

文章编号: 1001-0920(2005)07-0830-04

一种基于公平性MPLS的流控制技术的研究

夏利¹, 赵娟¹, 邹董董², 王光兴¹

(1. 东北大学 a 教育部流程工业综合自动化重点实验室, b. 信息科学与工程学院, 辽宁沈阳 110004; 2 温州师范学院 计算机科学与工程学院, 浙江温州 325035)

摘 要: 多媒体业务在 Internet 上的广泛应用带来了响应流和非响应流间带宽享用的公平性问题, 严重影响网络的性能。提出一种基于公平的 MPLS 流控制机制, 该机制通过在 MPLS 模型中添加特定的队列管理算法 CSFQ, 在一定程度上解决了 MPLS 域中流间的公平性问题, 并且其对于 MPLS 设备的扩展完全可以通过软件的升级来实现, 有较好的可行性。

关键词: 多协议标记交换; CSFQ; 公平性

中图分类号: TP393 **文献标识码:** A

A Fairness-based MPLS Flow-control Mechanism

XIA Li¹, ZHAO Juan¹, ZOU Dong-dong², WANG Guang-xing¹

(1a Key Lab of Process Industry Automation of Ministry of Education, 1b School of Information Science and Engineering, Northeastern University, Shenyang 110004, China; 2 Institute of Computer Science and Engineering, Wenzhou Normal College, Wenzhou 325035, China Correspondent: XIA Li, E-mail: xial-neu@126.com)

Abstract: With the comprehensive application of multimedia in Internet, there is unfair competition for network resources between responsive flows and unresponsive ones, which reduce the performance of network. To such problem, a fairness-based MPLS flow-control mechanism is presented, which resolves the unfairness issue in the MPLS domain by integrating an active queue management algorithm CSFQ with the MPLS model. After analyzing the principle of the mechanism, the procedure of it is described.

Key words: MPLS; CSFQ; fairness

1 引言

随着网络技术的快速发展和网络规模的不断扩大, 在高速网络中按照用户的要求提供服务质量(QoS)保证已成为一个普遍的要求, 也是当前计算机网络中研究的热点问题。多协议标记交换(MPLS)的强大的流量工程能力使其成为解决未来 IP 骨干网络中 QoS 问题的基本技术^[1]。主动式队列管理(AQM)是 IETF 推出的基于 FIFO 调度策略的队列管理机制, 它使路由器主动参与到拥塞控制之中, 成为拥塞控制机制的主要研究方向^[2]。由于多媒体业务的广泛应用, 目前 Internet 上充斥着越来越多的非响应流^[2], 此类流不使用 TCP 拥塞控制机

制, 在拥塞发生时继续维持高发送速率来抢占带宽, 对带宽资源的不公平竞争使遵循端到端控制机制的响应流陷入“饥饿”状态, 甚至会导致整个网络的拥塞崩溃^[4,5]。基于公平性的 AQM 算法是被普遍看好的解决该问题的基本手段^[6]。

本文提出一种新的机制, 它通过在 MPLS 域中的节点内部添加实现特定的流控制算法 CSFQ, 来解决 MPLS 域中响应流和非响应流之间带宽享用的公平性问题。

2 CSFQ 算法

针对带宽享用的公平性, 所提出的 AQM 算法包括 FRED、CSFQ、CHOKe 以及 SFB 等算法。其

收稿日期: 2004-09-03; 修回日期: 2004-10-22

基金项目: 国家 863 计划项目(708-4-5); 辽宁省自然科学基金项目(20042012)。

作者简介: 夏利(1962—), 女, 四川乐山人, 副教授, 博士生, 从事网络 QoS 理论及路由算法研究; 王光兴(1937—), 男, 哈尔滨人, 教授, 博士生导师, 从事第三代移动通信、ATM 网络管理等研究。

中, CSFQ 是一种分布式的主动式队列管理机制, 它把网络中的路由器区分为边缘路由器和核心(非边缘)路由器, 在其上分别执行不同的动作^[6]. 这种划分结构使得 CSFQ 较其他算法能够更好地应用于 MPLS 网络中, 以解决 MPLS 域中的公平性问题

