TCP 连接竞争资源的公平性研究

吴 杰,冯振明,袁 坚,刘文利
(清华大学电子工程系,北京 100084)

摘 要: Internet 的迅速发展使得计算机网络的资源分配成为数据通信研究领域的热点. 在 Internet 由单一服务质量过渡到支持多种服务质量的过程中,传输控制协议(TCP)将一直是主要的端到端的资源分配方案. 本文深入分析了具有不同传输延时的 TCP 连接竞争资源时的公平性. 研究表明,TCP 算法对具有较大传输延时的连接具有不公平性. 当竞争带宽的各连接的传输延时按比例减小时,小传输延时的连接对大传输延时的连接的抑制性相对减弱,这种不公平性将减弱. 结论对于 TCP 算法的改进具有重要意义,同时这种不公平性对于设计分级服务方案将有很好的应用前景.

关键词: 传输控制协议; 传输延时; 吞吐量; 公平性
中图分类号: TP393 **文献标识码**: A **文章编号**: 0372-2112 (2000) 08-0032-03

Study on the Fairness of TCP Connections Competing for Resources

WU Jie, FENG Zhen-ming, YUAN Jian, LIU Wen-li
(Department of Electronic Engineering, Tsinghua University Beijing 100084, China)

Abstract: The rapid growth of the Internet makes the research on resource allocation in computer networks a hot area in data communication research field. In the evolvement of the single quality of service to multiple levels of quality of service in Internet, the transmission control protocolb (TCP) will always be the main end-to-end resource allocation scheme. This paper studied the fairness of TCP connections with different propagation delays when competing for resources. It is showed that it is unfair for the connections with large propagation delays in TCP algorithm. When the propagation delays of the connections competing for bandwidth proportionally decrease, the unfairness becomes weaker. This conclusion will be very useful in the renovation of TCP algorithm. Meanwhile, the unfairness property has a promising interest to the design of differentiated services scheme.

Key words: TCP (Transmission Control Protocol); propagation delay; throughput; fairness

1 引言

目前的 Internet 中没有服务质量的保证,只许诺以单一的"Best-Effort"服务,但随着 Internet 的迅速发展和逐步商用化,提供多种不同服务质量的应用已经成为 Internet 发展的必然趋势. 如何在 Internet 中实现服务质量的分级和保证是当前数据通信研究领域十分关注的问题[1], IPv4[2]、Integrated Services/ RSVP[3,4]和 Differentiated Services[5,6]等多种 Internet 体系模型被相继提出. 一个合理的 Internet 的资源分配方案是实现服务质量分级和保证的关键技术. 目前 Internet 的资源分配主要是在用户端进行,即利用传输控制协议(如 TCP, Transmission Control Protocol)在用户端进行端到端的流量控制. TCP 控制算法主要包括四个子算法,即慢启动(Slow Start)、拥塞预防(Congestion Avoidance)、快速重传(Fast Retransmit)和快速恢复(Fast Recovery)[7,8]. 而在网络内部(即路由器处),虽然已经提出了 Fair Queueing[9]、Random Early Dropping[10]等方案,但还不能很好满足服务质量分级和保证的要求,而且其大规模应用还有相当长的时间. 由于普遍性和简洁、有效的控制算法,作为目前 Internet 的主要资源分配控制方案的 TCP 算法必将长期存在和发挥其作用. 在目前的 Internet 和将来支持服务质量分级和保证的 Internet 体系模型中,TCP 算法在分配资源时的公平性都是一个极为重要的问题. 文[11]用流体流的方法深入分析了两个连接经过一个路由器时窗口和吞吐量的动态特性,着重于讨论两个连接慢启动过程和拥塞预防过程的瞬态相互作用. 本文在文[12]的基础上,给出了一种多个连接抢占资源的稳态简化模型,对多个具有不同特性的连接抢占资源的稳态情况进行了仿真,并深入讨论了 TCP 算法中的公平性问题. 本文结论对于 TCP 算法的改进具有重要意义,同时对于设计分级服务方案将有很好的应用前景.

