核心路由器中业务引擎交换吞吐量研究

孙长华, 刘 斌, 李文杰
(清华大学计算机科学与技术系, 北京 100084)

摘 要: 本文给出了核心路由器中业务引擎吞吐量分析的一般模型, 并应用此模型研究了实际设计的用于 OC-48c 接口的业务引擎. 由此得出: 当采用 GigaStream 交换芯片时, 每路 CSIX(Common Switch Interface) 接口至少需用 4 路 HSSL(High Speed Serial Link) 链路; 在 CSIX 接口处形成 CFrame 帧时需要对 IP 包最后一个信元按照实际大小形成 CFrame 帧, 才能保证交换网络与 CSIX 接口处线速转发所有长度的 IP 包. 最后, 本文给出了业务引擎吞吐量和 IP 包存储之间的一般性关系, 该关系可用于指导核心路由器交换网络的设计.

关键词: 吞吐量; GigaStream 交换芯片; CSIX 接口; 业务引擎
中图分类号: TP393 文献标识码: A 文章编号: 0372-2112 (2005) 07-1242-05

Research on Switching Throughput of Traffic Manager in Core Routers

SUN Chang-hua, LIU Bin, LI Wen-jie
(*Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China*)

Abstract: We made a general model to analyze switching throughput of traffic manager in core routers. By designing a real traffic manager which uses the OC-48c interface, we analyze the whole system and point out that at least four HSSLs should be employed per CSIX interface when using Vitesse's GigaStream switch chip set. Meanwhile, at the CSIX interface, the CFrame should be constructed according to the actual size of the last cell of each IP packet. Above principles can guarantee forwarding IP packets at line rate. We give a general relationship between throughput and buffering scheme of IP packets in the external memory, which is useful in the design of switch fabric in core routers.

Key words: throughput; GigaStream; CSIX interface; traffic manager

1 引言

核心路由器中广泛使用交换结构(Switch Fabric)来取代共享总线与共享内存, 交换结构与业务引擎(Traffic Manager)之间一般采用 CSIX 接口[1](Common Switch Interface). CSIX-L1 规范规定了交换结构与业务引擎之间的数据交换格式和流量控制等信息, 它使用变长 CFrame 作为数据格式, 同时要求交换结构内部采用定长信元进行处理. 变长 CFrame 的最大长度取决于交换结构内部定长信元的长度[2], 这样, 数据格式的转换以及可能的填充会浪费传输带宽, 影响路由器的吞吐量. 文献[3~5]都讨论过这个问题, 但仅仅给出了定性的描述与分析, 设计路由器时, 需要准确分析它们对交换吞吐量的影响. 另外, IP 包的存储效率也是影响路由器吞吐量性能的一个重要因素, 它主要涉及到 IP 包在外存中如何缓存. 本文将对影响交换吞吐量这两个因素进行分析与建模, 给出一般性的结论.

图 1 是实际设计的用于 OC-48c 接口的业务引擎(TM, 整合在 THNPU 中)的框图. 交换结构采用 Vitesse 公司的 GigaStream 交换芯片组 VSC872/882; 外存采用减少延时的 RLDRAM II. 输入方向上, 变长的 IP 包经网络处理器(NP)处理后传输至业务引擎, 被缓存到外存中进行排队等待调度(图中 C 处). 当被调度至 B 处后, 形成 CFrame 帧, 传送到交换结构, 通过交换结构传输到目的输出端口中. A、B、C 三处都会对 IP 包进行处理, 增加额外开销, 降低 IP 包可获得的有效带宽. 这样会导致部分长度的 IP 包不能线速转发, 降低了整个系统的吞吐量. 而 A、B 两处的转发性能与 CFrame 最大净荷长度密切相关, 本文通过建模与分析, 并结合 THNPU 得出如何选择 CFrame 最大净荷长度的一般性结论.

