

基于分布式调度的 FC 核心交换电路设计与实现

王玉欢,田 泽,杨海波,李 攀,霍卫涛

(中国航空计算技术研究所,陕西 西安 710119)

摘 要:随着 FC 交换网络结构和规模日趋复杂和庞大,其对于 FC 交换网络核心芯片的功能、性能要求也越来越高。因此,FC 交换机芯片交换结构以及调度策略的选择已成为影响其效率、时延等关键指标的重要因素。文中对 FC 交换机芯片的交换结构及调度策略进行了研究,在对传统 FC 交换机芯片调度电路进行设计优化的基础上,提出一种基于交叉开关和输入输出端分布式调度的交换核心电路设计和实现方法。该电路适用于多端口交换机芯片设计,易于实现且能保证较高的交换性能。

关键词:FC;交换机;芯片;分布式调度

中图分类号:TP31

文献标识码:A

文章编号:1673-629X(2015)03-0214-04

doi:10.3969/j.issn.1673-629X.2015.03.049

Design and Implementation of Core FC Switching Circuit Based on Distributed Scheduling

WANG Yu-huan, TIAN Ze, YANG Hai-bo, LI Pan, HUO Wei-tao

(Aeronautical Computing Technique Research Institute, Xi'an 710119, China)

Abstract: As the structure and the size of the FC exchange network is increasingly complex and huge, the demand on the performance and function of the FC exchange network core chips is higher and higher. So the structure of FC chips and the choice of the scheduling policy has become the important factors of affecting the key indicators of efficiency and delay. In this paper, study on the structure and scheduling policies of FC chips, on the basis of the design optimization of traditional FC switch chips, come up with a method of design and implementation of exchange core circuit based on cross switches and input/output distributed scheduling. The circuit is suitable for design of multiport switch chips, it is easy to implement and assure high switching function.

Key words: FC; switch; chip; distributed scheduling

0 引 言

FC 交换机是以光纤作为传输介质,通过路由和地址管理等功能完成 FC 网络系统中各节点的信息交换的设备^[1]。根据 FC-SW 协议^[2-3],FC 交换机应具备一个或多个交换端口,具有可实现多帧交换或者电路交换的交换结构,能够以预定机制和速率将 FC 帧发送至它们的最终目的地址,能确定从一个源节点到一个目标节点的最佳路径选择机制,同时能够使每个交换单元在交换网络中都能具有相应的端口的标识;另外,FC 交换机还应该包含负责接收执行指令并产生响应,实现对内部各个交换单元的控制的控制器。

FC 交换机芯片是实现 FC 交换机功能的核心器件,它实现光纤通道网络多端口全双工核心无阻交换,

能够完成 FC 网络高速交换、数据监控、通信配置、时钟同步及网络管理等功能。随着 FC 交换网络结构和规模日趋复杂和庞大网络技术的发展,其对于 FC 交换机及其核心芯片的功能、性能要求也越来越高。其性能评价的关键要素如可同时连接的输入/输出端口数、端到端延迟、吞吐率等均需实现较高的指标,而 FC 交换机芯片选择何种交换结构以及调度策略对于上述性能参数影响重大。因此,FC 交换芯片核心电路的交换结构和调度策略的选择至关重要。

文中基于 FC 交换网络的特点,结合多端口 FC 交换机芯片的设计,针对传统包交换结构和集中式调度的弊端,提出了一种基于交叉开关和输入输出端分布式调度的交换核心电路设计和实现方法。首先对适用

于FC交换机芯片的交换结构进行了研究分析,在此基础上重点描述了基于交叉开关结构和分布式调度的FC交换机核心交换电路的设计和实现。文中的研究成果,为后续FC交换机芯片的成功投片奠定良好技术基础,对基于其他网络协议的交换机芯片实现也具有一定的借鉴意义。

1 FC 交换机芯片结构

FC交换机芯片端到端的时延是最重要的指标^[4-5]。因此,交换机芯片内部交换网采用何种结构是整个交换机芯片的关键所在。目前的主流交换结构

有共享总线式结构、共享缓存式结构和带VOQ(虚拟输出队列)的Crossbar结构和共享缓存等结构。

由于FC协议采用基于Buffer到Buffer信用管理的流控机制,即可以通过反馈机制来调整当前网络流量,使网络流量与网络可用资源相协调,是一种闭环控制算法的网络机构机制,因此比较适合采用Crossbar交换结构。