CSFQ 算法在边缘路由器处保持每个流的状态信息, 对于每个进入网络的分组, 边缘路由器判断它所属的流并根据所记录的流状态估计该流的速率, 然后在分组头部插入一个标签来记录这个估计值. 核心路由器不再保存流的状态信息, 当有分组到达时, 它和边缘路由器都使用 FIFO 队列进行分组调度, 分组的丢弃概率依赖分组头的标签和此时路由器处平均份额速率. 同时, 核心路由器还负责根据本地的业务流量聚合信息, 并重写分组头标签的内容. 边缘路由器和核心路由器的功能结构如图 1 所示, 其中 $r_i(t)$ 和 $a(t)$ 分别表示 t 时刻流 i 的估计速率和路由器处的平均份额速率

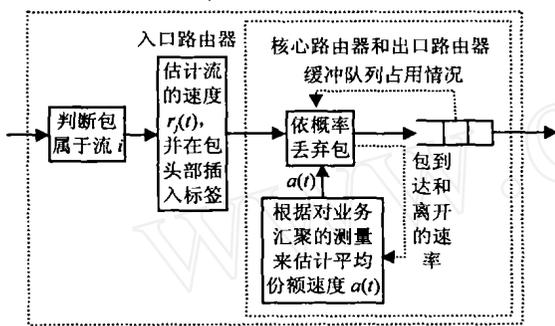


图 1 CSFQ 算法中边缘路由器和核心路由器的功能结构

2.1 流到达速率的估计

CSFQ 在边缘路由器处使用指数平均来估计流的速率 r_i , 指数平均能够更准确的反映流的变化, 并且不受具体分组结构的影响. 每次有新的分组达到, r_i 则作如下更新:

$$r_{i, \text{new}} = (1 - e^{-T_i^k/K}) \frac{t_i^k}{T_i^k} + e^{-T_i^k/K} r_{i, \text{old}} \quad (1)$$

其中: $T_i^k = t_i^k - t_i^{k-1}$, 是前后两个分组到达的时间差, K 是常数

2.2 平均份额速率 $a(t)$ 的估计

$a(t)$ 依赖于业务在该处的汇聚流量. 为了计算 $a(t)$ 的估计值 $a(t)$, 先定义分组的接收速率 $F(a(t))$, 本地的总到达速率 $A(t)$ 和链路输出速率 C . 即

$$F(a(t)) = \min_{i=0}^n (r_i(t), a(t)), \quad (2)$$

$$A(t) = \sum_{i=0}^n r_i(t). \quad (3)$$

F 和 A 分别表示接收速率和总到达速率的估

计值. 当链路拥塞 ($A(t) > C$) 时, $F(x) = C$ 的唯一解就是 $a(t)$; 否则, 当链路不拥塞 ($A(t) < C$) 时, $a(t) = \max_{i=0}^n (r_i(t))$. 有分组到达时, F 和 A 也采用类似式(1)的方法进行更新

2.3 丢弃概率的计算

对属于流 i 的分组, 记其丢弃概率为 p , 则

$$p = \max \left(0, 1 - \frac{a(t)}{r_i(t)} \right). \quad (4)$$

2.4 重写标签

如果某个流的一些分组由于网络拥塞被丢弃, 则余下分组的标签中便不再是对流速率的准确估计. 这时需要重新估算流速来更新标签. 此时已经不知道每个流的状态信息, 因此无法用原来的估计算法. 由以上的描述可知, 出队速率是入队速率和平均速率中的较小者. 所以, 新的标签 L_{new} 由下式给出:

$$L_{\text{new}} = \min (L_{\text{old}}, a(t)). \quad (5)$$

3 基于公平性的 MPLS 机制 f-MPLS

1998 年, IETF 推出了 MPLS 技术. 其基本思想是取消 ATM 等第 2 层信令, 将 IP 协议和 ATM 等下层协议紧密地结合在一起, 并使用 IP 和 MPLS 协议来控制 ATM 等交换结构. 同时, 在 MPLS 网络的边缘对收到的分组依据转发等价类 (FEC) 进行分类, 根据分类的结果给分组加上一个定长的标签, 该标签对应于与该分组的处理方式 (使用的路由、资源预留、业务量参数、业务等级等) 有关的一切信息

