高速网络拥塞控制协议 VCP 的研究

邢国稳 薛胜军
(南京信息工程大学计算机与软件学院 南京 210044)

摘 要 互联网正在逐步进入一种高带宽延时积的高速网络时代。当网络的带宽或者时延增大时,TCP协议的性能严重下降,最显著的就是网络瓶颈处带宽利用率很低。在高速拥塞控制方面比较理想的XCP协议却存在部署方面的问题。变结构拥塞控制协议(VCP)可有效地解决上述问题。VCP协议联合使用ECN机制的两个二进制来编码拥塞信息。根据来自接收端的拥塞信息,VCP协议的发送端选择控制算法来响应拥塞信号。仿真实验表明VCP协议与TCP协议、XCP协议相比不仅具有较高的链路利用率,并且对现有的协议改动非常小,有利于逐步地实施。

关键词 拥塞控制,高带宽时延积,明确拥塞通知,变结构拥塞控制协议
中图法分类号 TP393 **文献标识码** A

Study on High-speed Network Congestion Control Protocol

XING Guo-wen XUE Sheng-jun
(Computer and Software Institute, Nanjing University of Information Science & Technology, Nanjing 210044, China)

Abstract The Internet is shifting to the age of high-speed network, which is called as High Bandwidth-Delay product (BDP) network. TCP's performance degrades significantly as either bandwidth or latency increases. The most noteworthy aspect of this poor performance is the low bandwidth utilization in bottleneck. And XCP protocol has some applied problems. Variable-structure Congestion control Protocol (VCP) addresses this problem efficiently. VCP protocol encodes congestion information with two binary bits in ECN. According to congestion information from receiver, the sender chooses control arithmetic to reply congestion signals. The paper indicates VCP protocol has high link usage, and it changes present protocol a bit. This is helpful to put VCP into practice.

Keywords Congestion control, High bandwidth-delay product, ECN, VCP

1 VCP 协议产生背景

当网络中存在过多的数据包时,网络的性能就会下降,这种现象称为拥塞。拥塞控制机制实际上包含拥塞避免和拥塞控制两种策略。拥塞避免是一种预防措施,维持网络的高吞吐量、低延迟状态,避免进入拥塞;拥塞控制是一种恢复措施,使网络从拥塞中恢复过来,进入正常的运行状态。拥塞虽然是由于网络资源的稀缺引起的,但单纯增加资源并不能避免拥塞的发生。例如增加缓存空间到一定程度时,只会加重拥塞,而不是减轻拥塞,这是因为当数据包经过长时间排队完成转发时,它们很可能早已超时,从而引起源端超时重发,而这些数据还会继续传输到下一路由器,从而浪费网络资源,加重网络拥塞。单纯地增加网络资源之所以不能解决拥塞问题,是因为拥塞本身是一个动态问题,它不可能只靠静态的方案来解决,而需要协议能够在网络出现拥塞时保护网络的正常运行。

新的协议的主要思想大体可以归为两大类:(1)改进当前的TCP协议,(2)终端与路由器辅助联合的拥塞控制协议。第一类的主要思想是考虑与当前的TCP共存并兼容,只需要在终端(发送端、接收端)进行适当的修改以适应高速网络。这一类的工作包括:High-Speed TCP(HSTCP),Fast TCP,BIC,TCP-RAB以及Scalable TCP(STCP)。另一类协议的主要思路是终端的拥塞控制与路由器辅助进行联合的全新设计,并且新的协议能与当前的TCP协议并存[3],如图1所示。

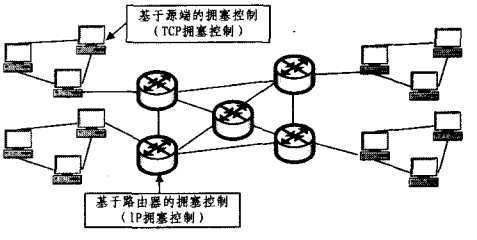

图1 互联网上的拥塞控制

1994年Floyd提出将明确拥塞通知(explicit congestion notification,ECN)应用在TCP/IP中。每个路由器都使用自己的主动式队列管理(active queue management,AQM)策略来实现ECN机制[5]。2002年Katabi等提出明确控制协议(explicit control protocol,XCP)。XCP扩展了ECN机制,另外它还引入了一个新的将效率控制与公平控制分离的概念。