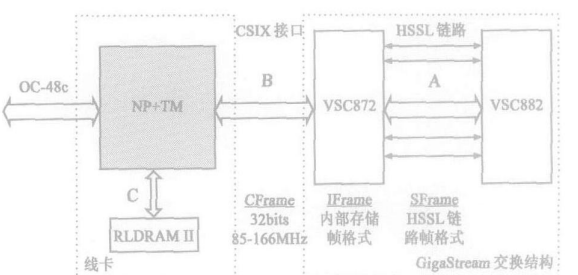


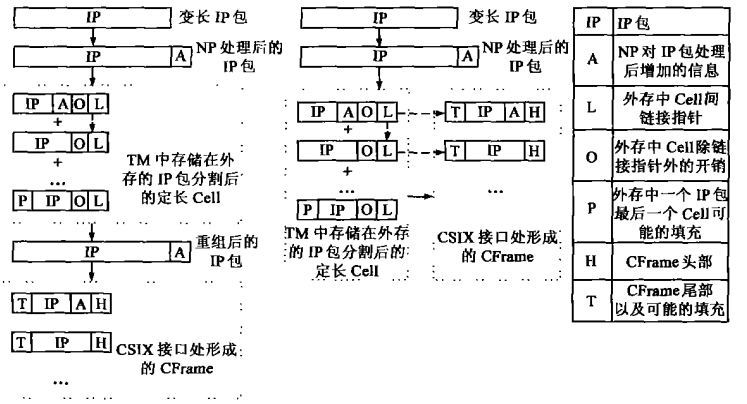
图 1 OC-48c 线卡系统 THNPU 框图

收稿日期: 2004-10-25; 修回日期: 2005-04-26
基金项目: 国家 863 高科技研究发展计划(No. 2002AA103011-1; No. 2003AA115110); 国家自然科学基金(No. 60173009; No. 60373007); 中国-爱尔兰政府间国际科技合作项目(No. CI-2003-02)

IP 包在外存中缓存时, 需要将 IP 包分割成定长信元, 这样能更好的管理外存并可实现动态队列. C 处的存储与 CFrame 最大净荷长度关系取决于不同的设计方式:

(1) B、C 离散方式设计: 如图 2(*a*), 变长 IP 包经 NP 处理后, 被分割成定长信元存储到外存中, 在 CSIX 接口处, 先将属于同一个 IP 包的所有信元重组成完整的 IP 包后再分割形成具体的 CFrame. 此时 B 与 C 处联系不大;

(2) B、C 耦合方式设计: 如图 2(*b*), 变长 IP 包经 NP 处理后, 被分割成定长信元存储到外存中, 在 CSIX 接口处, 将 IP 包在外存中的每一个信元单独形成 CFrame 帧. 此时 B 与 C 关系紧密, 共同影响交换吞吐量.



(*a*)B、C离散方式　　(*b*)B、C耦合方式　　(*c*)缩写含义表

图 2　TM 系统存储设计与 CSIX 接口关系

两种方式的选择涉及到系统吞吐量、设计的难易、系统的灵活性、调度方式等因素. 本文主要从吞吐量与设计难易的角度对 A、B、C 三处进行综合分析.

广泛使用的 Intel 公司 IXP2400 网络处理器, 在与 Vitesse 公司 GigaStream 交换结构相连时, CSIX 接口处 CFrame 最大净荷长度为 64– 120 B(Bytes, 字节)[5], 用户可以配置, 但如何选择, 可以参照本文给出的思路进行分析. 本文结构如下: 第 2 部分对整个系统进行建模和分析; 第 3 部分对实际系统 THNPU 进行分析并讨论提高吞吐量的途径; 第 4 部分对全文作了总结.

2　业务引擎和交换系统的建模与分析

将图 1 中 OC-48c 接口的线卡系统适当抽象, 可以得到一般的系统框图(输入部分), 如图 3, 外存可采用一般的 DRAM 或者特殊的如 RLDRAM 等. IP 包进入业务引擎后, 被分割成定长信元到外存缓存(C 处), 再经过两次转发(B 与 A 处), 最终从交换结构中出来. 在 A、B、C 三处都有可能达不到线速转发, 导致最终存在部分长度的 IP 包不能线速转发. 这个传输过程近似为无反馈的, 最终不能转发的 IP 包为三处不能转发 IP 包的并集.