FC交换芯片结构的主要功能就是把源N端口的帧交换到各个目的N端口。每个N端口都有一个3字节的标识符,类似于以太网的MAC地址。图1给出了一个FC交换机芯片的简单协议构架^[6]。

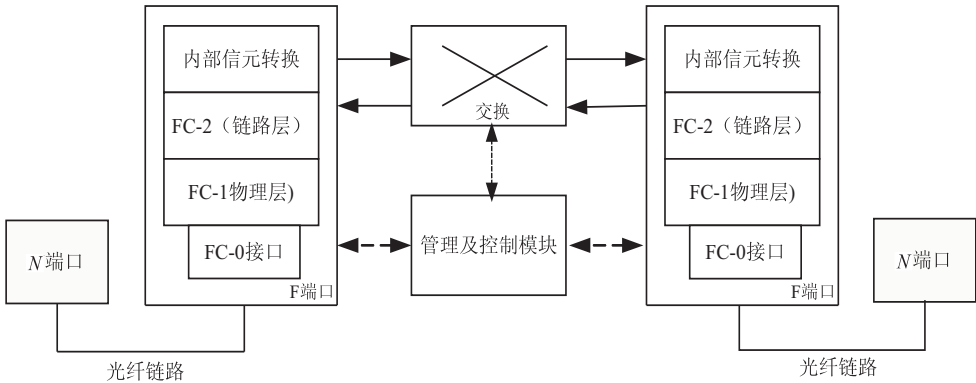


图1 FC 交换机芯片协议架构

交换端口在交换机中的线路接口卡(线卡)实现^[7-8]。线卡是构建FC交换机底层数据通道的核心部分,主要由光收发模块、物理层芯片、MAC处理、转发引擎和串行收发器几个部分组成。交换机芯片主要是实现时钟提取,串并转换,编码解码等FC-0、FC-1和FC-2层的功能。数据经过物理层后,被提交给MAC层处理模块,该模块负责光纤通道帧的定界,CRC校验,EOF检测以及更新和处理各种可能的链路服务命令等。之后,数据进入转发引擎模块,将数据通过串行收发器传送给交换网结构进行交换。

2 基于分布式调度的核心交换电路设计

文中设计和实现的FC交换机芯片核心电路如图2所示。包括FC MAC模块、FC帧接收控制及输入调

度模块、Crossbar模块、FC帧发送控制及输出调度模块等^[9]。

网络数据帧首先由通信端口的Serdes及PCS模块进行串并转换和解码,然后经FC MAC进入FC帧接收发送控制模块的帧接收缓冲区,在帧头进入时即提取其D_ID、优先级等信息,根据D_ID获知其目的端口,通过VOQ排队、轮询及优先级(2级)调度等手段,将此帧转发到对应FC交换机端口,再由PCS及Serdes模块编码、并串转换后发出。

FC MAC控制器为FC交换机的各个端口分别调用,每个端口分别连接FC MAC来实现FC网络数据通信。FC MAC实现FC-FS标准规定的FC-1和FC-2层的部分功能,如FC端口状态机,对数据帧进行解析、封装等功能。

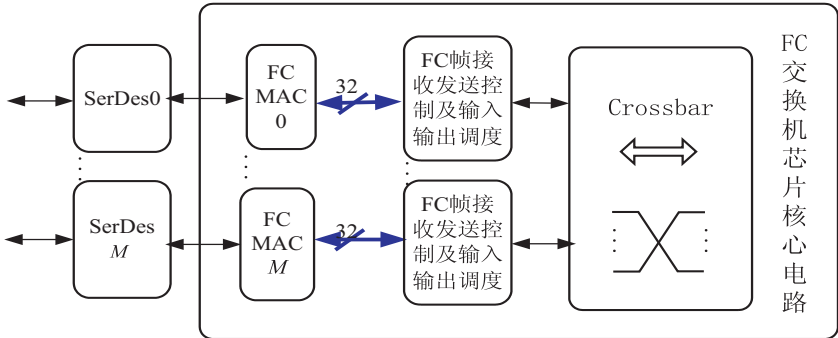


图2 FC 交换机芯片核心电路

FC 帧接收发送控制及输入输出调度模块又可分为帧接收控制及输入调度模块和帧发送控制及输出调度模块。FC 数据帧分别在输入和输出时进行分布式调度。帧接收控制及输入调度模块对接收到的数据帧进行排序和 VOQ 入队管理以及 VOQ 调度。帧发送控制及输出调度模块实现数据帧请求调度及输出等功能。

