多性能目标分组调度策略*

江 勇, 吴建平

(清华大学 计算机科学与技术系,北京 100084)
E-mail: yong98@mails.tsinghua.edu.cn
http://netlab.cs.tsinghua.edu.cn

摘要: 在高速分组交换网络中,分组调度策略和算法的设计是一个关键问题.由于网络分组调度策略有着多方面性能的要求,如何同时满足多个性能目标是当前的研究难点.基于比例公平性原则,提出了一种分组网络中的比例公平调度策略(proportional fairness scheduling,简称 PFS),该调度策略综合考虑了网络效率、用户 QoS 要求和系统公平性等多维目标,对该策略进行了详细的分析和论证.比例公平调度策略可以广泛应用于分组网络调度策略的设计研究和改进.

关 键 词: 比例公平性原则;分组调度;QoS;公平性
中图法分类号: TP393 文献标识码: A

未来的高速分组交换网络需要支持多样的服务类型.这些服务类型之间在流特性和服务质量要求等方面有很大的不同.各种新的网络应用对分组交换网络中调度策略的转发效率、带宽、延迟、丢失率以及系统公平性等都提出了新的要求,如何同时满足这些要求是当前的研究难点.任何一种实际的分组调度策略都只能对多个服务质量要求进行折衷.

这就引出了一个问题,如何设计一种分组交换网络中的调度策略,或者说资源分配机制,使其能够有效而公平地分配网络资源,而且更重要的问题在于不应该只考虑单个性能目标(如吞吐率或者延迟等等)的要求,而应该考虑多个性能目标的、综合的要求,而这在目前的网络技术研究中还是开放问题.

以前对网络中调度策略的研究工作主要集中在问题的某个方面,比如某一个性能目标的要求或者某些特定领域的综合性能研究.例如,有几种新的流模型,确定的[1,2]或随机的[3~5],被提出来作为端到端网络的分析和解决,诸如延迟、吞吐率和存储空间等性能参数的方法;在文献[6]中,作者比较了吞吐率和延迟抖动对不同 IP 分组的影响,进而提出了一种非对称尽力发送服务模型,对两种 IP 分组提供不同的吞吐率和延迟抖动;文献[7]研究了传统强迫优化算法(classical constrained optimisation)和遗传算法(genetic algorithm)在吞吐率、公平性和时间复杂度方面的性能差异,作者随后提出了一种综合的折衷方案,但它主要关注的是带宽分配问题.

我们注意到,到目前为止还缺乏有效的综合多种性能目标的分组调度策略.本文力图在这方面做一些理论上的探讨.我们提出了基于丢失率和转发延迟的比例公平调度策略(proportional fairness scheduling,简称 PFS),该调度策略在综合考虑网络效率、用户的服务质量(QoS)要求和系统公平性等多方面的性能目标的基础上提出了一种综合方案,本文对比例公平调度策略给出了严格的证明.我们的研究工作也提供了对网络调度策略及性能研究有益的理论探索.

本文第 1 节介绍了比例公平调度策略的相关知识背景.第 2 节详细介绍了 PFS 采用的延迟比例函数和丢失率比例函数.第 3 节提出和证明了基于丢失率和转发延迟的比例公平调度策略.第 4 节对比例公平调度策略进

* 收稿日期: 2001-06-20; 修改日期: 2002-01-14
基金项目: 国家自然科学基金资助项目(90104002)
作者简介: 江勇(1975－),男,重庆人,博士,主要研究领域为计算机网络体系结构,高性能交换结构,调度算法;吴建平(1953－),男,山东巨野人,博士,教授,博士生导师,主要研究领域为计算机网络体系结构,计算机网络协议测试,形式化技术.

行了性能模拟和测量.第 5 节总结全文并指出了进一步的研究方向.

