IP 分组的 AAL3/4 适配及性能分析

游 骅, 刘增基, 陈 鹏
(西安电子科技大学 ISN 国家重点实验室, 陕西西安 710071)

摘 要: 本文提出一种用 AAL3/4 适配 IP 分组, 支持 MPLS(多协议标签交换)的方法. 给出单个分组经过基于信元的交换网络的端到端时延分析方法—"等价长度"法以及相应的结果, 进行了仿真验证; 并对此条件下接收端所需缓存容量进行了估计. 研究表明, 在实现 VC 合并时, 采用 AAL3/4 适配能够获得较小的分组端到端时延, 简化中间结点的操作并减小结点所需缓存容量. 与其它方式(AAL5)相比, 这是一种更好的适配方式.

关键词: AAL3/4; IP 分组; 多协议标签交换; 端到端时延
中图分类号: TN915.01 文献标识码: A 文章编号: 0372-2112 (2002) 10-1425-03

An Approach on IP Packet Adaptation Using AAL3/4 and Performance Analysis

YOU Hua, LIU Zeng-ji, CHEN Peng
(*National Key Lab on ISN, Xidian University, Xi'an, Shaanxi 710071, China*)

Abstract: A new scheme of adaptation for IP packet over ATM and support of MPLS (Multi-Protocol Label Switching) using AAL3/4 is proposed in the paper. Aiming at the performance of end-to-end delay, we suggest a theoretical model, "Equivalent Length", for analysis and perform the corresponding simulation. An estimation of the needed buffer at the receiver is also presented. The result shows that our approach, compared with current fashion using AAL5, can be a better choice due to its obtainable smaller end-to-end packet delay, simpler operation and less buffer needed at each switching node under the situation of VC-Merging.

Key words: AAL3/4; IP packet; MPLS; end-to-end delay

1 引言

数据业务的迅猛增长极大地推动了因特网的发展, 同时也对其施加了巨大的压力. 传统的分组转发方式已经不能满足业务的需求. 吸收 ATM 等面向连接技术的优点, 综合几家国际大公司的建议, 因特网工程工作组(IETF)提出了一种新的快速分组转发(交换)机制——多协议标签交换(MPLS). 在此机制下, 业务流分组在网络边缘结点(LER)处被分成不同的转发等价类(FEC), 并被标以不同的标签, 网络内部的结点按照标签进行转发或交换, 从而大大提高分组在网内转移的速度. 既然如此, 用 ATM 交换机构来实现标签交换路由器(LSR)就是很自然的. 因为 AAL5 的 SAR PDU 和 CPCS PDU 较为简单, 现有的 ATM 对 IP 分组的适配采用 AAL5 方式[1]. 但是在有些方面, AAL5 对 IP 乃至 MPLS 的支持是不够的. 比如, 当一个 IP 分组由多个 SAR PDU(信元)传送时, AAL5 只能指示属于该分组的最后一个信元, 如果其中某个信元丢失引起整个分组无效时, 只有在接收端恢复 CPCS PDU 时才能发现; 另外, 当需要支持(MPLS)标签合并时, AAL5 不支持信元直接交织的 VC 合并, 就需要中间结点增加缓存[2], 或者修改原有的 ATM 协议[3], 或增加支持的业务[4], 反而增加了实现的复杂性.

AAL3/4 也是 ITU-T 在 B-ISDN 框架中为适配数据业务所定义的方式. AAL3/4 SAR PDU 中比 AAL5 SAR PDU 增加了一些开销, 但是前者具备一些后者所没有的优点, 可以提供更加灵活和有效的支持. 本文将提出一种 AAL3/4 适配 IP 分组业务的方法, 给出单个分组在此方式下的端到端时延性能分析, 并对此条件下接收端所需缓存容量进行估计, 最后指出这种适配方式的实际应用价值.

2 AAL3/4 IP 分组适配方法

基于 ATM 的 MPLS 域如图 1 所示. 到达的 IP 分组在 MPLS 域的入口结点被适配成信元, 在网络中以信元为单位进行交换和传输, 在出口结点或其它需要的结点处恢复成原来的分组. 参照建议[5,6], 这里采用如图 2 的适配协议栈. 在入口结点处, IP 分组作为 CPCS 子层的净荷(Payload), 在 CPCS 子层为其生成 CPI、Btag、Etag 以及 BASize 域后, 形成 CPCS PDU; 然后由 SAR 子层以 44 字节为单位进行分段, 分配 MID(属于同一个 CPCS PDU 将具有相同的 MID), 组成 SAR PDU; 最后交由 ATM 层形成信元.