Crossbar 模块实现 M 路交换机端口之间用于数据交换的交叉开关通路的建立和关闭,由多个多路选择器构成。

2.1 FC MAC 模块

FC MAC 实现 FC-FS 标准规定的 FC-1 和 FC-2 层的部分功能,核心是 FC 端口状态机,对数据帧进行解析、封装等功能^[10-11]。

FC MAC 模块包括管理接口模块、信用管理模块、介质访问控制器 (MAC) 模块、链路控制单元模块、统计模块等。其中 MAC 是最主要的功能模块。主要实

现的功能有:端口状态机 (PSM)、帧控制和数据检测,包括 PSM、RXFSM、发送数据仲裁和发送数据选择。链路控制单元实现了 FC-FS 协议所描述的输入数据的字同步和对齐功能。这个模块同时提供帧发送时的 CRC 校验码。

2.2 FC 帧接收控制及输入调度模块

文中研究和设计的 FC 交换机采用分布式调度,即在输入和输出端分别调度;输入端采用时间窗 (固定时钟周期) 调度;输出端采用公平轮询与优先级相结合的方法^[12]。

FC 交换机所接收和发送的 FC 帧可划分为单播、多播及广播三种类型。帧数据从 FC MAC 输入到帧接收模块后,根据该帧的 D_ID 字段判断该帧属于单播、多播或是广播,通过 CS_CTL/Priority 字段、F_CTL 字段、OX_ID 字段得到其优先级、E 端口路由等信息,根据相关信息决定路由。FC 交换机支持的 FC 帧格式定义如图 3 所示。

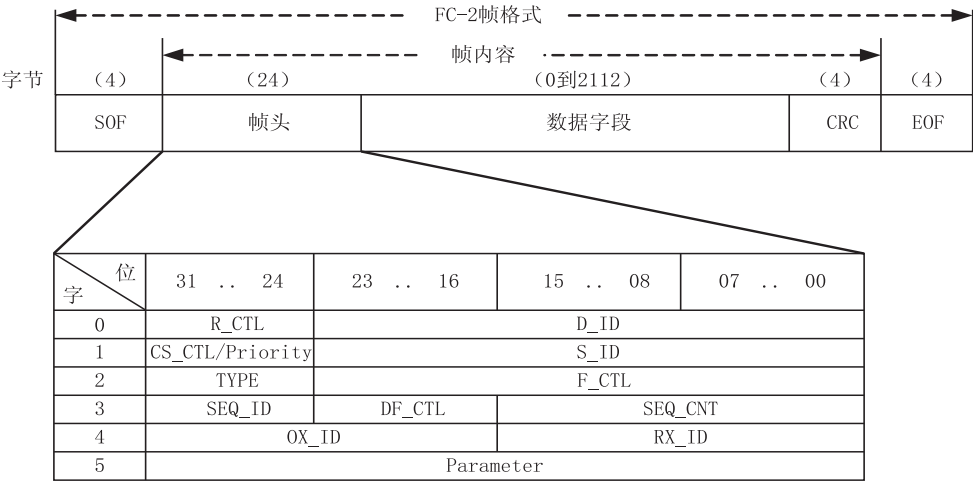


图 3 FC 帧格式

路由查找的 4 种情形如表 1 所示,通过对 FC 帧头的 D_ID 字段进行路由选择,可以得到帧路由的目的地址。

表 1 路由查找表

D_ID 字段	D_ID 字段的值	路由查找结果
D_ID[23:8]	等于多播预定义值	多播,查找多播配置表
D_ID[23:0]	广播预定义值	广播,查找广播掩码
D_ID[23:16]	等于单播预定义区段内值	根据 D_ID[23:0] 低位值查找单播路由
D_ID[23:0]	以上情形之外的其他值	丢弃