1 相关背景

1.1 比例公平原则

Internet 应用之间以及使用者之间有着非常不同的服务要求,这使得目前的同一服务模型在某些情况下存在着较大的局限性.在相对区分服务(relative differentiated services)[8]中,网络流被组合成一个个的服务类(service classes),这些服务类按照其分组转发质量要求进行排序以决定其排队延迟、分组丢失率等转发行为.假定网络流被分类组合成排序的服务类,类 i 应有更好于(或者至少不差于)类$(i-1)(1<i<N)$的服务质量(排队延迟和分组丢失率).注意,“至少不低于”是必须的,因为在低负载的情况下,所有分组的服务质量是相同的,即满足所有流的服务要求.这样 Internet 用户或者应用程序能够选择最符合它们要求的质量和价格限制的服务类.在这种情况下,由于没有准入控制和资源预留,Internet 应用程序和使用者不能获得绝对的服务质量保证,如端到端的延迟界限和带宽等,但网络能保证高级别的类享有比低级别类相对更好的服务.在文献[9]中,我们提出了一种比例公平性原则(proportional fairness principle).

比例公平性原则按照网络管理者给定的区分参数按比例分配网络资源,从而得到相应的服务性能.若用 q_i 代表服务类 i 的性能量值,则比例公平性原则给每对服务类加入如下的限制:

$$q_i / q_j = c_i / c_j, i, j = 1,..., N, \tag{1}$$

其中 $c_1 < c_2 < ... < c_N$ 是一般的服务质量区分参数.因此,即使每个类的服务质量随着其负载发生变化,但类间的服务质量比值是不变的,与负载无关.

考虑基于排队延迟的公平性,采用排队延迟作为比例公平性原则的性能参数.若用 $\hat{d}_i$ 代表服务类 i 分组的队列延迟界限,比例公平性原则可以表述为对所有服务类对 i 和 j 有

$$F = \hat{d}_i / \delta_i = \hat{d}_j / \delta_j, i, j = 1,..., N, \tag{2}$$

其中参数 $\{\delta_i\}$ 是用户要求的延迟区分参数(delay differentiation parameters,简称 DDPs),由于高级别类有更好的服务性能,故有 $\delta_1 > \delta_2 > ... > \delta_N > 0$.

同样,我们用 $\hat{l}_i$ 代表类 i 的丢失率限制,基于分组丢失率的比例公平性原则要求服务类间丢失率满足

$$F = \hat{l}_i / \sigma_i = \hat{l}_j / \sigma_j, i, j = 1...N, \tag{3}$$

其中 σ_i 是丢失率区分参数(loss rate differentiation parameters,简称 LDPs),有 $\sigma_1 > \sigma_2 > ... > \sigma_N > 0$.

1.2 服务描述函数

服务描述函数的概念最早出现在 Parekh 和 Gallager 的工作[10]中,他们在其调度算法中引入了一种通用的服务描述函数,这种方案的一个优点在于,它能把对服务质量的要求通过一个简单的描述函数来表示,并将一个网络的连结的服务特性与其他网络连结区别开来;它的另一个重要特点是,给了服务器更大的灵活性来为达到不同的延迟和吞吐率要求分配系统资源.

本文中采用的服务描述的定义也可见文献[11,12].

2 比例公平函数

在本文中我们假定系统时间分成一个个的小时间片,编号为 0,1,2,…,考虑将所有分组分为 M 个服务类,并假定在每个时间片内网络服务器(交换机或路由器)能服务 c 个分组,c 称作服务器的服务能力.注意,在本文中我们考虑了分组丢失.

我们用 $R_i^{\text{in}'}[t]$ 代表在时间片 t 服务类 i 中到达服务器的所有分组数, $L_i[t]$ 代表该服务类在 t 中丢失的分组数,令 $R_i^{\text{in}}[t] = R_i^{\text{in}'}[t] - L_i[t]$ 表示到达服务器并最终得到服务的分组数;用 $R_i^{\text{out}}[t]$ 代表在时间片 t 服务类 i 离开服务器的分组数, $Q_i[t]$ 代表在 t 时暂存在服务器中的分组数,其中 t 是一个非负整数.不失一般性,令

$R_i^{\text{in}}[0]=R_i^{\text{in}'}[0]=0$, $R_i^{\text{out}}[0]=0$, $Q_i[0]=0$, $L_i[0]=0$.