到稿日期:2008-04-10 本文受国家自然科学基金(60572015)资助。
邢国稳 女,讲师,研究方向为高速网络通信、实时控制与拥塞控制,E-mail:xgw@nuist.edu.cn;薛胜军 男,教授,博士生导师,研究方向为高速网络通信、网格计算。

XCP 不仅在传统的网络下性能比 TCP 好，而且在大的 BDP (High Bandwidth Delay Product)网络下也保持了高效性、公平性、可扩展性以及稳定性[2]。但是，为了实现运行在路由器的效率，控制器和公平控制器与终端之间交换拥塞反馈信息需要大量的二进制位，而 IP 报文头无法提供足够的二进制空间。这影响了 XCP 的部署和推广应用。2005 年 Xia 等人提出可变结构拥塞控制协议(variable-structure congestion control protocol，VCP)。VCP 使用 ECN 的两个二进制位交换网络拥塞反馈信息，同时取得了与 XCP 相类似的性能[1]。

2 ECN 概述

2.1 ECN 在 TCP/IP 包头的设置

ECN 需要在 IP 包头设置一个两位(bit)的 ECN 域，一个是 ECT(ECN-Capable Transport)位，由源端设置以显示源端节点的传输协议是支持 ECN 的；另一个是 CE(Congestion Experienced)位，由路由器设置，以显示是否发生了拥塞。IPv4 中 TOS 字节的第 6 位被设置为 ECT 位，第 7 位被设置为 CE 位。IPv4 中 TOS 字节和 IPv6 中的流类型字节(traffic class octet)是相对应的，它们的前 6 位被设置为区分服务(Different-iated Services)中的区分服务标记域(DS field)。后两位保留未用，因此可用来作为 ECN 域。除了在 IP 头中设置 ECN 域外，ECN 还需要传输协议的支持。在 TCP 头中需要设置两个标志位：ECN-Echo 和 CWR(Congestion Window Reduced)。ECN-Echo 是目的端，用来通知源端收到一个 CE 包；CWR 是源端，用来通知目的端拥塞窗口已减小[4]。

2.2 VCP 协议中的 ECN 编码

VCP 协议联合使用 ECN 机制的两个二进制来编码拥塞信息。根据来自接收端的拥塞信息，VCP 协议的发送端选择 MI，AI 和 MD 其中的一个拥塞窗口控制算法来响应拥塞信号。两个二进制位可以表达 4 种拥塞级别。在 VCP 协议中，00，01，10，11 二进制信息分别解释为：ECN 无效、低拥塞级别、高拥塞级别和拥塞发生。从低拥塞级别到高拥塞级别的拥塞跃迁点设置为 80%。

3 VCP 协议的工作原理

VCP 协议的设计是根据两个方针：1)效率控制与公平控制分离。路由器负责计算拥塞指数，终端选用积性增加(multiplicative increase，MI)、加性增加(additive increase，AI)两个算法中的一个来增加拥塞窗口。*MI* 和 *AI* 都是以拥塞指数为变量的函数。VCP 协议将链路利用率分为不同的负载区间，并选择适合某一区间的控制方法。当网络利用率低的时候，VCP 的目标是提高效率而不是公平性。另一方面，当利用率很高时，相对效率控制，VCP 给公平性控制更高的优先级。2)使用链路的负载指数作为拥塞信号。XCP 协议使用剩余带宽作为拥塞度的一个衡量，而 VCP 协议采用负载指数作为拥塞信号，比如：带宽的需求与链路带宽容量的比率。

简要地说，VCP 协议选择下面的 3 个区间编码负载指数 ρ_t：低负载区间：

简要地说，VCP 协议选择下面的 3 个区间编码负载指数 ρ_t：低负载区间：$\hat{\rho}_t=80\%$，当 $\rho_t\in[0\%,80\%]$；高负载区间：$\hat{\rho}_t=100\%$，当 $\rho_t\in(80\%,100\%)$；超载区间：$\hat{\rho}_t>100\%$，当 $\rho_t\in(100\%,+\infty)$；并负载指数的拥塞跃迁点为 80%。每一个时间周期 t_p 路由器估计它的出口链路 ι 的负载指数 ρ_t，根据下面的公式

$$\rho_t=\frac{\lambda_t+\kappa_q\cdot\tilde{q}_t}{\gamma_t\cdot c_t\cdot t_p} \tag{1}$$