FC 帧接收控制及输入调度模块如图 4 所示。FC MAC 接收到新数据,将其写入当前可用的缓冲区。同时,提取得到 24 bit 的 D_ID,判断帧类型。若此帧为广播,优先级为最高。若为多播则根据 D_ID 查询多

播表,将查得的目的端口掩码写入多播 VOQ;若为单播,则根据其目的端口使其进入对应 VOQ。

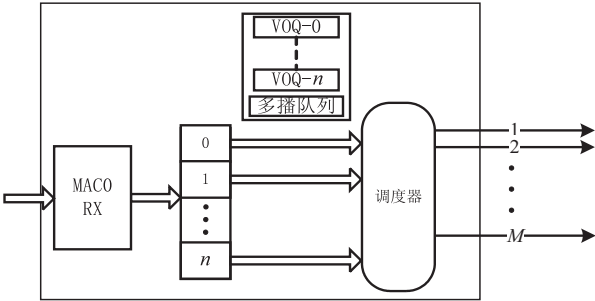


图 4 FC 帧接收控制及输入调度模块

2.3 FC 交换机交叉开关模块

每个输入端口可以通过 Crossbar 连接至 M 个输出端口,Crossbar 模块实现 M 路交换机端口之间用于数据交换的交叉开关通路的建立和关闭,由各端口数据线的连接及其多路选择开关构成^[13-14]。

Crossbar 在内部是无阻塞的,它可以支持所有端口同时线速交换数据。一个 Crossbar 只要同时闭合多个交叉节点(crosspoint),多个不同的端口就可以同时传输数据。

文中涉及到的 FC 交换机的 Crossbar 的交换结构组成如图 5 所示。

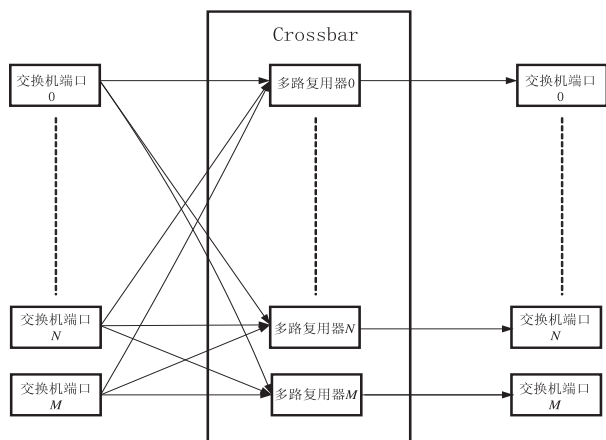


图 5 Crossbar 的交换结构组成

Crossbar 采用多个多路复用器实现,多路复用器的选通控制信号根据交换机端口的数据发送请求以及该多路复用器对应的交换机端口的状态生成,即当某一路端口发生请求且与其对应的目的端口对其调度有效时,多路复用器打开,该端口与多路复用器对应的通信端口建立数据通路连接。此时,建立连接端口的数据通过多路复用器发送给通信端口,从而实现其选择输出的功能。

2.4 FC 帧发送控制及输出调度模块

FC 帧发送控制及输出调度模块如图 6 所示。每个输出端口按优先级与公平轮询结合的方法调度。

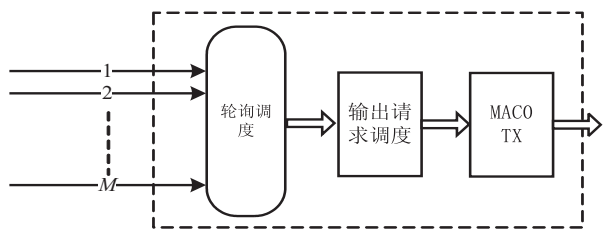


图 6 FC 帧发送控制及输出调度模块

有数据帧输出请求的时候,调度器按照请求优先级从高到低依次处理。如有高优先级请求,则查询其中是否有多播或广播请求,如有多播或广播请求,则向以公平轮询方式指向的当前高优先级多播或广播请求发送允许,如没有多播或广播请求,则向以公平轮询方式指向的当前高优先级单播请求发出允许;如没有高优先级请求,如存在一般优先级请求,则查询其中是否有多播或广播请求,如有多播或广播请求,则向以公平轮询方式指向的当前多播或广播请求发送允许,如没有多播或广播请求,则向以公平轮询方式指向的当前