定义 $R_i^{\text{in}}[s,t]$ 为在时间间隔 $[s,t]$ 内到达的分组数, $R_i^{\text{in}}[s,t]=\sum_{m=s}^{t}R_i^{\text{in}}[m]$.若 $s>t$,定义 $R_i^{\text{in}}[s,t]=0$.同样, $R_i^{\text{out}}[s,t]$ 定义为在时间间隔 $[s,t]$ 内离开的分组数.为了简化起见,在后面我们集中讨论一个给定的服务类并省略下标 i.

假定在一开始服务器中没有分组,因而在时间片 t 结束时服务器中的暂存分组数为

$$Q[t]=R^{\text{in}}[1,t]-R^{\text{out}}[1,t]\geq 0. \tag{4}$$

相对于 t 的丢失率 $l(t)$ 定义为

$$l(t)=\frac{L[0,t]}{R^{\text{in}'}[0,t]}. \tag{5}$$

相对于 t 的延迟 $d[t]$ 定义为

$$d[t]=\min\{\Delta:\Delta\geq 0 \text{ and } R^{\text{in}}[1,t]\leq R^{\text{out}}[1,t+\Delta]\}. \tag{6}$$

注意到,如果该服务类的分组离开服务器的顺序和其到达顺序一致(FIFO),那么 $d[t]$ 的上限为在 t 中到达的分组的延迟.

定义 **1**(突发性限制). 给定一个非减函数 $b(\cdot)$ 作为到达函数,我们说输入网络流 R^{in} 是 b 形的,若对所有满足 $s\leq t$ 的 s 和 t 有 $R^{\text{in}}[s+1,t]\leq b(t-s)$ 成立.对于某些特殊情形,如 $b(x)=\sigma+\rho x$,我们说 R^{in} 是 (σ,ρ) 形的.

由前面的延迟比例公平性原则[9]有, $d_i/\delta_i=d_j/\delta_j=\tilde{d}$,其中 $\tilde{d}$ 为延迟比例公平参数,其值我们将在后面的比例公平调度策略的可行性分析中加以讨论,在此我们将其看做是一个已知量.容易发现,若服务器对每个服务类 i 均满足延迟条件 $d_i^{\max}=\delta_i\tilde{d}$,则服务器在服务类间满足延迟比例公平性原则(见式(2)).

定义 **2**(延迟比例函数). 假设服务类 i 的网络流符合某种 b 形,要求服务器产生的延迟最大为 $d_i^{\max}=\delta_i\tilde{d}$ 个时间片.假定 $P^{D}{}_i(\cdot)$ 是一非减函数,

$$P^{D}{}_i(t)=\begin{cases}0, & \text{if } 0\leq t\leq d_i^{\max}-1\\ b(t-d_i^{\max}), & \text{if } t\geq d_i^{\max}\end{cases}. \tag{7}$$

若对任意 t 存在 $s\leq t$ 使 $Q_i[s]=0$ 且 $R_i^{\text{out}}[s+1,t]\geq P^{D}{}_i(t-s)$,则我们说服务器保证延迟比例函数 $P^{D}{}_i(\cdot)$.

容易发现, $P^{D}{}_i(t-s)$ 描述了在给定的时间间隔 $[s+1,t]$ 中必要的最小离开服务器分组数,其中 t 是任意给定的时间片,s 是某个不大于 t 的时间片,且在该时间片末暂存分组数为 0.

接下来,我们分析服务器的丢失率控制.

定理 **1**(丢失率控制). 假定 $R^{\text{in}}[t]$ 符合到达函数 $b(\cdot)$,并且服务器保证延迟比例函数 $P^{D}(\cdot)$.若分配给服务类 i 的缓冲区为 B 且存在一个常数 $\hat{l}$ 使

$$P^{D}(t)\geq b(t)(1-\hat{l})-B,\ \forall t\geq 0, \tag{8}$$

则丢失率 $l(t)$上限为 $\hat{l}$.

