第 35 卷 第 5 期 计 算 机 工 程 2009 年 3 月
Vol.35 No.5 Computer Engineering March 2009
·网络与通信· 文章编号：1000—3428(2009)05—0117—02 文献标识码：A 中图分类号：TP273

路由器高效并行存储访问机制

王逸欣[1]，吴纯青[2]
(1. 南方医科大学网络中心，广州 510515；2. 国防科技大学计算机学院，长沙 410073)

摘 要：访存速率远滞后于传输速率的发展，成为影响路由器性能的最大瓶颈。该文提出一种高效并行存储设计方案，通过简洁的软硬件协同技术实现，硬件实现 2 个同样大小的输出端缓冲区，消除由于单个存储接口可能引起的系统瓶颈，软件设计为成对存储信元的结构。试验结果表明，该机制有效提高了存储访问速度和路由器的吞吐率。

关键词：存储带宽；并行存储机制；路由器

Efficient Parallel Storage Access Mechanism of Routers

WANG Yi-xin[1], WU Chun-qing[2]
(1. Network Center, Southern Medical University, Guangzhou 510515;
2. Computer School, National University of Defense Technology, Changsha 410073)

【Abstract】As the progress of storage access speed is much slower than the communication transport rate, the storage bandwidth becomes the bottleneck of the router's performance. This paper proposes a high efficient parallel storage access mechanism implemented by a simply approach combining hardware and software. Two same size output buffers are designed for alleviating the bottleneck caused by single storage interface. A paired access structure is designed by software to take use of the two same size output buffers parallely. The test results indicate that this mechanism effectively increases the router's throughput and storage access speed.

【Key words】storage bandwidth; parallel storage mechanism; router

1 概述

路由器是互联网的核心设备，路由器在很大程度上决定了网络的性能和功能。网络应用的发展对路由器的性能提出了很高的要求。近年来的研究成果和技术的发展使路由器的性能获得了很大的提高。在传输领域，以密集波分多路复用(dense wavelength division multiplexing)为代表的光纤通信技术得到了飞速发展，光纤传输带宽在 1995 年后以约每 7 个月提高 1 倍的速度增长[1]。目前，1.6 Tb/s(160×10 Gb/s)的 DWDM 系统进入商用阶段，10.9 Tb/s(273×40 Gb/s)的系统已完成长距离传输试验[2]。可以认为，传输链路不是提高网络性能的瓶颈。

相比传输速率的发展，存储器访问速率的发展就要慢得多。存储器的访存速度每 18 个月仅提高 1.1 倍[3]，这与网络流量每年增长 1 倍的发展速度形成较大的落差，也成为提高路由器性能的最大瓶颈。

2 相关研究

在提高路由器存储带宽方面，人们已提出了许多好的设计思想，并取得了一定的成效。

2.1 多级存储结构

报文缓冲速度直接影响高性能路由器的处理能力。对共享内存交换单元更是如此，其内存操作速度必须几倍于线卡速度。随着链路速度增加，问题变得更加严重。

文献[4]提出了交换单元或路由器报文缓冲的多级存储结构，它由速度较慢、成本较低但容量大的若干个并行的 DRAMs 和速度快但容量较小的 SRAM 组成，所有 DRAM 由一条地址总线控制，从而可以完成多个信元的并行操作。

Iyer S 等人证明了用多级存储结构构造的报文缓冲能够达到相当好的性能。

2.2 乒乓缓冲结构

高速交换和路由器的设计通常受限于内存带宽。文献[5]分析了商用内存的结构和工作原理，即使用单端口内存，在一个时间槽内完成读和写 2 个操作，Joo Y 等人认为，如果使用双端口内存，每个端口分别完成读和写操作，则可将访存速度提高 1 倍，因此，提出了一种乒乓缓冲机制以提高存储带宽。