2 简化模型

本节给出了一种 TCP 算法简单的分析模型. 发送速度由发送端的拥塞窗口 cwnd(congestion window)和接收端的标定窗

收稿日期:1999-04-09;修回日期:1999-10-05

© 1994-2010 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

口 rwnd(receiver 's advertised window)的最小值决定,rwnd 的典型值是 64kB. 为了分析的简便,在模型中不考虑 rwnd 的影响,发送速度仅由 cwnd 决定. 这样 TCP 的慢启和拥塞预防算法可以描述如下:

```
Additive Increase
    every ACK do
      if cwnd < ssthresh then cwnd = cwnd + 1
      else cwnd = cwnd + 1/ cwnd
Multiplicative Decrease
    every Time-Out do
      ssthresh = 0.5 * cwnd
      cwnd = 1
```

如图 1 所示,假设有 K 个连接经过一个路由器,此路由器是这 K 个连接的共同瓶颈. 由于慢启过程远远短于拥塞预防过程,在以下对稳态的分析中略去了慢启动过程. 路由器的带宽为 μ,缓冲区大小为 B. 此 K 个连接用 $k = 1, 2, \cdots, K$ 表示,$W_k(t)$ 为连接 k 在时刻 t 对应的 cwnd 的大小,τ_k 为连接 k 的传输延时,$\lambda_k(t)$ 为连接 k 在时刻 t 的吞吐量大小,$RTTk(t)$ 为连接 k 在时刻 t 的往返时间. $X(t)$ 是此路由器缓冲区在时刻 t 的队长.

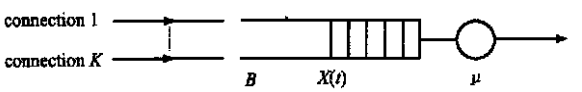


图 1　多连接共享一个路由器的简化模型

根据以上算法和假设,每个连接的窗口动态行为可以用以下方程刻划:

$$\begin{cases} \dfrac{dW_k}{dt} = \dfrac{1}{RTT_k(t)} \\ X(t) = B \Rightarrow W_k(t^+) = 0.5 \cdot W_k(t) \end{cases} \quad (1)$$

当 $X(t) > 0$ 时,连接 k 的往返时间 $RTT_k(t)$ 可以近似表示为:

$$RTT_k(t) = \tau_k + X(t)/\mu \quad (2)$$

在 $RTT_k(t)$ 时间内,连接 k 发出的数据量为 $W_k(t)$,故此时其吞吐量为:

$$\lambda_k(t) = W_k(t)/RTT_k(t) \quad (3)$$

又因为 $\sum_{k=1}^{K}\lambda_k = \mu$,于是可以得到下式:

$$\sum_{k=1}^{K} W_k(t)/(X(t)/\mu + \tau_k(t)) = \mu \quad (4)$$

根据此式,可以求取 $X(t)$.

当 $X(t) = 0$ 时,$RTT_k(t) = \tau_k$,此时有

$$\sum_{k=1}^{K}\frac{W_k}{\tau_k} \le \mu \quad (5)$$

反之,如果上式成立,也可以推出 $X(t) = 0$.

3　仿真结果

根据 TCP 算法和以上的简化模型对各连接的窗口和吞吐量的动态变化进行了大量的仿真实验,主要研究了具有不同传输延时的连接在竞争资源时的特点.

假设一个路由器参数为:μ = 2000packets/s, B = 120packets. 有 4 个连接共享此路由器,传输延时分别为 τ_1 = 10ms, τ_2 = 30ms, τ_3 = 400ms, τ_4 = 600ms,起始传输窗口为 $W_1(0) = 0$, $W_2(0)$ = 50packets, $W_3(0)$ = 150packets, $W_4(0)$ = 30packets. 通过仿真,得到 4 个连接的吞吐量的动态特性见图 2. 4 个连接的窗口的动态特性如图 3 所示.

其余条件不变,改变初始窗口大小,$W_1(0)$ = 20packets, $W_2(0)$ = 100packets, $W_3(0)$ = 80packets, $W_4(0)$ = 350packets. 仿真结果如图 4,可以看到,稳态吞吐量和初始窗口大小无关.

从图 2、图 4 中可以清楚地看到,传输延时分别为 10ms、30ms、400ms 和 600ms 的连接所获得的吞吐量分别约为 0.66μ、0.32μ、0.01μ 和 0.01μ. 也就是说,两个相对传输延时大的连接在 TCP 的控制下获得的吞吐量远远少于两个传输延时小的连接. 对窗口大小的动态仿真也可以得到类似的结论. 从以上的仿真结果可以看到,在其它条件相同的情况下,传输延时大的连接受到传输延时小的连接很强的抑制,得到的带宽非常少,TCP 对传输延时大的连接是很不公平的.