高优先级单播请求发出允许。若被允许的输入端口确认连接,则数据传输开始。等待数据传输结束,释放允许,可以开始下一次调度。

3 结束语

文中通过对 FC 交换机芯片的结构和路由方式的研究,针对 FC 交换机芯片设计中的核心交换电路的设计及实现,提出一种基于交叉开关和输入输出端分布式调度的交换电路设计和实现方法,并详细介绍了该电路包含的子模块及其实现。功能仿真验证表明,该核心交换电路功能正确,满足芯片基本的交换功能需求。该核心电路的设计和实现对 FC 交换机的设计奠定了有效的基础,同时也对基于其他网络协议的交换机实现具有一定的借鉴意义。

参考文献:

- [1] T11/1861-D Rev 0.90, Fibre Channel Framing and Signaling-3 (FC-FS-3) [S]. US: American National Standards Institute, 2008.
- [2] 赵文波, 黄士坦. Fiber Channel 协议分析[J]. 计算机技术与发展, 2006, 16(12): 35-38.
- [3] 林强, 熊华钢, 张其善. 光纤通道综述[J]. 计算机应用研究, 2006, 23(2): 9-13.
- [4] NCITS/Project 1331-D fibre channel avionics environment [S]. New York: American National Standards Institute, Inc., 2002.
- [5] 徐亚军, 张晓林, 熊华钢. 航空电子系统 FC 交换式网络的可靠性研究[J]. 航空学报, 2007, 28(2): 402-406.
- [6] ANSL. Fiber Channel Physical and Signaling Interface (FC-PH), X3 [S]. USA: ANSI, 1994.
- [7] 田泽, 韩炜, 蔡叶芳, 等. 基于 FC 接口的 SoC 软硬件协同设计验证平台构建与实现[C]//第十三届计算机工程与工艺会议论文集. 西安: 西北工业大学出版社, 2009.
- [8] 杨海波, 田泽, 蔡叶芳, 等. 基于 FPGA 的多功能 FC 协议分析仪设计[J]. 计算机技术与发展, 2013, 23(7): 214-217.
- [9] 林强, 熊华钢, 张其善. 光纤通道及其交换型拓扑结构分析[J]. 计算机工程, 2004, 30(10): 93-95.
- [10] ANSI INCITS. Fibre Channel Switch Fabric (FC-SW) [S]. 1997.
- [11] ANSI INCITS. Fibre Channel Switch Fabric-5 (FC-SW-5) [S]. 2006.
- [12] 陈晴, 吴俊, 罗军舟. 一种适合于多播和单播的集成调度算法[J]. 计算机学报, 2004, 27(6): 758-764.
- [13] 谢军, 涂晓东, 孟中楼. 多用途光纤通道交换机的设计与实现[J]. 计算机研究与发展, 2011, 48(S): 335-339.
- [14] 黎小玉, 田泽, 王泉, 等. 基于 SoC_FC 芯片的电源管理系统设计与实现[J]. 计算机技术与发展, 2010, 20(8): 247-249.

基于分布式调度的FC核心交换电路设计与实现

作者:

[王玉欢](#), [田泽](#), [杨海波](#), [李攀](#), [霍卫涛](#), [WANG Yu-huan](#), [TIAN Ze](#), [YANG Hai-bo](#)
[, LI Pan](#), [HUO Wei-tao](#)

作者单位:

[中国航空计算技术研究所, 陕西 西安, 710119](#)

刊名:

[计算机技术与发展](#) 

英文刊名:

[Computer Technology and Development](#)

年, 卷(期):

[2015\(3\)](#)

引用本文格式: [王玉欢](#). [田泽](#). [杨海波](#). [李攀](#). [霍卫涛](#). [WANG Yu-huan](#). [TIAN Ze](#). [YANG Hai-bo](#). [LI Pan](#). [HUO Wei-tao](#) [基于分布式调度的FC核心交换电路设计与实现](#) [期刊论文] - [计算机技术与发展](#) 2015(3)