还允许各种协议在相同的物理接口上并行地传输。

以下是被FC-4层指定或提出的网络和通道协议[5]:

* Small Computer System Interface (SCSI);
* Intelligent Peripheral Interface (IPI);
* High Performance Parallel Interface (HIPPI) Framing Protocol;
* Internet Protocol (IP);
* ATM Adaptation Layer for computer data (AAL5);
* Link Encapsulation (FC-LE);
* Single Byte Command Code Set Mapping (SBCCS);
* IEEE 802.2。

5 结 论

随着多媒体技术、可视化技术的发展,实时通信的数据量非常之大,这样就要求计算机有非常快速的数据通信速率。然而,现有的计算机与I/O设备之间不能以期望的带宽进行通信。目前通用的SCSI(小型计算机系统接口),已不能满足增加的设备和距离连通性的要求。这样,传输速率的不足就将导致通信通道成为系统性能的瓶颈。

(下转第42页)

表2 模糊推理结果

u		$e_{(cpu)}$						
		BN	MN	SN	ZE	SP(0.5)	MP(0.5)	BP
$e_{(responsetime)}$	BN	0	0	0	0	0	0	0
	MN	0	0	0	0	0	0	0
	SN(0.2)	0	0	0	0	u_{ZE}	u_{SP}	0
	ZE(0.8)	0	0	0	0	u_{SP}	u_{SP}	0
	SP	0	0	0	0	0	0	0
	MP	0	0	0	0	0	0	0
	BP	0	0	0	0	0	0	0

(2)计算可信度。

在同一条规则内,前提之间通过"与"的关系得到规则结论。前提的可信度之间通过取小运算,得到每条规则总前提的可信度为:

R1 前提的可信度为 $\min(0.5,0.2)=0.2$

R2 前提的可信度为 $\min(0.5,0.8)=0.5$

R3 前提的可信度为 $\min(0.5,0.2)=0.2$

R4 前提的可信度为 $\min(0.5,0.8)=0.5$

将前提的可信度和表2进行"与"运算,得到规则的可信度如表3所示。

表3 规则可信度表

u		$e_{(cpu)}$						
		BN	MN	SN	ZE	SP(0.5)	MP(0.5)	BP
$e_{(responsetime)}$	BN	0	0	0	0	0	0	0
	MN	0	0	0	0	0	0	0
	SN(0.2)	0	0	0	0	$\min(u_{ZE},0.2)$	$\min(u_{SP},0.2)$	0
	ZE(0.8)	0	0	0	0	$\min(u_{SP},0.5)$	$\min(u_{SP},0.5)$	0
	SP	0	0	0	0	0	0	0
	MP	0	0	0	0	0	0	0
	BP	0	0	0	0	0	0	0

模糊系统总的可信度为各条规则可信度推理结果的并集,即

$u_{agg}=\max\{\min(u_{ZE},0.2),\min(u_{SP},0.2),\min(u_{SP},0.5),\min(u_{SP},0.5)\}=\max\{\min(u_{ZE},0.2),\min(u_{SP},0.5)\}$

可见实际触发了2条规则。

(3)逆模糊化。

在计算精确的输出值时,通常有三种方法:最大隶属度法、取中位法和加权平均法。最大隶属度法利用的信息比较少,会引起一定的不精确性;而取中位法充分利用模糊子集提供的信息,但是计算太复杂。这里采用普通加权平均法。

从上面的计算可知,模糊系统的输出 u_{agg} 实际上是两个规则R1和R2推理结果的并集。对应R1规则,输出变量隶属度0.2,对照图4(ZE部分)可以计算出两个输出值为 -8 和8。对应R2规则,输出变量隶属度0.5,对照图4(SP部分)可以计算出两个输出值为5和15。则输出量

$$u=\frac{0.2*(-8)+0.2*8+0.5*5+0.5*15}{0.2+0.2+0.5+0.5}=50/7$$

则在原来链接数的基础上,再增加7个。

2.3 模糊控制合理性分析

当CPU利用率比较轻松(50% - 27.5% = 22.5%),响应时间刚好满足客户需求(0.7 - 0.73 = - 0.03)的时候,下一时刻增加7个连接数,可以提高CPU利用率,但是增加的不是很多,这样就不会影响服务质量。通过模糊控制器计算出的这个值是很符合现实环境的。

3 结论及展望

至此,模糊控制器已设计完成。其工作原理是:从后端的执行服务器获取CPU利用率,从前端集中期得到服务响应时间,为了使这两个指标保持在一定的水平,通过决定执行服务器与前端服务器的连接数来控制。比如,当CPU利用率较低、响应速度很快时,说明执行服务器比较空闲,于是增加它的连接数,即加重下一时间片内的负载,从而提高CPU资源的利用率;而当CPU利用率较高、响应时间较长时,就减少连接数,于是减轻了下一时间片的任务,从而可以加快响应,保证服务质量。

文中设计的模糊控制器,取得了很好的性能表现。下一步研究中,将考虑引进神经网络以增加学习能力,同时可以吸取遗传算法的思想。

参考文献:

[1] 章文嵩.可伸缩网络服务的研究与实现[D].长沙:国防科技大学计算机学院,2000.

[2] 金士尧.主动式集群服务器总体设计.中国,02114011.1[P].2003.

[3] 王晓川.主动式集群网络服务器调度机制的研究[D].长沙:国防科技大学计算机学院,2001.

[4] Zadeh L A. Fuzzy Sets[J]. Information and Control,1965,8:33-35.

[5] Raju G V S, Zhou Jun. Hierarchical fuzzy control[R]. Athens, Ohio: ECE Department,Ohio University,1999.

(上接第38页)

而新的串行接口标准——光纤通道使用简单的点到点互连,提供了双口传输能力,使数据在两个独立的数据通路上传输,增加了故障冗余,提高了可靠性,同时降低了电缆连接的复杂程度。

参考文献:

[1] ANSI. X3T9.3 Task Group. Fibre Channel Physical and Signaling Interface (FC-PH), Rev. 4.2[M]. US:ANSI,1993.

[2] Fiber Channel Association. Fibre Channel: Connection to the Future[M]. [s.l.]:Fiber Channel Association,1994.

[3] 颜浩南,华晓红.光纤通道浅述[J].高性能计算技术,2004,167(2):27-30.

[4] Kessler G. Changing channels[J]. LAN Magazine,1993(12):69-78.

[5] Meggyesi Z. Fibre Channel Overview[J]. High Speed Interconnect,1994(8):15-18.