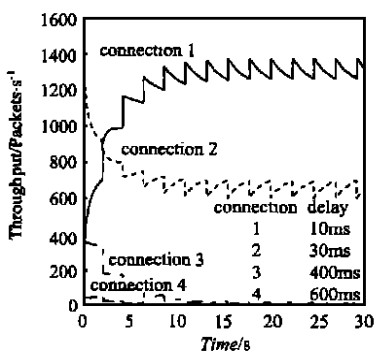


图 2　4 个连接的吞吐量的动态特性曲线

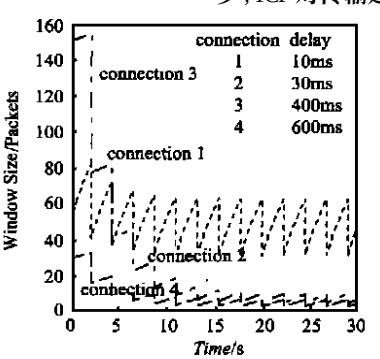


图 3　4 个连接的发送窗口的动态特性曲线

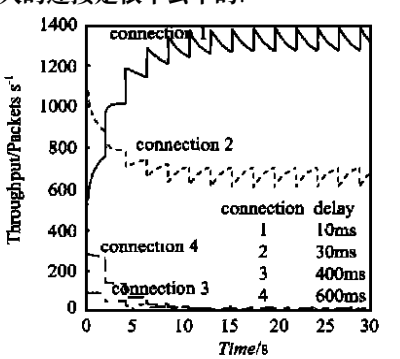


图 4　初始窗口变化时 4 个连接的吞吐量的动态特性曲线

下面将 4 个连接的传输延时分别降为原来的 1/10,即为 1ms、2ms、40ms 和 60ms. 其余条件不变,其仿真结果如图 5.

由图 5 中可见,传输延时分别为 1ms、3ms、40ms 和 60ms 的连接所获得的吞吐量分别为 0.405μ、0.395μ、0.105μ 和 0.095μ. 尽管各连接传输延时的相对比值不变,随着传输延时绝对值的减小,大的传输延时对应的连接获得的吞吐量相对增大. 换句话说,在竞争资源的各连接的传输延时都较小的情况下,传输延时小的连接对传输延时大的连接的抑制性相对减弱,TCP 对具有长传输延时的连接的不公平性减弱.

© 1994-2010 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

4 理论分析

首先考察传输延时为:(1) 10ms、30ms、400ms 和 600ms(图 6) (2) 1ms、3ms、40ms 和 60ms(图 7) 两种情况下缓冲区队列长度的动态变化情况.

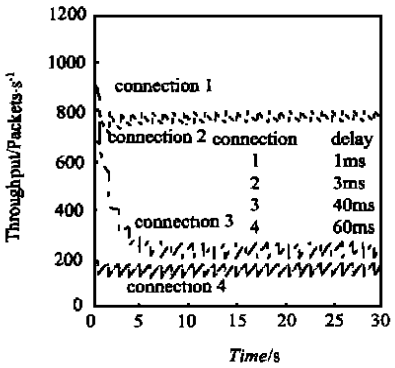


图 5 传输延时降低后的 4 个连接的吞吐量的动态特性曲线

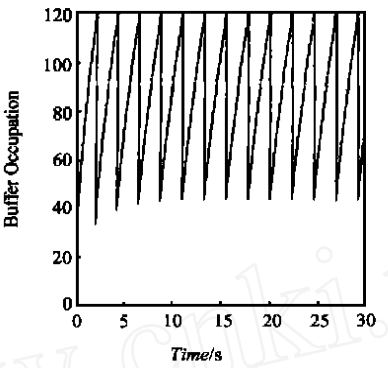


图 6 大传输延时下缓冲区中队长的动态变化曲线

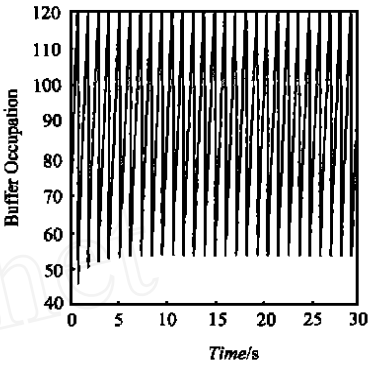


图 7 小传输延时下缓冲区中队长的动态变化曲线

由图 6、图 7 中可以发现,在两种情况下,缓冲区中队长的动态范围大致相同,均为约 50～120packets,因此 $X(t)/\mu$ 的范围是约为 0.025s～0.06s,为以下叙述方便,不妨取 $X(t)/\mu$ 为 0.031. 根据式(2)进行估算,见表 1、表 2.

