

用动态链路复用技术构建多业务交换体系结构

吴光斌¹, 李 军², 夏侯荔鹏²

(1. 深圳职业技术学院网络技术研究所 广东 深圳 518055; 2. 联想网络(深圳)有限公司 广东 深圳 518051)

【摘要】通过对交换体系结构及数据在交换芯片中的转发流程的深入分析,提出了一种用动态千兆链路复用技术构建多业务融合的交换体系结构的设计方法,解决了在传统的交换机中融合诸如防火墙等多种应用的硬件设计问题。该方法的原理是通过Cross-point 交换网络实现防火墙模块和业务板到交换板的链路共享,共享的数据链路在CPU的配置下,能动态地改变,从而实现多重并行数据传输。对样机的性能测试证明了设计方法的正确性。

关键词 动态链路复用技术; 多业务融合; 交换体系结构; 业务板; 交换板
中图分类号 TN919.72 文献标识码 A

A Multipurpose Switch Fabric with Dynamic Link Multiplex Technology

WU Guang-bin¹, LI Jun², XIAHOU Li-peng²

(1. Network Technology Research Institute, Shenzhen Polytechnic Shenzhen Guangdong 518055;

2. Lenovo Network (Shenzhen) Corporation Limited Shenzhen Guangdong 518051)

Abstract A new design method of building a multipurpose switch fabric with dynamic gigaspeed link multiplex technology is presented through analyzing switch fabric and the processing of frame forwarding in switch ASIC. The hardware designing problem how to combine multi-purpose such as firewall in the traditional switch is solved. The principle is that the firewall card and line cards share the link which link switch card through the cross-point switch. This link can be configed by CPU and changed dynamically, so it can carry out parallel communication. The results of tested pattern prove that this designing is correct.

Key words dynamic link multiplex technology; multipurpose; switch fabric; line card; switch card

网络应用越来越广泛的今天^[1],在交换机中实现更多业务融合的要求越来越强烈^[2],例如在交换机中融合防火墙功能,构成具有更高安全性的防火墙交换机,充分利用交换体系结构的高速数据交换提高交换机的性能^[3-4]。融合的系统还可以简化网络结构、减少设备数量、节省资金^[5]。

要实现多业务融合,可以采用动态千兆链路复用技术实现其他业务板与交换机的线卡共享高速交换板。本文主要介绍研制新型防火墙交换机的千兆链路复用技术。

1 交换体系结构分析

1.1 交换板硬件体系结构分析

交换板是实现各业务板之间数据转发的交换引擎,要达到“大容量、线速、低价”等目标,必须采用ASIC技术。下面以最新研制的Szpt4500防火墙交换机为例对交换板体系结构进行分析。

Szpt4500交换板以交换芯片为核心构成,支持12个Gigabit Channel和1个12 Gbps扩展接口。交换板可通过4个Gigabit Channel分别与每个业务板相连,任何两个业务板之间通过交换板实现高带宽无阻塞通道。

交换板的主要接口有面板接口、CPU接口、扩展链路接口、千兆链路接口、业务板控制接口、风扇和电源控制接口。交换板安装在机箱后部,交换板面板上提供电源指示灯、系统诊断指示灯、系统报警指示灯、复位按钮和CONSOLE接口。

1.2 数据包的转发流程分析

业务板源端口的数据选择一条Gigabit Channel到达交换板,经交换芯片完成数据的交换和转发,通过交换板的Gigabit Channel到达目的端口。数据包在交换芯片内按Ingress MMU Egress顺序转发,进入交换芯片的以太网数据包,在MAC层执行IEEE 802.3兼容处理,包括校正数据包的格式、检查数据包长、校验数据包字节对齐、检验数据包是否错误、更新统计结果、接收特定的控制包,以及其他的转换,直到被MAC完全接受送到Ingress Logic。Ingress Logic检查从MAC过来的数据包,以确定是否执行进一步的处理。对于Pause帧或控制帧不需要进一步处理。数据包从Ingress Logic进入MMU,缓存在CBP中,创建处理队列(XQ)的入口信息,通过数据包的传输调度,实现交换机先进的CoS功能。Egress Logic负责完成数据包的传送。一旦Egress Logic处于空闲状态,就向MMU请求数据包。如果MMU中有数据包需要传送,MMU就将数据包传给Egress Logic,并指示传送的数据包是否需要打上Tagged标记。

由以上分析可知,数据交换的主要流程中,对数据包的处理是在交换芯片中完成的,因此在数据转发流程中,线卡的所有端口在本地不能交换,必须通过特定的链路到达交换板,通过交换板进行集中交换,再经链路到达目的端口。因此,要实现交换线卡业务板和其他业务板(本文中特指防火墙业务板)在同一交换体系中融合,必须对链路采用动态复用技术。

2 千兆链路复用技术的设计

在Szpt4500防火墙交换机的研制中,为了实现嵌入式防火墙业务板与交换线卡业务板的多业务融合,使Szpt4500安全交换机在实现三层线速交换的同时,集成嵌入式的防火墙模块,实现完整的、高性能的企业级防火墙功能,需要硬件、软件、结构设计、配电等的配合。采用一体化设计,最关键的是设计一种动态的千兆链路复用技术,以实现多业务的高速交换。