定理 **2**(丢失率比例公平). 假定服务类 i 的 $R_i^{\text{in}}[t]$ 符合到达函数 $b_i(\cdot)$,并且服务器保证延迟比例函数 $P^{D}{}_i(\cdot)$.若分配给服务类 i 的缓冲区 B_i 满足

$$B_i=\max\{b_i(t)(1-\sigma_i\tilde{l})-P^{D}{}_i(t),\ t\geq 0\}, \tag{9}$$

则服务器对服务类的丢失率满足比例公平性原则.

由于所有服务类缓冲区 B_i 的总和要小于系统总的缓冲区空间 B_{total},即 $\sum_{i=1}^{M}B_i\leq B_{\text{total}}$,故丢失率比例公平参数 $\tilde{l}$ 应满足 $\tilde{l}\geq(\sum_{i=1}^{M}[b_i(t)-P^{D}{}_i(t)]-B_{\text{total}})/\sum_{i=1}^{M}\sigma_i b_i(t)$.

3 比例公平调度策略 PFS

考虑一个 M 个服务类的服务器.假定服务类 i 需要服务器保证丢失率限制以及延迟比例函数 $P^{D}{}_i(\cdot)$,i=1,…,M.我们希望设计一种调度策略按照比例公平性原则提供延迟界限保证和丢失率界限保证.

根据定理 2 中丢失率比例公平的条件,在输入侧由缓冲控制器和丢失率控制对每服务类队列的缓冲区大小和分组丢弃策略按照式(9)给予限制,从而对服务器内的服务类提供比例公平的丢失率界限保证.

而对调度器的设计遵从延迟比例公平的原则.为简化起见,我们假设 $P^{D_i}(\cdot)$ 取整数值.一个服务器称为空的,条件是所有服务类在服务器中的暂存分组数在时间片 t 末均为 0,也就是说, $Q_i[t]=0$,i=1,…,M.对于一个非负整数 t,定义 $\tau(t)$ 为不大于 t 的最后一个服务器为空的时间片,也即

$$\tau(t)=\max\{s: s\le t \quad \text{and} \quad Q_i[s]=0, \qquad i=1,...,M\}. \tag{10}$$

我们说服务器对服务类 i 保证目标输出函数 $Z_i(\cdot)$,若对每个 t 有

$$R_i^{\text{out}}[\tau(t)+1,t]\ge Z_i(t), \tag{11}$$

其中

$$Z_i(t)=\min_{\substack{s:\tau(t)\le s\le t\\ Q_i[s]=0}}\{R_i^{\text{out}}[\tau(t)+1,s]+P^{D_i}(t-s)\}, \tag{12}$$

我们思考比例公平调度策略(PFS).考虑一种策略给每个到达分组分配一个满足式(11)的限期(deadline),并按限期的顺序发送分组.目标输出 $Z_i(t)$ 在 t 前是不可知的,但在不考虑暂存分组数清 0 时可以作一个估计.其中每个服务类的缓冲区大小和丢弃策略按照定理 2 设计.

定义 **3**. 按照式(9)计算和分配各服务类的缓冲区,根据缓冲空间和丢弃策略决定到达分组丢弃或入队,对进入服务器的每个分组分配一个限期,按持续工作方式有最早限期的分组先得到服务,特别地,在某个时间片 u 内,有 $\sum_{i=1}^{M}\left(Q_i[u-1]+R_i^{\text{in}}[u]\right)$ 个分组可以离开服务器,并有最多 c 个最早限期的分组在时间片 u 中被选中离开.服务类 i 中在时间片 u 到达的分组被分配如下的限期 D_i

$$D_i=\min\{t: t\ge u \quad \text{and} \quad Z_i(t;u-1)\ge n_i\}, \tag{13}$$

其中

$$Z_i(t;u)=\min_{\substack{s:\tau(u)\le s\le u\\ Q_i[s]=0}}\{R_i^{\text{out}}[\tau(u)+1,s]+P^{D_i}(t-s)\}. \tag{14}$$

n_i 为分组的到达序号.