其中：λ_t 为在 t_p 时间内输入传输流的总量，$\tilde{q}_t$ 为在这一时间内持久性的队列长度，k_q 为控制多快清空队列的参数，γ_t 为链路的目标利用率，c_t 为链路的带宽。在时刻 t，一个 VCP 传输流的发送端根据从网络中反馈回的负载指数执行以下 3 个动作中的其中一个：

$$\begin{aligned}&MI:W(t+rtt)=W(t)\times(1+\xi(\hat{\rho}_t))\\&AI:W(t+rtt)=W(t)+\alpha\\&MD:W(t+\delta_t)=W(t)\times\beta\end{aligned} \tag{2}$$

其中：

$\zeta(\hat{\rho}_t)=\kappa\cdot(1-\hat{\rho}_t)/\hat{\rho}_t$

$W(t)$为 t 时刻拥塞窗口的大小。

如果负载指数属于低负载区间，发送端采用 MI 算法；如果负载指数属于高负载区间，发送端采用 AI 算法；如果负载指数属于超载区间，发送端采用 MD 算法。为了减少系统往返时间 RTT(round trip time)对公平性的影响，调整 MI/AI 的参数：

$$\begin{aligned}&MI:\zeta_S\leftarrow(1+\xi)^{\frac{rtt}{t_p}}-1\\&AI:a_s\leftarrow a\cdot\frac{rtt}{t_p}\end{aligned} \tag{3}$$

在许多主动队列管理算法(AQM)中，随着网络环境的不同，如带宽 *rtt* 的变化，甚至网络流量的变化，必须调整控制参数的值以达到控制目标。在 VCP 中，无论任何网络环境下，都使用上述的常量参数。从控制理论的角度来说，对于变化的网络环境，VCP 协议有较强的鲁棒性[5]。

在 VCP 协议中，路由器将网络的拥塞程度分为：低负载、高负载、过载 3 个区域。原主机根据拥塞程度的不同，分别执行下列算法：低负载区域执行 MI(Multiplicative Increase)、高负载区域执行 AI(Additive Increase)、过载区域执行 MD (Multiplicative Decrease)。在低负载区域 MI 算法能够以指数增长速度提高带宽利用率；而在高负载区域以上，AIMD 算法则可以提供数据流之间的公平性[6]。

4 仿真分析

VCP 协议联合使用 ECN 机制的两个二进制来编码拥塞信息。根据来自接收端的拥塞信息，VCP 协议的发送端选择 MI，AI 和 MD 其中的一个拥塞窗口控制算法来响应拥塞信号。两个二进制位可以表达 4 种拥塞级别。在 VCP 协议中，00，01，10，11 二进制信息分别解释为：ECN 无效、低拥塞级别、高拥塞级别和拥塞发生。从低拥塞级别到高拥塞级别的拥塞跃迁点设置为 80%。

考虑一种情况，瓶颈链路的负载指数超过 80%时，所有的 VCP 传输流都进入 AI 阶段。当有新的传输流启动时，该新的传输流也进入 AI 阶段。它的拥塞窗口是每一个 RTT 时间内增加一个报文包。它无法通过 MI 阶段快速增加它的拥塞窗口。当瓶颈链路发生拥塞时，所有传输流的发送端通过来自它们各自接收者的确认包得知链路发生拥塞的信息。紧接着，它们进入积性减少(Multiplicative Decrease，MD)阶段。从此以后，所有流不断更替地经历 AI 和 MD 阶段。因此，一个新的传输流需要较长的时间增加它的拥塞窗口来得

到它应得的公平分配的带宽。

用仿真软件 NS-2 来试验 RTT 对于链路利用率的影响如图 2 所示。采用一个单瓶颈的网络拓扑结构，设置瓶颈链路的带宽为 160Mbps，RTT 从 1ms 到 1500ms。通过图 2 可以看出 VCP 协议在大多数情况下，瓶颈链路的利用率保持在 90%左右。

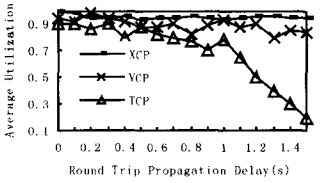

图 2　RTT 变化时 3 种协议的性能

RTT 非常小的时候比如 1ms，由于 VCP 的 RTT 参数的比例调整对于小的 RTT 值敏感导致了平均队列长度增加到大约为缓冲区大小的 15%，最大值为缓冲区大小的 45%。在 RTT 大于 800ms 时，VCP 协议的链路利用率为 85%～94%，这是由于路由器端的负载计算间隔为 200ms 与流的 RTT 值相比太小的原因。在所有情况下，VCP 协议的丢包率为 0。