3.1 f-MPLS 的设计思想

为了解决 MPLS 网络中响应流和非响应流之间的公平性, 本文提出一种基于公平性的 MPLS 机制 f-MPLS. 该机制结合了 CSFQ 算法和 MPLS 模型, 在 MPLS 域的入口路由器处保存网络中所有流的状态信息. 路由器根据这些信息估算流的速率并将其映射为分组的丢弃优先级, 且写入标记封装中. 因此, f-MPLS 需要对 MPLS 做以下扩展

3.1.1 LSR 的功能扩展

MPLS 标记交换路由器 (LSR) 是 MPLS 网络中的基本组成实体, 它由控制单元和转发单元构成. 其中, 控制单元负责路由的选择、标签分发协议 (LDP) 的实现、标签的分配与发布以及标签信息库 (LB) 的形成. 转发单元只负责依据 LB 建立标签转发信息库 (LFB) 和对已标记的分组进行简单的转发操作. f-MPLS 在 LSR 中添加一个流管理单元, 入口 LSR 的流管理单元包括流状态信息表和流速估算单元, 前者记录网络中所有流的状态, 后者根据这些状态利用式(1)估算流的速率. 核心 LSR 和出口 LSR 的流管理单元则只需使用式(2)完成本地平均份额速率估算, 并根据式(5)重写标签 (即重新

估算流速)。扩展后的LSR结构如图2所示

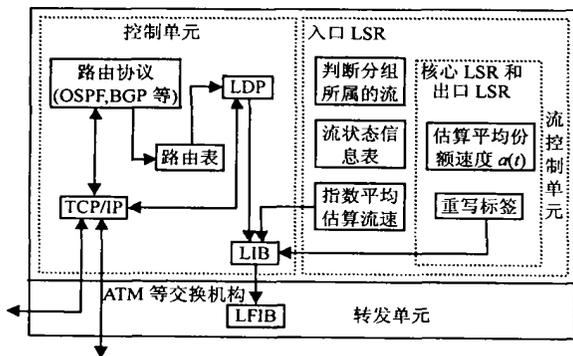


图2 f-MPLS中LSR的结构

在边缘路由器处, f -MPLS 对路由器的要求和本地流的数目有关。MPLS 具有流聚合功能, 且其边缘路由器本身就具有分类、标记等功能, 所以采用CSFQ 算法后, 路由器负担并没有明显增加

3.1.2 LB和LFB的扩展

MPLS 标记通用封装, 共4个字节。其中Label 字段长度为20 bit, 用来记录实际的标签值, 分组的转发依赖于该标签值; EXP 字段占3 bit, 保留作实现高级应用; S 字段1 bit, 用作栈底标记; 8 bit的TTL 字段为生存期字段。在传统的IP网络中, 分组头部既表示分组的下一跳也可以指示分组的优先级。类似地, 在MPLS域中, 标签不仅标识相同的FEC, 也可以指示分组的优先级或服务类型。此时, 标签便是分组FEC和优先级共同作用的产物^[7]。所以, 将CSFQ 中对流速的测量映射到标签可以有两种方法: 一种是扩展标准的标记分发协议(LDP), 将对流速的测量结果结合FEC一起映射到Label 字段; 另一种不需要改动LDP, 而将流速的测量结果映射为一定的优先级值, 记录在标签头部的EXP 字段中。第1种方法需要对现有的MPLS 标准协议作重大修改, 难以实现。本文中采用第2种方法, 不对LDP 进行改动, 但需要扩展LB 和LFB。

LDP 结合路由表根据特定的路由协议生成LB, LB 的每行记录包括Network, LSR 和Label 三个字段, 分别记录子网号, LSR 名及其对应的标签。下层的交换单元依据该LB 生成LFB, 它的每行记录包括Label, Action 和Next hop 三个字段, 分别表示入标记、执行的处理(出标记或出栈)以及下一跳LSR 名。LFB 是MPLS 转发的关键, 分组在MPLS 中依据LFB 的信息交换转发, 它的功能相当于IP网络中的路由表。在 f -MPLS 中, LB 不仅由LDP 和路由表生成, 还需要考虑该分组所属的流的发送速率。因为EXP 字段只有3位, 所以将分组的估算流速量化为8个等级(0~7), 量化的粒度可以根据网络的实际情况进行具体配置。因此, f -