注意, $Z_i(u;u)=Z_i(u)$ 为式(14)中的目标输出函数.另外, $Z_i(t;u)$ 是 $Z_i(t)$ 在时间片 u 的估计值.

4 模拟与分析

4.1 模拟实验环境

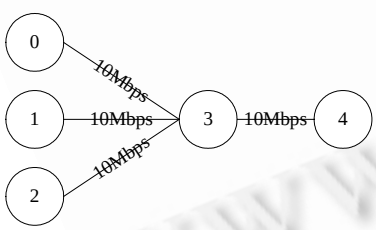

Fig.1 Topology in test
图 1 模拟实验拓扑图

本节通过扩展模拟实验评价 PFS 算法.所有模拟实验都运行在网络模拟系统 ns-2[13]上,该系统提供了在分组级别的网络协议、缓冲管理和调度算法的不同实现.我们将结合传输协议和流量模型考察 PFS 算法的行为特性.

我们的模拟网络的拓扑结构如图 1 所示.节点 0,1,2 通过 10Mbps 的链路与节点 3 相连,而节点 3 和节点 4 通过 10Mbps 链路连结.我们有 3 个服务类 0,1 和 2.从节点 0,1,2 中出来的数据流分别对应于上述的 3 个服务类,也就是说,从节点 i 出来的数据流为服务类 i.节点 3 中的缓冲区大小为 200K 字节.

4.2 模拟结果及分析

考察 3 个服务类数据流的网络模型.在我们的实验中,每个服务类包括一个或多个数据流,其分布和分配的带宽见表 1.音频(audio)流每 20ms 发送 160 字节分组,而视频(video)流每 33ms 发送 8K 字节的分组,其他数据流发送 4K 字节分组,而 FTP 数据流是持续发送的.

要描述 PFS 在延迟比例区分和降低实时应用延迟的能力,我们设音频数据流延迟参数为 5,给视频数据流设定延迟参数为 10,其他数据流延迟参数为 100.

Table 1 Traffics in the test

表 1 实验中的数据流

Class ID①	Flow type②	Assigned rate③(bps)	Interval④	Packet length⑤(Bytes)	Input⑥
0	Audio	64K	20ms	160	0
1	Video	2M	33ms	8K	1
	On/Off	5M	5s	4K	2
2	Poisson	2M	None	4K	2
	FTP	5M	None	4K	2

①服务类 ID,②流类型,③分配速率,④时间间隔,⑤分组长度,⑥入结点.

图 2 显示了各服务类数据流的延迟分布.从图中可以看出,PFS 算法能够较好地满足延迟的比例区分.其中延迟的周期性变化表现了 ON-OFF 数据流的影响,在 ON-OFF 数据流激活时各服务类的延迟均增加了,但音频流和视频流相应于其实时性要求增加的幅度较小.这反映了 PFS 算法能够根据网络拥塞情况实时调节各服务类的延迟,保证其比例公平性,同时我们看出,它对实时性要求较高的音频流和视频流也能在网络拥塞时提供优先的延迟性能.

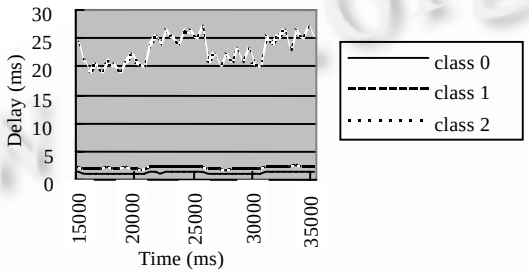

Fig.2 Traffic delay distribution in service classes

图 2 服务类数据流的延迟分布

5 结 论

本文研究了高速分组交换网络中调度策略的综合性能要求,提出了一种有效地综合网络效率、用户服务质量(QoS)要求和系统公平性等多性能目标的综合调度策略,并对比例公平调度策略进行了详细的分析论证.本文的主要贡献包括:(1) 提出了基于比例公平性原则的丢失率和延迟比例函数;(2) 提出了综合考虑分组延迟和丢失率的用户要求与系统公平性的比例公平调度策略,并给出对其详细的分析论证;(3) 分析并讨论了比例公平调度策略的可行性.由于多目标性能调度策略研究的复杂性,目前还缺乏有效的分组网络综合性能调度策略,本文的研究是对这方面研究的有益的理论探索,研究成果可应用于对分组网络调度策略的设计、实现和性能分析优化.