实现 ATM LSR 有两大问题, 一是为了支持 MPLS 标签合

收稿日期: 2001-05-21; 修回日期: 2002-02-25

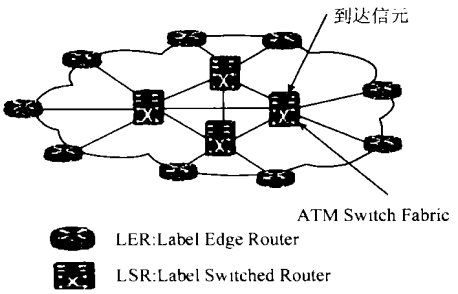


图 1 基于 ATM 的 MPLS 网络

并(Aggregation),就要求 ATM 交换机具有 VC 合并(VC Merge)的功能;再就是支持 TTL,以避免环路和满足控制命令要求. AAL3/4 的 CPCS PDU 和 SAR PDU 的协议控制信息分别有 8 字节和 4 字节之多,充分利用这些信息域能提供比其它方式更有力的适配,能够较为简单地解决这两个问题. 采用 AAL3/4 消息方式(Message Mode),利用 SAR PDU 中的 MID 域,在发送端已经将属于不同 FEC 分组的 SAR PDU 及承载它们的信元区分开,那么在需要进行合并的结点,只要仍旧使得属于不同分组的信元具有不同的 SAR PDU MID 就可以了. 另外,AAL3/4 CPCS PDU 的头(Header)和尾(trailer)分别居于此 CPCS PDU 的第一个 SAR PDU 和最后一个 SAR PDU,而且位于信元中相对固定的位置,利用 CPCS PDU 头或尾中信息域可以携带 MPLS 的 TTL.

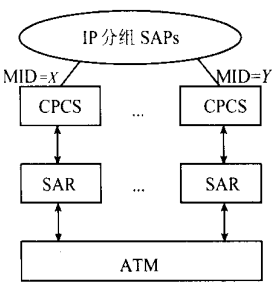


图 2 适配协议栈

3 性能分析

3.1 端到端时延

经过 AAL3/4 适配的分组在网络中以信元为单位传输和交换,该分组端到端的时延由发送端分段形成信元的时间 D_s、信元在网内交换和传输时延 D_x、以及接收端由信元重组成分组的时间 D_r 三部分组成. 若设 L_U 为 CPCS PDU U 的长度,T_s 为由 U 形成 ATM 信元或将信元恢复成 U 的单位时延,N 为信元在网络中传递的跳数,T_x 为经过每一跳所需的交换和传输时延,则显然有

$$D_s = \frac{L_U}{44} \cdot T_s \tag{1}$$

$$D_x = N \cdot T_x \tag{2}$$

应该注意到的是,D_r 并不一定等于 D_s. 按照一般的情形,在网络的中间结点有可能出现多个输入端口的信元竞争相同的输出端口的情况(如图 3 所示),具有相同目的结点,但来自不同(输入端口)CPCS PDU 的信元将交错在一起. 这样,在接收端为恢复 U 所需的时间是等待所有属于 U 的信元到达的等待时间加上信元重组成分组的时间. 为此,本文提出 CPCS PDU 或分组的"等价长度 EL"概念,U 的端到端时延主要取决于其"等价长度",所以下面主要对 EL 进行讨论分析.

所谓"等价长度 EL",就是指由于信元交织转移导致在进行 CPCS PDU U 恢复之前,从属于 U 的第一个信元到达一直到 U 的最后一个信元到达所需要等待的时间. 可以看出,随着经过的中间结点数的增加,由于别的信元插入,属于 U 的首尾信元间的信元个数也会增多,而 U 在中间结点转移的时延以及接收端恢复 U 所需的时间取决于 U 首尾信元的到达间隔,即 EL. 在任何交换结点,信元由输入端口被交换到输出端口的过程可用随机服务系统来刻划;属于 U 的信元以及与其具有相同输出端口的信元被交换到相应输出端口,可由一个"单服务员定长服务"的排队系统"GI/D/1"描述. 设第 n 跳结点不同输入端口竞争相同输出端口的信元流为相互独立的到达过程;令 t_k^n 为这些到达流的复合流中第 k 个信元到达的时刻,$\tau_k^n = t_k^n - t_{k-1}^n$ 为相继到达的信元间隔,记其分布函数为 $A_n(x) = P(\tau_k^n < x)$, $k = 1, 2, \cdots$;其统计平均值为 $1/\lambda_n \equiv \int_{0^-}^{\infty} x\, dA_n(x)$. 另一方面,服务时间为常数 $\frac{1}{\mu_n} \equiv \frac{53 \times 8}{u_n}$, u_n 为此结点为这些信元的服务速率(比特率).