MPLS 中LB 和LFB 都需要增加一个precedence 字段来记录量化后的流速, 在转发时指示丢弃的优先级。LB 和LFB 扩展后的具体结构如图3所示。其中precedence 的值从0~7依次表示优先级别的上升。

Network	LSR	Label	precedence

(a) LB 表记录格式

Label	Action	Next hop	precedence

(b) LFB 表记录格式

图3 扩展后的LB和LFB

3.2 f-MPLS的工作过程

由图2可以看出, f -MPLS 的标记交换过程主要有以下5个步骤:

1) LDP 和传统路由方法(OSPF, BGP 等)一起, 在各个LSR 中为有业务需求的转发等价类建立路由表, LSR 根据本地的路由表向下游LSR 发送标记请求消息, 下游LSR 记录下这一请求消息, 继续向下转发标记请求消息。

2) 当标记请求消息到达目的节点或是MPLS域的出口节点时, 如果还有可供分配的资源, 则该节点为相应的FEC 分配标记, 并完成LB 和LFB (此时precedence 字段为空), 然后向上游发出标记映射消息。收到标记映射消息的上游LSR 向上重复此过程直到入口LSR。此时, LSP 即告建立。

3) 入口LSR 接收分组, 完成第3层功能。首先判断分组所属的流, 根据所记录的流的状态估算该流的速率并将其量化结果填入LB 和LFB 的precedence 字段; 然后判定分组所属的FEC, 给分组加上标记, 其中流速的量化结果写入EXP 字段。

4) 核心LSR 不再对标记分组进行任何第3层处理, 只是依据分组上的标记以及LFB 通过交换单元对其进行转发。当有拥塞发生时, 依据EXP 字段所记录的优先级别对分组进行丢弃。

5) 在MPLS 出口的LSR 上, 将分组中的标记去掉继续转发。

4 结 语

MPLS 是下一代IP 骨干网络的主要技术。而随着大量多媒体等业务的出现, 非响应流和响应流之间带宽享用的公平性问题也是未来IP 网络需要解决的一大问题。本文提出了一种基于公平的MPLS 机制, 并详细描述了其原理和工作过程。该机制通过在MPLS 网络的节点内部加入一种流控制算法CSFQ, 从理论上可以保证一定程度的流间公平性。而且, 它对于MPLS 设备的扩展完全可以通过软件

的升级来实现, 有较好的可行性

参考文献(References)

- [1] 吴江, 赵慧玲. 下一代的 IP 骨干网络技术——多协议标记交换[M]. 北京: 人民邮电出版社, 2001.
(Wu J, Zhao H L. *Next Generation IP Backbone Networking Technique: multiprotocol Label Switching* [M]. Beijing: Posts & Telecom Press, 2001.)
- [2] RFC2309. Recommendations on Queue Management and Congestion Avoidance in the Internet[S]. 1998
- [3] Floyd S, Fall K. Promoting the Use of End-to-end Congestion Control in the Internet[J]. *IEEE/ACM Trans on Networking*, 1999, 7(4): 458-472

- [4] Jacobson V. Congestion Avoidance and Control[A]. *Proc of ACM SIGCOMM* [C]. 1988: 314-329.
- [5] Vern Paxson. End-to-end Internet Packet Dynamics [J]. *IEEE/ACM Trans on Networking*, 1999, 7(3): 277-292
- [6] Stoica I, Shenker S, Zhang H. Core-Stateless Fair Queueing: A Scalable Architecture to Approximate Fair Bandwidth Allocations in High Speed Networks [J]. *IEEE/ACM Trans on Networking*, 2003, 11(1): 1432-1446
- [7] RFC3031. *Multiprotocol Label Switching Architecture* [S]. 2001