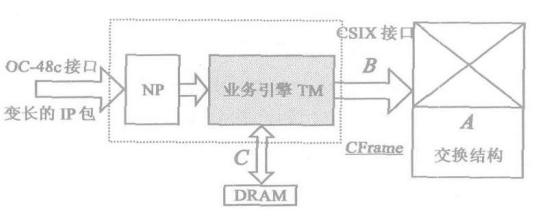


图 3　OC-48c 接口线卡系统抽象框图

同时, 计算 A、B、C 三处 IP 包的实际转发速率时, 不考虑其它两处不能被线速转发的情形, 仅考虑从 OC-48c 接口进入的 IP 包在此处获得的实际转发速率. 这样建模, 虽然忽略了一些次要的影响因素, 但仍能反映系统的实质, 且简单易行. OC-48c 数据接口的速率 $R(IP@OC48c)$ 为:

$$R(\mathrm{IP@OC48c}) = 155.52 \times 16 \times (260/270) = 2.39616\ \mathrm{Gbps} \quad (1)$$

根据 PPP 协议[6], OC-48c 的 HDLC 帧中, 每个 IP 包有 9 B 的开销: Flag 为 1B, Address 为 1B, Control 为 1 B, Protocol 为 2B, FCS 为 4B. 因此对于长度为 y B 的 IP 包, 它的实际数据接口速率 R(IP)为:

$$R(\mathrm{IP}) = \frac{R(\mathrm{IP@OC48c}) \cdot y}{(y+9)} \quad (2)$$

在 A、B、C 三处, 因均对 IP 包进行了分割, 并添加了一定的开销, 因此, IP 包实际获得的转发速率可表示为:

$$\text{IP 包实际获得的传输带宽} = \text{总的传输带宽} \cdot \frac{\text{IP 包的总字节数}}{\text{IP 包的总字节数} + \text{额外开销}} \quad (3)$$

IP 包被线速转发, 需满足不等式:

$$\text{IP 包实际获得的传输带宽} \geqslant R(\mathrm{IP}) \quad (4)$$

将 IP 包按照长度分类, 如果某个长度的 IP 包在 A、B、C 三处的任一处不满足式(4), 则认为此类 IP 包不能达到线速转发. 这样假设有一定局限性, 但可认为是系统测试的极端情况. 系统设计的最终目标是使达不到线速转发的 IP 包的类数最少, 从而取得较大的交换吞吐量.

3　THNPU 系统吞吐量分析

该部分将通过对实际系统 THNPU 进行分析, 讨论提高业务引擎中交换吞吐量的途径.

图 1 为 THNPU 的框图. VSC872 与 THNPU 的 CSIX 接口相连, 接收来自业务引擎的 CFrame, 去掉 CFrame 的头部(CH)和尾部(CT), 若 CFrame 净荷长度达不到最大净荷长度, 则填充 CFrame 净荷到最大净荷长度, 然后转换为用于内部存储的定长数据帧 IFrame; VSC872 与 VSC882 之间由 HSSL 链路(high-speed serial link, 高速串行链路)相连, 每路 CSIX 接口可以配置 2 路或 4 路 HSSL, 每路 HSSL 链路可运行于 2.125 或 2.64384 Gbps 的速率, 去掉链路开销后的实际数据传输速率分别为 2.0 与 2.5 Gbps. HSSL 链路上的数据格式为定长数据帧 SFrame, 其大小为 CFrame 最大净荷长度加上 12 B. 图 4 表示了上述数据格式的转换关系. 其中, CPayload 表示 CFrame 的净荷, 其中也包含为维护 CFrame 数据格式而增加的填充; IH、IT 为 IFrame 的头部和尾部, CP 为 IFrame 中将 CFrame 净荷填充到 CFrame 最大净荷长度, IFP 为 IFrame 中的填充; SH 为 SFrame 的头部. GigaStream 支持的 CFrame 最大净荷长度为 40– 120 B,

并且为4的整数倍[7,8].