表 1 大传输延时下各连接 RTT 的估算表

	Connection 1	Connection 2	Connection 3	Connection 4
$\tau(s)$	0.01	0.03	0.4	0.6
$X/\mu(s)$	0.031	0.031	0.031	0.031
$RTT(s)$	0.041	0.061	0.431	0.631

表 2 小传输延时下各连接 RTT 的估算表

	Connection 1	Connection 2	Connection 3	Connection 4
$\tau(s)$	0.001	0.003	0.04	0.06
$X/\mu(s)$	0.031	0.031	0.031	0.031
$RTT(s)$	0.032	0.034	0.071	0.091

从表中可以看到,随着 τ_k 的减少,$X(t)/\mu$ 对 $RTT_k(t)$ 的影响增大,使得各连接的 $RTT_k(t)$ 之间的差距明显减小. 而由式(1)和式(3)可知,$RTT_k(t)$ 是决定吞吐量的主要因素,$RTT_k(t)$ 小的连接的吞吐量大. 因此,尽管各连接传输延时的比值一定,当它们的绝对值减小时,各连接对应的 RTT 将逐渐接近,使得其中具有长传输延时对应的连接的吞吐量增大. 也就是说,在竞争资源的各连接的传输延时都较小的情况下,TCP 对具有长传输延时的连接的不公平性将减弱.

下面讨论两种特殊情况.

(1) 如果对于 $k=1,2,\cdots,K$,有 $\tau_k \ll X/\mu$,由式(2)知,此时所有连接的 RTT 近似相等,都等于 X/μ. 因此,在稳态时,各连接的对应的发送窗口大小近似相等. 于是有:

$$\forall l,m=1,2,\cdots,K,\frac{\lambda_l(t)}{\lambda_m(t)}\approx\frac{RTT_m(t)\cdot W_l(t)}{RTT_l(t)\cdot W_m(t)}\approx 1 \qquad (6)$$

即各连接对应的吞吐量近似相等.

(2) 如果对于 $k=1,2,\cdots,K$,有 $\tau_k \gg X/\mu$,由式(2)知,此时连接 k 的 RTT_k 近似等于各自的传输延时 τ_k,这样在稳态时,有:

$$\forall l,m=1,2,\cdots,K,\frac{W_l(t)}{W_m(t)}\approx\frac{RTT_m(t)}{RTT_l(t)}\approx\frac{\tau_m}{\tau_l} \qquad (7)$$

则有 $$\frac{\lambda_l}{\lambda_m}\approx\frac{RTT_m(t)\cdot W_l(t)}{RTT_l(t)\cdot W_m(t)}\approx\left(\frac{\tau_m}{\tau_l}\right)^2 \qquad (8)$$

即各连接对应的吞吐量近似与传输延时的平方呈反比.

5 结论

从以上对 TCP 的基于一种流体流模型的仿真和理论分析的结果可以看到,TCP 算法对传输延时较大的连接具有不公平性,但当竞争带宽的连接的传输延时按比例减小时,这种不公平性将减弱. 这增大了在 Internet 中对具有不同传输延时的连接实现服务质量的分级和保证的难度,而在同一个局域网内部或 ISP 内部,由于传输延时较小且差距不大,保证服务质量相对容易一些,这说明在 Internet 从单一服务体系过渡到多级服务体系的过程中,首先在一个局域网内部或 ISP 内部实现服务质量是很好的过渡方案.

通过对 TCP 算法的剖析可以看到,由于 TCP 算法是一个端到端的算法,RTT 对于连接的发送窗口和吞吐量起着决定性的作用,所以这种不公平性是 TCP 算法的一种内在属性. 这一结论对于 TCP 算法的改进具有重要意义. 如对具有长传输延时的连接,可以使其发送窗口的增长速度相对更大,使其能够得到更多的带宽. 利用这种不公平性,如果在路由器调整排队延时(如对优先级低的连接采取增大其排队延时的方法),就可以直接利用 TCP 来实现一定程度的服务质量的分级. 这是下一步的研究方向.

作者简介:

吴 杰 1997 年本科毕业于清华大学电子工程系,现在为清华大学电子工程系硕士生,主要从事宽带网络性能分析、流量控制、资源分配等方面的研究工作.

冯振明 清华大学电子工程系副教授. 1970 年毕业于清华大学无线电与电子学系. 1982 年在清华大学电子工程系获硕士学位,多年来一直从事信号处理方面的教学与科研工作,从 1995 年至今主要从事 HFC 网络的多媒体通信方面的研究工作.

(下转第 31 页)

© 1994-2010 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

[2] David J. DeWitt ,Jeffrey F. Naughton ,John C. Shafer ,and Shivakumar Venkataraman. Parsets for parallelizing OODBMS traversals: Implementation and performance [A]. In Proceedings of the Third International Conference on Parallel and Distributed Information Systems [C] ,1994 , 111 - 120.