在实际当中,ATM 交换机缓存信元的缓冲器容量是有限的,为保证一定的 QoS,无论有多少输入端口竞争同一输出端口,交换机将为信元提供一定的交换时延保证(比如最大时延). 为此,交换机对各个输入端口信元服务速率应大于总的信元到达的强度,即

$$\lambda_n < \mu_n \tag{3}$$

满足如上假设,有如下定理.

定理 1 若各个结点对信元的服务满足式(3),则 U 经过各个结点的 EL 的均值不变,且与经过第一跳后的 EL 均值相等.

证明 考虑第 n 跳交换结点. 令 w_k^n 为输入信元流第 k 个信元等待的时间,显然有

$$w_{k+1}^n = \max(0, w_k^n + \mu_n^{-1} - \tau_{k+1}^n), \quad k = 0, 1, 2, \cdots \tag{4}$$

其分布函数为 $W_k^n(x) \equiv P(w_k^n < x)$. 另设 t_B^{n-1} 和 t_B^n 为 U 的第一个信元到达前一交换结点和到达本结点的时刻,t_E^{n-1} 和 t_E^n 为 U 的最后一个信元到达相应结点的时刻,w_B^{n-1} 和 w_E^{n-1} 分别为第一个和最后一个信元在前一交换结点相应的等待时间. 则第 $n-1$ 和第 n 跳结点 U 的 EL 为 $D_{n-1} = t_E^{n-1} - t_B^{n-1}$,以及 $D_n = t_E^n - t_B^n = (t_E^{n-1} + w_E^{n-1} + D_C) - (t_B^{n-1} + w_B^{n-1} + D_C) = (t_E^{n-1} + w_E^{n-1}) - (t_B^{n-1} + w_B^{n-1})$. 这里 D_C 为固定传输时延.

根据已有的排队论结论[7],当式(3)满足,即 $\rho = \frac{\lambda_n}{\mu_n} < 1$ 时,极限分布 $\lim_{k\to\infty} W_k^n(x) = W(x)$ 存在,且独立于初始分布 $W_1^n(x)$. 故而 U 在此结点的 EL 均值为

$$\begin{aligned} E[D_n] &= E[(t_E^n - t_B^n)] = E[(t_E^{n-1} + w_E^{n-1}) - (t_B^{n-1} + w_B^{n-1})] \\ &= E[(t_E^{n-1} - t_B^{n-1})] + E[w_E^{n-1}] - E[w_B^{n-1}] \\ &= E[(t_E^{n-1} - t_B^{n-1})] = E[D_{n-1}] = \cdots E[D_1] \end{aligned}$$

可见,此交换结点 U 的 EL 的均值与前一结点的相同. 当每一跳都如此时,显然 U 的 EL 均值都相等,且等于经过第一跳后的值. 证毕

U 在网络中转移时,其首尾信元之间的信元数目会增大,

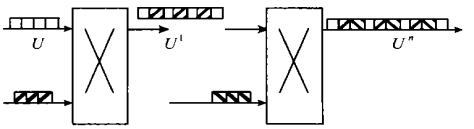


图 3　信元交织转移

尤其是在进行 VC 合并这类多点对一点通信时, 会单调增大. 但是, 由于式(3)的约束, 使得靠近信宿的交换机必须提供更高的服务速率, 因而 U 的 EL 不会不断加大, 继而整个 U 经过各个交换结点的平均时延 T_x 与单个信元的相同. 若设 d 为单个信元经过一跳交换结点的平均时延, 则有

$$D_r = d + T_s \qquad (5)$$

综上所述, 结合式(1)、(2), 则 U 的端到端平均时延为

$$D = D_s + N \cdot d + D_r = 2 \cdot \frac{L_U}{44} + (N+1) \cdot d \qquad (6)$$

3.2　缓存估计

前面对 CPCS PDU 的"等价长度 EL"进行了分析, 从而得到分组的端到端的时延性能. 应用得到的结果, 下面对接收端(需要恢复分组的结点)所需的缓存容量进行估算. 由于属于不同分组的信元会交织在一起到达, 要求在接收端设立不同的缓存队列用以暂存这些属于不同分组的信元. 按照最差的情形, 所需的队列数为通过不同输入端口至接收端的连接数 $C = \sum_{j=1}^{N} c_j$, 这里 N 为总跳数, c_j 为途经各个结点至本接收端连接的输入端口的数量. 则所需的最大缓存容量(以字节记)为

$$M = C \cdot \left(\left\lceil \frac{L}{53} \right\rceil + 1\right) \cdot 53 \qquad (7)$$

其中, L 为 IP 分组的最大长度. 极端情况下, C 取 1024(MID 允许的最大值), L 取 9000 字节(网络实测数据[8]中分组 MTU 值)时, M 约为 9 兆字节.