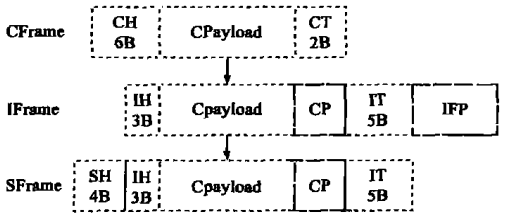


图4 VSC872/882数据格式转换图

如图1, CFrame最大净荷长度直接影响A、B处IP包可获得的实际转发速率, 因此, 本节先从A、B处分别分析如何选择CFrame最大净荷长度, 然后从B、C两处的互动设计方式来分析C处与吞吐量的关系, 最后, 对A、B、C三处进行总结.

3.1 A处选择 CFrame 最大净荷长度

交换结构处(A)总是传输定长帧, 因此容易计算出IP包在A处获得的实际转发速率.

理想情况下, CFrame的净荷没有额外的开销, 全部用来传送IP包. 并且不考虑NP在IP包上附加的信息, 则当CFrame最大净荷长度为 x B时, 长为 yB的IP包在A处获得的传输率为:

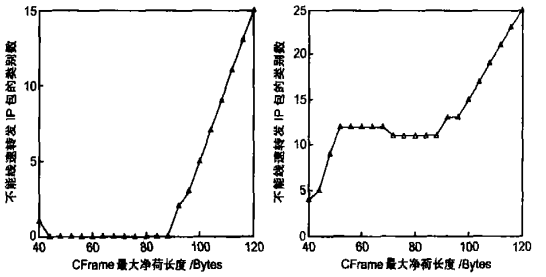
$$R(\mathrm{IP@HSSL}) = R(\mathrm{HSSL}) \cdot \frac{y}{(x+SF)\lceil \frac{y}{x} \rceil} \qquad (5)$$

式(5)中, $\lceil n \rceil$表示不小于 n 的最小整数, $R(\mathrm{HSSL})$表示HSSL链路的实际数据传输速率, $(x+SF)$表示SFrame的大小, $SF=12$.

当每路CSIX接口选用2路2.64384 Gbps的HSSL链路时, $R(\mathrm{HSSL})=2*(2.64384*16/17)$. 此时 $R(\mathrm{HSSL})$约为 $R(\mathrm{IP@OC48c})$的两倍, 故一般认图4VSC872/882数据格式转换图为选用2路HSSL链路应能线速转发所有类别的IP包. 本文将指出这种理解是不正确的.

对于每一种可能的CFrame最大净荷长度, 计算出所有类别的IP包(长度为40–65535 B, 依据长度分类)的R(IP)和 $R(\mathrm{IP@HSSL})$, 根据式(4)统计出不能线速转发IP包的类数如图5(a)所示. 由图知应选择 x 为44–88, 不能转发的IP包的类数均为0. 但实际上如图2, CFrame中的净荷是需要一定开销的, 包括IP包重组、为维护CFrame的格式而加入的填充等, 记此开销为 a B. IP包经NP处理后, 需增加一些信息, 如NIP(下一跳IP地址)等, 记这些信息长为 b B. 这样, HSSL处传输总的IP包长为 $(y+b)$B, 而CFrame净荷中有效IP为 $(x-a)$B. 实际系统中IP包在交换结构获得的传输率如式(6)所示. THNPU设计中: $a=4$(2B用于IP包重组, 2B用于填充); $b=8$(4B用于NIP, 其它用于线卡上的输出端口、流分类信息等). 同样, 对于每一种可能的CFrame最大净荷长度, 统计出不能线速转发IP包的类数如图5(b). 此时任何一种CFrame最大净荷长度都不能线速转发所有IP包. 从图5(b)给出的结果判断, x 为40时是最佳选择, 但此时连最小的IP包都需要两个CFrame帧封装, 会导致NP工作量增加很大.

$$R(\mathrm{IP@HSSL}) = R(\mathrm{HSSL}) \cdot \frac{y}{(x+SF)\lceil \frac{y+b}{x-a} \rceil} \qquad (6)$$