[3] Russell F. Haddleton. An implementation of a parallel object oriented database system [R]. Technical Report ,University of Virginia ,1995.

[4] Russell F. Haddleton. Parallel set operations in complex object oriented queries [D]. PhD thesis ,University of Virginia ,1998.

[5] N. Bassiliades , I. Vlahavas. Hierarchical query execution in a parallel object-oriented database system [J]. Parallel Computing ,1996 ,22(3) : 1017 - 1048.

[6] Michael Gesman. Mapping a Parallel Complex-Object DBMS to Operating System Processes [A]. In Proceedings of the Second International Euro-Par Conference [C]. Euro-Par '96 ,1996 ,852 - 861.

[7] Daniel F. Lieuwen ,Daoid J. Dewitt ,and Manish Mehta. Parallel Pointer-based Join Techniques for Object-Oriented Databases [J]. In Proc. of the Second Int 'l Conf. on Parallel and Distributed Information Systems ,January 1993 :172 - 181.

[8] Leung ,C. H. c and Taniar ,D. Parallel Query Processing in Object-Oriented Databases Systems [A]. Proceedings of the Sixth Austrakasian Database Conference (ADC '95) [C]. Adelaide , Australia ,January 1995 :119 - 131.

[9] 王意洁. 面向对象数据库的并行查询处理与事务管理 [D]. 博士学位论文 ,国防科技大学研究生院 ,1998 ,10.

[10] 王意洁 ,王勇军 ,胡守仁. 并行面向对象数据库中的基于对象类的混合式数据放置策略 [J]. 计算机学报 ,1998 ,21 (增刊) :178 - 183.

[11] 何炎祥 ,郑振楣 ,石树刚. 面向对象数据库 [M]. 武汉大学出版社 ,1995 ,6.

作者简介:



王意洁　1971 年 9 月生 ,1998 年获博士学位 ,现为国防科技大学计算机学院并行与分布处理国家重点实验室助理研究员 ,研究领域为数据库和并行分布处理技术.



王勇军　1971 年 1 月生 ,1998 年获博士学位 ,现为国防科技大学计算机学院网络与信息安全研究所助理研究员 ,研究领域为虚拟现实、数据库和并行分布处理技术.

(上接第 34 页)

参考文献:

[1] Lefelhocz C ,Lyles B ,Shenker S ,et al. Congestion control for best-effort service: Why we need a new paradigm [J/OL]. IEEE Network ,1996 , January/ February: 10 - 19.

[2] Almquist P. Type of service in the internet protocol suite [J/OL]. Internet RFC 1349 ,July 1992.

[3] Braden R. Integrated services in the internet architecture: An overview [J/OL]. Internet RFC1633 ,July 1994.

[4] Braden B. Resource ReServation Protocol (RSVP)-Version 1 functional specification [J/OL]. Internet RFC 2205 ,September 1997.

[5] Blake S ,Black D ,Carlson M ,et al. An architecture for differentiated services [DB/OL]. Internet Draft < draft-ietf-diffserv-arch-02. txt > . October 1998.

[6] Black S ,Bernet Y ,Binder J ,et al. A Framework for Differentiated Services [DB/OL]. Internet Draft < draft-ietf-diffserv-framework-01. txt > . October 1998.

[7] Jacobson V , Kerals M. Congestion avoidance and control [A]. In Proceedings of SIGCOMM '88 [C] ,Stanford ,CA ,Aug. 1998 ,ACM.

[8] Allman M ,Paxson V ,Stevens W ,et al. TCP Congestion Control [DB/OL]. Internet Draft < draft-ietf-tcpimpl-cong-control-03. txt > . Decemrber ,1998.

[9] Demers A , Keshav S , Shenker S. Analysis and simulation of a fair queueing algorithm [A]. In Proc. ACM SIGCOMM '89 [C] ,1989 :3 - 12.

[10] Floyd S ,Jacobson V. Random Early Detection Gateways for Congestion Avoidance [J]. IEEE/ ACM Transactions on Networking ,1998 ,1 (4) : 397 - 413.

[11] Alt-Hellal O ,Altman E , Elouadghiri D ,et al. Performance of TCP/ IP: the case of Controlled Sources [A]. International Conference for Comrputer Communications-ICCC '97 [C] , Cannes , France , November 19-21 ,1997

[12] Bonald T. Comparion of TCP Reno and TCP Vegas via Fluid Approximation [R]. In Workshop on the modeling of TCP ,December 1998.

© 1994-2010 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net