References:

[1] Cruz, R.L. A calculus for network delay, part I: network elements in isolation. IEEE Transactions on Information Theory, 1991,37(1):114~131.

[2] Cruz, R.L. A calculus for network delay, part II: network analysis. IEEE Transactions on Information Theory, 1991,37(1):132~141.

[3] Kurose, J. On computing per-session performance bounds in high-speed multi-hop computer networks. In: Proceedings of the ACM Sigmetrics and Performance'92. New York, 1992. 128~134.

[4] Yaron, O., Sidi, M. Performance and stability of communication networks via robust exponential bounds. IEEE/ACM Transactions on Networking, 1993,1(3):372~385.

[5] Chang, C.S. Stability, queue length, and delay of deterministic and stochastic queueing networks. IEEE Transactions on Automatic Control, 1994,39(5):913~931.

[6] Hurley, P., Boudec, J.L. A proposal for an asymmetric best-effort service. In: Proceedings of the IWQOS'99. London, 1999. 129~132.

[7] Pitsillides, A., Stylianou, G., *et al*. Bandwidth allocation for virtual paths (BAVP): investigation of performance of classical constrained and genetic algorithm based optimisation techniques. In: Proceedings of the INFOCOM 2000. Tel Aviv, 2000. 1379~1387.

[8] Dovrolis, C., Stiliadis, D. Relative differentiated services in the Internet: issues and mechanisms. In: Proceedings of the ACM SIGMETRICS'99. Atlata , 1999. 204~205.

[9] Jiang, Y. Lin, C., Wu, J. Integrated performance evaluating criteria for network traffic control. In: Proceedings of the IEEE Symposium on Computers and Communications 2001. Tunisia: IEEE Communications Society Press, 2001.

[10] Parekh, A.K., Gallager, R.G. A generalized processor sharing approach to flow control in integrated services networks: the single-node case. IEEE/ACM Transactions on Networking, 1993,1(3):344~357.

[11] Le, J-Y., *et al*. Connectionless data service in an ATM-Based customer premises network. Computer Networks and ISDN Systems, 1994,26(11):1409~1424.

[12] Agrawal, R., Rajan, R. Performance bonds for flow control protocols. IEEE/ACM Transactions on Networking, 1999,7(3):310~323.

[13] Ucb/lbnl/vint network simulator-ns (version 2). 2001.

A Packet Scheduling Strategy for Multiple Performance Objects*

JIANG Yong, WU Jian-ping

(*Department of Computer Science and Technology, Tsinghua University, Beijing* 100084, *China*)

E-mail: yong98@mails.tsinghua.edu.cn

http://netlab.cs.tsinghua.edu.cn

Abstract: The design of packet scheduling strategy and algorithm is one of the most important issues for the high-speed packet-switched networks. Because the packet scheduling strategy has multiple performance objects, how to reach multiple objects simultaneously is a difficult problem. Based on the proportional fairness principle, a proportional fairness scheduling (PFS) strategy in packet-switched networks is provided. The PFS integrates several objects, such as network performance, user's QoS requirement and system fairness. And the proposed strategy is analyzed and proved in detail. Moreover, the proportional fairness scheduling strategy can be applied to design and improve the packet scheduling strategy and algorithms in packet-switched networks.

Key words: proportional fairness principle; packet scheduling; QoS; fairness

* Received June 20, 2001; accepted January 14, 2002
Supported by the National Natural Science Foundation of China under Grant No.90104002