乒乓缓冲结构是一种加倍内存带宽的技术。它的优越性很明显，即使用传统存储设备，通过限制每个内存在单个时间槽内只进行一种内存操作，使得缓冲区的整体操作速度加倍；或者对于给定的速度，它可以使用速度较慢的较低成本的内存。但是这种结构会浪费一部分内存，在最坏情况下，有一半内存会浪费掉。

另外由于 2 个 memory 的占有率的差异性，溢出速率也会增加，虽然可以通过附加内存解决，但这无疑增加了路由器的成本。

3 并行存储访问机制

3.1 并行存储设计

由于访存速率的发展远远滞后于传输速率的发展，成为影响路由器性能的最大瓶颈。在研究路由器的高效 QoS 机制时，如何提高系统的整体访存速率仍然是不可回避的问题。

作者简介：王逸欣(1975－)，男，工程师，主研方向：自动控制；吴纯青，研究员

收稿日期：2008-10-13 **E-mail：**xixiwu2001@yahoo.com.cn

多级存储结构和乒乓结构的设计主要是依靠并行存储访问技术提高系统的整体访存速率。但这些结构设计复杂，同时软件也需要设计复杂的访问机制，实现难度较大。本文提出一种高效并行存储设计方案，通过简洁的软硬件协同技术实现，达到了提高路由器吞吐率的目的。

在硬件方面，设计实现了 2 个同样大小的输出端缓冲区，即 EB0, EB1(Egress Buffers)。EB0 和 EB1 的大小和地址编址完全相同，但它们是 2 个独立的存储器，其用途不是增加输出端的报文缓冲容量，而是专门用于并行存储访问，从而消除由于单个存储接口可能引起的系统瓶颈。

在软件方面，将存储结构设计成成对存储信元的结构 SCP(Store Cells in Pairs)，如图 1 所示。

图 1　SCP 结构

一个信元 54 Byte(包括信元头 6 Byte)，加上帧头 10 Byte，共 64 Byte。因此，每个 SCP 为 128 Byte 的连续存储单元，可以存放 2 个完整的信元，前 64 Byte 称为 F_SCP，后 64 Byte 称为 B_SCP。F_SCP 的头 6 Byte 未使用，B_SCP 的头 6 Byte 用于存放下一个 SCP 的首地址，使得同一个帧的 SCP 形成一个链，从而将信元重组为帧。

输出端缓冲区被划分成多个 SCP，SCP 的个数由 EB0 或 EB1 的大小确定，如一个 64 MB 的 EB0 或 EB1 有 512 000 个 SCP。

输出缓冲区的申请和释放由一个先进先出队列结构 Free_SCP 来管理。由于 EB0 和 EB1 总是成对使用，因此只需一个 Free_SCP。如图 2 所示。

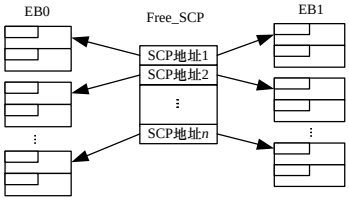

图 2　输出缓冲区管理结构

当申请一个 SCP 时，从队列头获得一个空闲 SCP 块，当释放一个 SCP 时，则将其加入到队列尾。

前面提到每个 SCP 可以存放 2 个从交换单元到来的信元，为了提高系统存储带宽，利于并行存储访问的硬件实现，从交换单元到来的信元分别按到达顺序存放在 EB0 和 EB1 中，如图 3 所示。其中，C_hd 为信元头；T_hd 为帧头；F_data 为帧数据；N_SCP 为下一个 SCP 首地址。图中 1 个数据长度为 120 Byte 的帧由 3 个信元组成，信元 1 存放在 EB0 的 F_SCP 中，信元 2 存放在 EB1 的 B_SCP 中，EB0 的 F_SCP 和 EB1 的 B_SCP 构成了一对 SCP，该帧由 2 个 SCP 组成，由于 EB0 和 EB1 的编址完全相同，因此只需设计一个 N_SCP。在输出端 QoS 处理过程中，读取帧数据通过硬件并行访存部件一次读取这样一对信元，从而提高了系统的处理速度。