2.1 千兆链路复用技术的设计原理

为了充分利用交换板的交换能力和背板带宽,必须将分配给防火墙模块的链路与千兆线卡业务板实现共享,根据用户实际使用情况,自动将带宽分配给防火墙板或其他业务板。

系统通过Cross-point交换网络实现防火墙模块和业务板到交换板的链路共享。共享的数据链路在CPU的配置下,能动态地改变,从而实现多重并行数据传输。每一个连接都是一个直接的点到点的数据通路,允许更高的工作频率和更宽的带宽。

交换体系结构和数据包的转发设计采用集中式的点对点交换架构,即中央交换板集中交换路由,业务板实现端口扩展,防火墙板通过交换板实现与业务板的数据交换,任何两个业务板之间通过交换板实现高带宽无阻塞通道。这样的设计可以充分利用千兆链路复用技术的高速带宽。

2.2 防火墙和业务板的状态检测

要实现防火墙业务板和交换线卡业务板之间的高速链路复用,交换体系结构必须能够对它们进行状态监测和统一监控。这就要求对基于传统计算机架构(本系统中为Intel架构)的嵌入式防火墙业务板的硬件设计进行改造,使之能够接收和识别交换体系结构的控制信号,这是链路复用技术在多业务板之间实现带宽复用的重要前提。在本文所述的Szpt4500安全交换机的研制中,对防火墙业务板的接口进行了专门的设计。

状态检测通过交换体结构中的CPU来完成。CPU通过控制总线对防火墙和业务板的状态进行监测。根据链路的运行状态和当前参数,如通断情况、吞吐量等,可以做出正确的调度决策。控制器检测防火墙模块是否在位,并根据检测结果,决定是否给出控制信号,将交换板的千兆通道由线卡切换给防火墙模块。通过这种链路的自动切换和共享,可实现防火墙与交换系统的无缝集成。

2.3 链路的管理和调度

在防火墙和业务板的链路同时使用共享链路时,对物理层的操作将产生冲突,必须对链路的数据输入进行串行化(排队)处理,保证在同一时刻只处理一个输入。因此需要对链路进行管理和调度。

链路管理和调度必须维护一张链路表,以记录链路的情况。链路表包括链路的标识、链路的优先级、链路的空闲状态等信息。在链路调度时,首先激活优先级高的链路。若优先级高的链路无法激活,则激活优先级低的链路,直到激活连接成功。若激活了低优先级链路,则保持高优先级链路的“试图连接状态”。

若均无法激活,则停止主动激活,等待被动激活。在所有激活的链路上,链路管理和调度单元主动发出“心跳信号”,侦测链路状态,包括链路是否中断、心跳信号回应时间等,并实时反映到链路表上。

2.4 链路的切换

在防火墙(或业务板)与交换板通信的过程中,当优先级高的主用链路断链时,必须自动切换到优先级低的链路上进行数据通信,称为自适应切换。

在自动切换模式下,切换原则是保证高优先级链路进行通信。下面分5种情况进行讨论:(1)交换板或业务板交给链路层待发送消息,此消息结构中包含了目的链路ID。链路调度通过查询链路表,获得此ID的链路优先级,优先级高的链路中进行通信。(2)当正在通信的链路中断时,保持该链路的试图连接状态。激活优先级低的链路。如果没有链路可以被激活,则认为通信中断;若激活优先级低的链路,则从未被ACK的消息开始发送。(3)在通信中,优先级高的链路被激活时,中断正在使用的链路,转移到优先级高的链路上,从未被ACK的消息开始发送。(4)由心跳信号侦测得知,在通信中的链路中断时,保持该链路的试图连接状态,激活优先级低的链路。如果没有链路可以被激活,则认为通信中断。(5)在通信中,优先级高的链路被激活时,中断优先级低的链路。

3 防火墙与交换机多业务融合的实现

通过千兆动态链路复用技术,能够实现防火墙与核心交换机的有机整合,构成高度集成化、模块化的独立网络安全交换设备。

3.1 防火墙业务板与线卡的千兆动态链路复用实现

防火墙模块通过千兆NIC实现4个千兆以太网端口,其中的3个被引到面板提供给用户,1个作为千兆链路和背板相连,与线卡共享交换板的千兆通道。控制器检测防火墙模块是否在位,并根据检测结果,决定是否给出控制信号,将交换板的千兆通道由线卡切换给防火墙模块。通过这种链路的自动切换和共享,实现了防火墙与交换系统的无缝集成。交换系统以交换芯片ASIC为核心构成,支持12个千兆通道和1个扩展接口。从交换板分别通过4个千兆通道与每个线卡插槽相连,任何两个线卡模块之间通过交换板实现高带宽无阻塞通道,其中一个线卡插槽与防火墙模块复用一条千兆通道,交换板通过监测防火墙模块是否在位,控制高速交叉开关选择将某条通道分配给线卡或者防火墙模块。