(上接第 814 页)

- [3] 谢元旦, 夏淑艳. PD 继电器整定几种算法的比较[J]. *控制与决策*, 1993, 8(1): 77-79.
(Xie Y D, Xia S Y. The Relay Self-tuning of PD Regulators Parameters[J]. *Control and Decision*, 1993, 8(1): 77-79.)
- [4] 王肖峰, 徐科军, 沈峰. 小波滤波器的幅频特性研究[J]. *合肥工业大学学报*, 2003, 26(2): 171-175.
(Wang X F, Xu K J, Shen Y. Research on the Amplitude-frequency Characteristic of Wavelet Transform [J]. *J of Hefei University of Technology*, 2003, 26(2): 171-175.)
- [5] 孙立宁, 李志鹏, 曲东升. 小波变换在数字滤波中的实现[J]. *锅炉制造*, 2004, (1): 33-35.
(Sun L N, Li Z P, Qu D S. Application of Digital Filter Technology Based on Wavelet Transform [J]. *Boiler Manufacturing*, 2004, (1): 33-35.)
- [6] 张玉磷, 陈伟民, 杨建春, 等. 小波变换自适应滤波器及在主动噪声控制中的应用[J]. *控制与决策*, 2002, 17

(1): 107-110

(Zhang Y L, Chen W M, Yang J C, et al. Wavelet Transformation Adaptive Filter and Its Application in Active Noise Control[J]. *Control and Decision*, 2002, 17(1): 107-110.)

- [7] Donoho D L. De-noising by Soft-thresholding [J]. *IEEE Trans on Information Theory*, 1995, 41(3): 613-627.
- [8] 杨福生. *小波变换的工程分析与应用*[M]. 北京: 科学出版社, 2000: 42-69.
(Yang F S. *The Engineering Analysis and Application of Wavelet Transformation* [M]. Beijing: Science Press, 2000: 42-69.)
- [9] 徐晨, 赵瑞珍, 甘小冰. *小波分析与应用算法*[M]. 北京: 科学出版社, 2004: 91-118.
(Xu C, Zhao R Z, Gan X B. *The Wavelet Analysis and Application Algorithm* [M]. Beijing: Science Press, 2004: 91-118.)

(上接第 829 页)

进而求得 $kT + 16$ 时刻的控制输入. 取设定值 $r = 1$, 当控制器参数 p_1 分别取为 6 和 10 时, 系统的输出和输入如图 1 所示. 当 p_1 取 6 和 10 时, 经计算 Ψ 的特征值均在单位圆内, 其中特征值最大幅值分别为 0.9853 和 0.9332, 由定理 1 知闭环系统稳定. 可以看出, 控制器参数 p_1 的大小会影响控制的稳定性和快速性.

6 结 语

本文针对一类在实际生产过程中有着广泛应用的多速率采样系统, 提出了一种基于单值预报的多速率广义预测控制算法. 通过在系统周期 T 内根据不同的采样时刻构造扩展模型序列的方法, 解决了多速率采样预测控制系统中的因果约束问题. 在此基础上, 应用线性矩阵不等式的方法得到了有关多速率闭环系统稳定性的新结论. 本文的结果为多速

率采样预测控制系统的设计提供了一个新的思路

参考文献(References)

- [1] Jie Sheng, Tongwen Chen, Sirish L S. Generalized Predictive Control for Non-uniformly Sampled Systems [J]. *J of Process Control*, 2002, 12(3): 875-885.
- [2] Masahiro O. Multirate Multivariable Model Predictive Control and its Application to a Polymerization Reactor [J]. *Int J Control*, 1994, 59(3): 731-742.
- [3] Ling K V, Lim K W. A State Space GPC with Extensions to Multirate Control[J]. *Automatica*, 1996, 32(7): 1067-1071.
- [4] Colaneri P, Scattolini R, Schiavoni N. Stabilization of Multirate Sampled-data Linear Systems [J]. *Automatica*, 1990, 26(2): 377-380.
- [5] Kranc G M. Input-output Analysis of Multirate Feedback System [J]. *IRE Trans Automatic Control*, 1957, 3: 21-28.