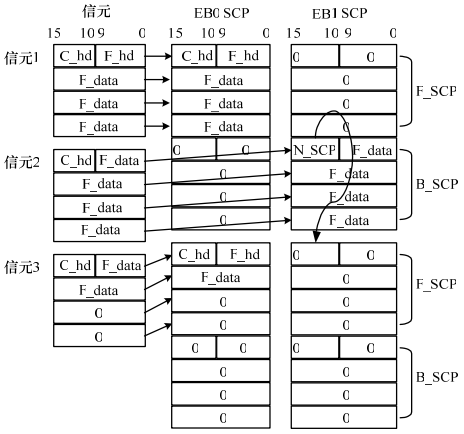

图 3　帧的信元在 SCP 中的存放方式

3.2　测试结果

用网络性能及协议测试仪 SmartBits6000(以下简称 SMB)的 2 个千兆模块与路由器千兆板相连，如图 4 所示。

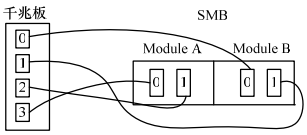

图 4　实验环境

由 SMB 的 4 个千兆口同时分别发送一百万个报文，并将路由器 4 个千兆口设为双向工作状态(同时收发)，在发送端带宽占用率为 100%, 50%, 5%时，记录 4 个千兆口收到的报文数。第 1 个实验将从交换单元到来的信元仅存放在 EB0 中，测试各端口接收到的报文数；第 2 个实验将从交换单元到来的信元按照并行存储访问的方式存储在 EB0 和 EB1 中，测试各端口接收到的报文数。图 5 是发送端带宽占用率为 100%时 2 种模式下各端口接收到报文数比较(深色为并行模式，浅色为非并行模式)。

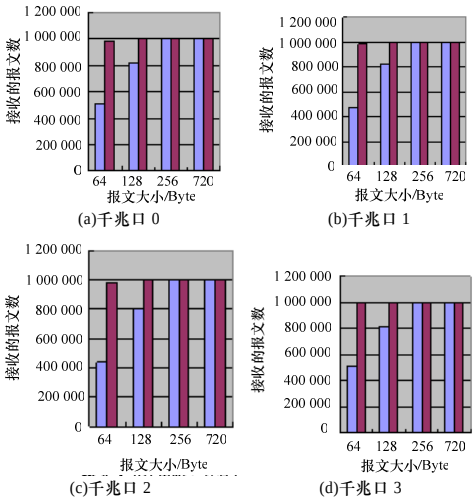

图 5　2 种模式下各端口接收到报文数比较

(下转第 130 页)

删除的内容:

(上接第 118 页)

4　结束语

本文针对路由器存储访问速度瓶颈问题，提出了一种基于软硬件协同技术的并行存储访问机制。实验结果表明，该机制能够有效提高访问速度，进而提高路由器的吞吐率。下一步将在高性能路由器研制中实际应用这一机制，进一步验证其有效性。

参考文献

[1] Andrew M. Internet Traffic Growth: Sources and Implications[C]// Proc. of Optical Transmission Systems and Equipment for WDM Networking II. Orlando, FL, USA: SPIE Press, 2003: 1-15.

[2] 杨　壮, 杨　名. 波分复用系统的发展和面临的挑战[J]. 通讯世界, 2002, 8(12): 52-54.

[3] Patterson D, Hennessy J. Computer Architecture: A Quantitative Approach[M]. 3rd ed. San Francisco, USA: Morgan-Kauffmann Press, 2002.

[4] Iyer S, Kompella R R, McKeown N. Analysis of a Memory Architecture for Fast Packet Buffers[C]//Proc. of IEEE Workshop on High Performance Switching and Routing. Dallas, TX, USA: IEEE Press, 2001: 368-373.

[5] Joo Y, McKeown N. Doubling Memory Bandwidth for Network Buffers[C]//Proc. of the IEEE INFOCOM'98. [S. l.]: IEEE Press, 1998: 808-815.

编辑　任吉慧

分节符(连续)