3.2 系统的交换体系结构设计

本系统采用“中央交换板集中交换路由,线卡实现端口扩展,防火墙板通过交换板实现与线卡数据交换”的体系结构,如图1所示。

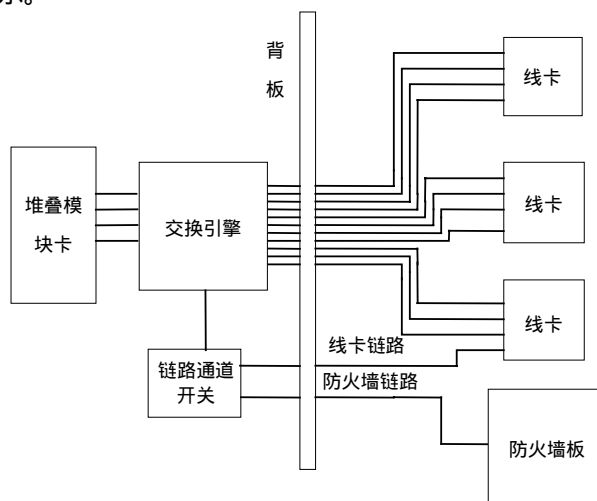


图1 防火墙交换机系统体系结构图

在图1中,线卡通过4个千兆通道与交换板相连,防火墙板通过链路开关与线卡实现链路共享,并由此经交换板实现与线卡的互通。扩展模块通过4个3.125 Gbps链路 with 交换板相连,将各线卡的数据汇聚后向上

传送。防火墙板与交换线卡共享32 G的高速交换背板。高速背板主要实现交换板和业务板之间的信号连接,以及提供交换板、业务板、扩展模块、防火墙模块、风扇模块所需的直流电源。背板的信号不仅包括交换板对各业务板的控制信号,还包括交换板与各业务板的高速数据链路通道。

4 性能测试与分析

在多业务融合的交换体系中,非交换线卡的业务板利用高速背板和交换板的高速数据转发,性能比在传统的计算机架构中大大提高,且不影响原有的交换体系的性能。为了验证这一点,必须对研制的样机进行严格的测试。

4.1 测试方法

为了验证在多业务融合时的系统性能,需要测试:(1)多业务工作状态下交换机的性能;(2)多业务工作状态下防火墙业务板的性能。

所有测试都使用Spirent Communications SmartBits6000C测试设备和Spirent Communications SmartFlow软件生成和测量网络流量。每项测试分别采用了64、128、256、512、1 024和1 518字节的分组;针对每种分组大小的测试都要进行两次,以验证结果是否一致。

4.2 测试结果与分析

在多业务工作状态下对交换机的性能测试,采用全密度全网状的配置,在Full-Mesh模式下测试所有12个千兆端口的吞吐量,测试结果见表1。

从测试结果可以看出,在最严格的全密度和全网状配置下,每个千兆端口在针对每种大小的帧的二层转发测试中都实现了线速的吞吐量,帧丢失率为零,表现出很高的交换性能。

在多业务工作状态下对防火墙业务板的性能测试,重点测试在防火墙的路由工作模式下,配置单条防火墙规则和多条防火墙规则时发送各种大小分组时的吞吐率,吞吐率测试结果见表2。

表1 千兆以太网端口的第二层吞吐量测试结果

帧长/字节	每秒发出的帧	每秒收到的帧	每秒丢失的帧	帧丢失率/(%)
128	10 135 140	10 135 140	0	0
256	5 434 788	5 434 788	0	0
512	2 819 544	2 819 544	0	0
1 024	1 436 779	1 436 779	0	0
1 518	975 283	975 283	0	0

表2 防火墙业务板吞吐率测试结果

帧长/字节	吞吐率/(%)		
	1条规则	100条规则	200条规则
64	35.07	31.58	30.94
128	59.80	58.85	53.72
256	99.46	95.50	94.68
512	99.91	99.91	99.91
1 024	100.00	100.00	100.00
1 280	100.00	100.00	100.00
1 518	100.00	100.00	100.00

从以上的测试结果可以看到,在多业务融合的工作状态下,防火墙业务板和交换业务板都表现出了很高的性能,证明千兆链路复用技术的设计是正确的,完全达到了设计目标。

参 考 文 献

- [1] 郭向勇, 吴光斌, 赵怡滨. 千兆以太网组网技术[M]. 北京: 电子工业出版社, 2002.
- [2] SafeNet Inc. Security trends in the networking industry[C]//4th Embedded System Conference-Asia, Hsinchu, Taiwan, 2004.
- [3] Ahmadi H, Denzel W. A survey of modern high performance switching techniques[J]. IEEE J. Selected Areas Comm, 1989, (7): 1091-1103.
- [4] Oie Y, Suda T, Murata M, et al. Survey of switching techniques in high-speed networks and their performance[J]. IEEE Proceedings of IEEE Infocom., 1990, (6): 1242-1251.
- [5] SRIDHAR T. The Evolving Ethernet Switching Landscape[C]. 4th Embedded System Conference-Asia, Hsinchu Taiwan 2004.

编 辑 熊思亮