(a)理想系统 (b)THNPU系统

图5 HSSL处不能转发的IP包的类数

如果最小的IP包仅需一个CFrame封装, x 最小应为52*. 故合理的 x 为72–88, 52–68也可以接受. 经计算, 当 $a=2$, $b=0$(或 $a=0$, $b=4$)时, 任何一种CFrame最大净荷长度都不能线速转发所有IP包. 因此, 实际系统中A处每路CSIX接口使用2路HSSL链路是不能线速转发所有的IP包的, 必须使用4路HSSL链路才能线速转发所有IP包**, 此时, 选择任何一种CFrame最大净荷长度都是可以的***. 图6中4路HSSL链路不能转发的IP包的类数为0(每路工作在2.64384 Gbps). 从达不到2倍加速比的IP包的类数最少角度(包括一个CFrame即可封装最小的IP包的因素)考虑, 选择72–88B最佳, 52–68B较佳.

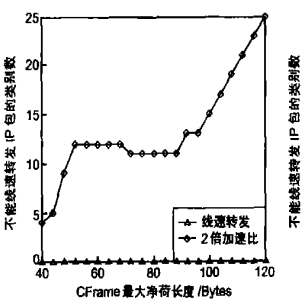


图6 四路HSSL不能转发IP包类数

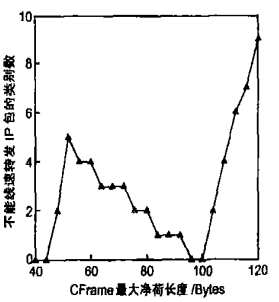


图7 当B处固定成帧时不能线速转发IP包的类数

3.2 B处选择 CFrame 最大净荷长度

CSIX接口处, GigaStream支持时钟频率为85–166MHz. 有两种方式形成CFrame帧. 固定成帧法: 将每一个IP包分割成定长的信元, 若最后一个信元不够定长则填充至定长, 再将每一个信元封装成CFrame; 浮动成帧法: 将IP包的最后一个信元按照实际大小封装成CFrame. 固定成帧设计简单, 但最后一个信元中可能的填充浪费了部分CSIX接口带宽, 当IP包较短时, 这种浪费尤其严重; 浮动成帧设计上复杂些, 但能尽可能保证IP包获得最大传输带宽.

固定成帧时, IP包在B处获得的传输率计算方式与式(6)相同, 只需将R(HSSL)换成R(CSIX), 表示CSIX接口的传输速率, SF 换成 CF, $(x+CF)$表示CFrame的大小, $CF=8$. 当CSIX接口时钟频率为166MHz时, 对于每一种可能的CFrame

* 以下讨论最佳CFrame最大净荷长度时均考虑这个因素, 即认为一个CFrame能封装最小的IP包.

** 一般配置每路CSIX接口只能选择2路或4路HSSL链路.

*** 计算知HSSL链路工作在2.125 Gbps速率也可以, 计算过程略.

最大净荷长度, 统计出不能线速转发 IP 包的类数如图 7. 从图 7 中得出, 此时选择 x 为 96 或 100 最佳. 但通过进一步的计算分析知, 此时转发性能与 CSIX 接口时钟频率关系太紧密. 当频率为 165MHz 时, x 为 96 或 100 时已经不能线速转发所有的 IP 包. 当频率为 100MHz(CSIX-L1 推荐的时钟频率, 此时 CSIX 接口可提供 3.2Gbps 的带宽)时, 最少不能转发的 IP 包的类数达到 251 种; 当频率为 125MHz 时, 最少的类数也达到 191 种. 浮动成帧时, 长度为 yB 的 IP 包在 B 处获得的传输率为:

$$R(\text{IP@CSIX}) = R(\text{CSIX}) \cdot \frac{y}{(x+CF)(n-1)+last}$$

$$last = \left(\left\lceil \frac{y}{4} \right\rceil 4 + b\right) - (x-a)(n-1) + CF + a \qquad (7)$$