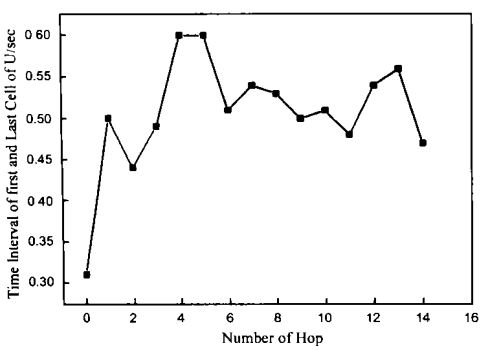


图 4　属于 U 的首尾信元时间间隔(样本)

3.3　仿真结果

为验证理论讨论的结果, 使用网络仿真软件对 U 在网络中转移过程进行了仿真. 分组经过的总交换结点数为 14 跳, 每个交换结点的端口数目为 16×16, 端口链路速率为 100 信元/秒, 各输入端口去向相同输出端口的业务强度为 5 信元/秒. 得到 U 经过各个交换结点后首尾信元的时间间隔(EL)如图 4 和图 5. 图 4 所示为某个分组经过各个交换结点的单个仿真样本的情形, 图 5 所示为仿真样本数为 10000 时的统计平均结果. 图中, 第 0 跳对应于由 CPCS PCU 形成 ATM 信元时的 EL, 即在业务源端的时间间隔; 其余为经过交换结点的 EL.

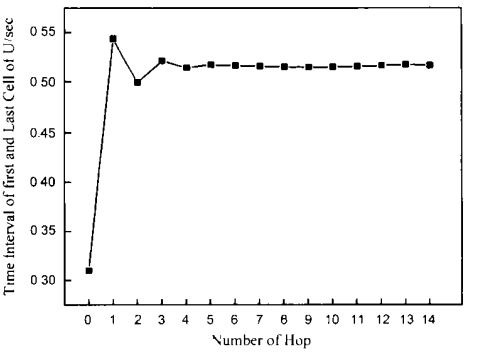


图 5　属于 U 的首尾信元时间间隔(平均)

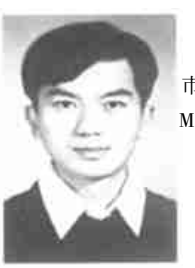
4　结论及实际应用

适当使用 AAL3/4 CPCS PDU 和 SAR PDU 的控制信息域, 能够较好地用 MPLS 实现对 IP 分组的传送. 经过分析和仿真发现, 在满足保证信元交换时延的前提下, CPCS PDU 的转移平均时延与信元的相同. 因而在实现 VC 合并时, 采用 AAL3/4 方式不需要缓存属于同一分组的所有信元, 并待这些信元收齐后才向下一跳发送; 因而分组在每一跳交换结点的时延较小, 端到端时延也较小. 虽然在接收端有可能需要较多的缓存容量, 但是中间结点的操作却得以简化, 对缓存的要求也会降低, 充分利用了 ATM 的优点. 因此, 与其它适配方式(AAL5)相比, 采用 AAL3/4 适配 IP 分组, 支持 MPLS 是一种更好的方式.

参考文献:

[1] RFC3035[S].
[2] M Grossglauser, et al. SEAM: Scalable an efficient ATM multicast [A]. Infocom'97 [C]. Japan: Infocom, 1997.
[3] Mario Baldi, et al. AAL5X: ATM adaptation layer 5 eXtension for efficient VC merging over ATM networks [A]. InfoCom98 [C]. USA: Infocom, 1998.
[4] Sridhar Komandur, et al. CRAM: Cell re-labeling at merger points for ATM multicast [A]. Proc. ATM'98 [C]. Fance: ATM, 1998.
[5] ITU-T Recommendation I.363.3[S].
[6] ITU-T Recommendation I.363.5[S].
[7] 徐光辉. 随机服务系统(第二版) [M]. 北京: 科学出版社, 1988.
[8] 12SDC-973265996.tsh.enc.gz, Internet traces [Z].
[9] available at http://moat.nlanr.net/Traces/Traces.

作者简介:



游　骅　男, 1974 年 1 月生于河南省新乡市, 西安电子科技大学博士生, 研究方向为 ATM, MPLS 及下一代信息网络业务服务算法.

刘增基　男, 1937 年 11 月生于浙江省丽水县, 西安电子科技大学教授, 博士生导师, 中国通信学会会士.