式(7)中, $n = \left\lceil \frac{y+b}{x-a} \right\rceil$, $n-1$ 个信元按照 CFrame 最大净荷长度封装成 CFrame, 最后一个信元按照其实际大小封装. $\left(\left\lceil \frac{y}{4} \right\rceil 4 + b\right)$ 为维护 CFrame 格式, 可能加入填充后在 CSIX 接口处实际传输的 IP 包的长度. 当 B 处工作在 166MHz 时, 任何一种 CFrame 最大净荷长度都能线速转发所有的 IP 包. 图 8 是时钟频率为 166MHz 时, 每一种 CFrame 最大净荷长度转发 IP 包的最小速率. 图中表明, B 此时工作在较低的频率下仍能够线速转发所有的 IP 包, 进一步计算知 B 处工作在 125MHz 时选择任何一种 CFrame 最大净荷长度都能转发所有的 IP 包. 因此, B 处需要选用浮动成帧法, 才能尽可能地保证吞吐量, 同时设计又具有灵活性.

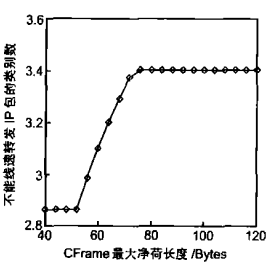


图 8　*B* 处浮动成帧时最小转发 IP 速率

在不考虑 C 处情况下, 综合 A、B 两处, 可以得出: A 处需使用 4 路 HSSL 链路, B 处需使用浮动成帧法, 才能保证设计的稳定性和可扩展性(即对 B 处时钟频率要求不严格, 而且 A 处能提供一定加速比), 同时保证任何一种 CFrame 最大净荷长度都能线速转发所有的 IP 包. 此时选择任何一种 CFrame 最大净荷长度都是可以接受的, 但是选择 72-88 B 最佳, 此时交换结构处达不到 2 倍加速比的 IP 包种类数较少.

通过分析 Internet 上的实际流量 Auckland-II[9], 我们知道网络中存在大量长度小于 64B 的 IP 包, 这类包个数约占所有包个数的 60%[10]. 因此, 如果能将这些 IP 包封装成一个 CFrame, 会减少网络处理器(业务引擎)的处理的时间. 这个意义上, 76B(64+12=76)及以上是更好的选择.

3.3　C 处的吞吐量考虑

本文第 1 部分中已经提到, 存在两种 B 与 C 处的互动设计方式: 离散方式和耦合方式. 离散方式增加了一个重组分割(SAR)操作, 但每个信元在外存中的开销小些, 耦合方式设计简单, 但每个信元在外存中增加了更多的开销. 两种设计方式中, IP 包在外存中获得的实际带宽计算方式与式(6)相同, 为:

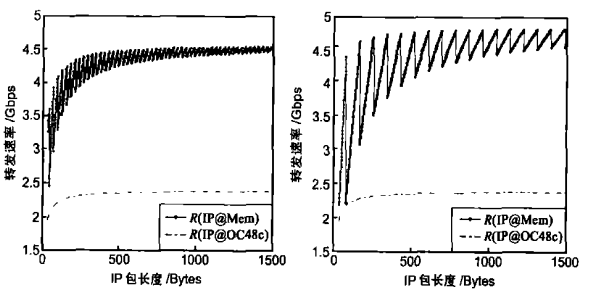
$$R(\text{IP@}Mem) = R(Mem) \cdot \frac{y}{Mem \cdot \left\lceil \frac{y+b}{Mem-c} \right\rceil} \qquad (8)$$

式(8)中, $R(Mem)$ 表示所有数据一次访存获得的总带宽; Mem 表示外存管理单位即定长信元的大小; c 表示每个信元中的额外开销; b 表示 NP 在 IP 包上附加的信息. 离散方式时 c 的值小些; 耦合方式时 c 的值大些, 因形成 CFrame 帧时需要额外信息. 此时 Mem 的大小直接影响到 CFrame 最大净荷长度, 进而影响 A、B 两处的转发性能.

THNPU 中外存采用 RLDRAM II[11,12], 读写分开的数据总线, 配置为 16M × 18bits, 数据有效位是 16bits. 与一般的 DRAM 仅有 4 个 Bank 不同, RLDRAM 有 8 个 Bank. 每次读写 8 个 Bank 效率最高, 加上至少为连续读或写 2 次(Burst2), 故每次读写 32B 效率较高. 因此 Mem= 32、64 或 96(32 整数倍)较合适. RLDRAM II 工作在 166MHz, 采用读写分开的数据线; 每 64 周期需要 1 个周期刷新; IP 包在外存中读写各一次, 共用总的存储器带宽; 因此:

$$R(Mem) = 0.166 \cdot 16 \cdot 2 \cdot \frac{63}{64}\text{Gbps} \qquad (9)$$

THNPU 设计中, 若采用离散方式设计, 则式(8)中 b= 12(存储 NIP、流分类、目的端口、长度等), c= 4(链接指针、分段信息). 对于各种 Mem, 计算出所有种类 IP 包的 $R(\text{IP@}Mem)$, 统计出达不到线速的类数 Num. Mem= 32 或 96 时, Num= 0; Mem= 64 时, Num= 1. 综合带宽大小及存储器利用率考虑, Mem 为 32 优于为 96. 图 9(*a*)为 Mem= 32 时 IP 包获得的有效带宽.



(*a*) *Mem*=32　　(*b*) *Mem*=96

图 9　IP 包在外存中获得的有效带宽

若采用耦合方式设计, 则 b= 8(存储 NIP、流分类、长度等), c= 8(链接指针、分段信息、目的端口). 此时, Mem= 64 时, Num= 1; Mem= 96 时, Num= 0. (Mem= 32 时不满足 CFrame 最大净荷长度大于 40B 要求, 不予考虑). 实际系统中, 允许有个别种类的 IP 包达不到线速转发, 故选择 Mem 为 64 或 96 均可. Mem 直接影响到 A、B 处 CFrame 最大净荷长度(为 Mem 减去 4B), 根据已有的分析结果, 选择 Mem= 96 时更佳, 能使 A、B、C 三处的情况达到最优, 缺点是外存的利用率稍低. 图 9(*b*)为 Mem= 96 时 IP 包获得的有效带宽.

B、C 采用离散方式设计时, 需增加一个重组分割的操作. 该操作稍微增大了系统设计的复杂性和代价. 图 10 中虚线框内为增加的功能模块. Reassembly 模块将来自 RLDRAM II 的同一个 IP 所有信元重组为原来的 IP 包, 并去掉可能的填充; Segment 模块根据选择的 CFrame 最大净荷长

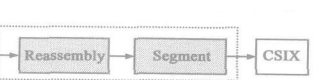


图 10　B、C 离散方式增加操作

度对 IP 包进行分割. 利用 Quartus 4.2 在 Altera 的 EP1S80 上综合,得到 Reassembly 模块占用 LE(Logic Element,逻辑单元)数为 415,Memory(FPGA 内嵌的存储)数为 38944 bits; Segment 模块占用 LE 数为 689,Memory 数为 39200 bits. EP1S80 共有 LE 数 79040,Memory 数 7427520 bits[13].

可见,离散方式只是稍微增加了系统设计的复杂性和占用资源数. 因此,需根据系统资源和性能灵活性综合决定 B、C 处的设计方式.

3.4 A、B、C 三处综合考虑

A、B 综合考虑已经得到较优的结论,A、B、C 综合时,主要取决于 B、C 两处的互动设计方式.

当 B、C 处采用耦合方式设计时,CFrame 最大净荷长度选择决定于外存管理单元的大小,同时需要考虑到 A、B 两处得到的 76 B 及以上是更好的选择的结论(若仅从吞吐量角度考虑,其它的选择也是可以的);当 B、C 处采用离散方式设计时,仅仅需要单独保证 C 处能线速存储 IP 包,可以从 A、B 两处考虑灵活选取 CFrame 最大净荷长度,甚至可以让用户配置.

最终 THNPU 设计中,A 处每路 CSIX 接口使用 4 路 HSSL 链路,B 处使用浮动成帧,并采用 B、C 耦合方式,CFrame 最大净荷长度为 92 B,能保证这三处都能达到 100% 吞吐量,同时适当降低了整个系统的设计复杂度,即减少一次 SAR 操作.

4 结论

本文通过分析影响业务引擎系统中交换吞吐量瓶颈的因素,给出了分析提高吞吐量的模型及方法. 并通过分析具体设计的 THNPU 系统(OG-48c)中的提高吞吐量的途径,指出 OG-48c 系统中采用 VSC872/882 交换芯片时,每路 CSIX 接口采用 4 路 HSSL 链路;并在 CSIX 接口处采用浮动成帧法,这样能保证选取任何 CFrame 最大净荷长度都能线速转发所有的 IP 包,但如需保证最小的 IP 包仅需一个 CFrame 封装以及交换结构处达不到 2 倍加速比的 IP 包种类数最少,选择 76- 88B 更佳. 另外,外存也与吞吐量有着密切的关系,对于不同的设计以及采用的不同外存,需要仔细分析它对吞吐量以及 CSIX 接口处的影响,做出最佳选择.

本文分析了线卡输入方向,输出方向可以类似建模分析. 同时,VSC872/882 为 Crossbar 结构的交换系统,若采用共享储存的交换系统,也可以进行类似分析,得出交换系统中必须提供的带宽.

致谢 感谢胡成臣、李竞对本文的有益建议.

参考文献:

[1] CSIX-L1 Specification V1.0[S]. http://www.npforum.org/techinfo/csixL1.pdf, Aug. 5, 2000.

[2] CSIX-L1 Frequently Asked Questions[DB/OL]. http://www.npforum.org/techinfo/CSIX_FAQ_D1.0.pdf, Dec. 5, 1999.

[3] Why Modern Switch Fabrics use a Fixed-Size Frame Format, Vitesse Semiconductor Corporation, white paper, V1.0[DB/OL]. http://www.vitesse.com, Jan. 27, 2004.

[4] Collier, Grudsky. The Switch Fabric Multiservice Dilemma, Vitesse Semiconductor, CommsDesign[DB/OL]. http://www.commsdesign.com/story/OEG20020702S0035, July 2, 2002.

[5] Intel IXP2400 Network Processor/Vitesse GigaStream Switch Fabric Solution White Paper, Intel Corporation, V1.0[DB/OL]. http://www.intel.com/design/network/papers/25214201.pdf, November, 2002.

[6] Perkins D. PPP: The Point-to-Point Protocol for the Transmission of Multi-Protocol Datagrams Over Point-to-Point Links[S]. IETF RFC 1171, 1990.

[7] GigaStream Intelligent Switch Fabric VSC872/VSC882 Design Manual[DB/OL]. http://www.vitesse.com, 2002.

[8] VSC872 and VSC882 data sheet[DB/OL]. http://www.vitesse.com, 2003.

[9] National Laboratory for Applied Network Research (NLANR), Auckland-II[DB/OL]. http://pma.nlanr.net/Special, 2004.

[10] 李文杰, 刘斌. 输入排队中抢占式的短包优先调度算法[J]. 电子学报, 2005, 33(4): 577- 583.
Li Wenjie, Liu Bin. Preemptive short-packet-first scheduling in input queueing switches[J]. Acta Electronica Sinica, 2005, 33(4): 577- 583 (in Chinese).

[11] RLDRAM II Data Sheets and Technical Notes[DB/OL]. http://www.rldram.com/datasheets/index.html, 2003.

[12] 288Mb SIO REDUCED LATENCY (RLDRAM II) Datasheet[DB/OL]. http://www.micron.com, 2003.

[13] Stratix Device Handbook[DB/OL]. http://www.altera.com, April, 2004.

作者简介:

孙长华 男,1982 年 8 月出生于湖北公安县,2004 年 7 月毕业于西安交通大学计算机系,获工学学士学位,现在清华大学计算机系攻读博士学位,主要研究方向为计算机网络,交换技术,网络流量控制等. E-mail: sch04@mails.tsinghua.edu.cn.

刘 斌 男,生于山东,清华大学计算机系教授,主要研究领域为交换技术,路由查找,流分类,网络安全等,发表论文 100 多篇,申请发明专利 10 项,获批 5 项.