结束语　VCP 协议联合使用 ECN 机制的两个二进制来编码拥塞信息。根据来自接收端的拥塞信息，VCP 协议的发送端选择 MI，AI 和 MD 其中的一个拥塞窗口控制算法来响应拥塞信号。通过原理与仿真分析，我们发现 VCP 协议在体系结构上是路由器检测拥塞，源主机调整算法，VCP 协议在高的链路利用率，小的队列长度和丢包方面的性能接近于 XCP。同时也要指出 VCP 协议收敛速度比较慢，这个问题有待进一步完善解决。

参 考 文 献

[1] Xia Y, Subramanian L, Stoica I, et al. One More Bit is Enough// Proceedings ACM SIGCOMM'05. Auguest 2005

[2] Katai D, Handley M, Rohrs C. Congestion Control for High Bandwidth-delay Product Networks. ACM SIGCOMM Computer Communications Review, 2002, 32(4): 89-102

[3] Golestani S J, Sabnani K. Fundamental Observations on Multicast Congestion Control in the Internet. INFOCOM, New York, NY, Mar. 1999

[4] Ramakrishnan K, Floyd S. A Proposal to add Explicit Congestion Notification (ECN) to IP. IETF RFC2481. Jan. 1999

[5] Floyd S, Fall K. Promoting the use of End - to - End congestion control in the Internet. IEEE/ACM Transaction on Networking, 1997

[6] Yang Y, Lam S. General AIMD Congestion Control// ICNP'00. November 2000

[7] Speakman T, et al. PGM Reliable Transport Protocol Specification. IETF RFC 3208. Dec. 2001

(上接第 41 页)

最为明显。这是因为引入了 BACnet 网络节点拥塞控制机制和 BACnet 路由器主动拥塞控制算法。BACnet 网络节点在检测到报文丢失率达到一定阀值的时候，开始主动降低报文的发送速率；BACnet 路由器在拥塞情况发生的时候，主动发送 Router-Busy-To-Network 报文和 Reject-Message- To-Network 报文，降低从其它路由器转发过来的报文的速率，从而减轻了 BACnet 路由器的负荷，加快了拥塞恢复的进度，降低了 BACnet 报文的丢失率。

3）拥塞控制策略对网络吞吐量的改善

第三组实验显示了拥塞控制策略对网络吞吐量的改善，实验结果如图 7 所示。

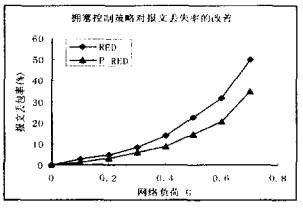

图 6　拥塞控制策略对报文丢失率的改善

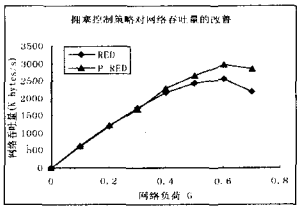

图 7　拥塞控制策略对网络吞吐量的改善

从图中可以看到，当 G 的值在 0～0.3 之间的时候，这两种拥塞控制算法下的网络吞吐量大致相等，随着网络负荷 G 的增大，都是呈上升趋势。当 G 的值逐渐增大到 0.3 以后，基于 P-RED 的新拥塞控制策略对网络吞吐量的改善逐渐明显；当 G 的值增大到 0.6 以后，网络吞吐量开始下降。这是因为当 G 的值大于 0.3 时，该拥塞控制策略对报文丢失率的改善开始变得较为明显；当 G 的值大于 0.6 时，报文丢失率变得越来越大，很多报文都被丢弃了，所以网络负荷开始逐渐减小。

结束语　基于 NS2 的 BACnet 网络拥塞控制策略，从路由器和 BACnet 网络节点两个方面入手，引入了基于 BACnet 协议端到端的流量控制方法，提出了基于 RED 的 BACnet 路由器拥塞控制算法。在保证 BACnet 网络吞吐量的同时，尽可能地提高 BACnet 网络的实时性和可靠性。这不但很好地解决了 BACnet 路由器的拥塞控制问题，而且对于研究其它面向无连接协议的拥塞控制机制有着重要的启示作用。

参 考 文 献

[1] Liu Quan, Ren Ping. Study on the Congestion Control Arithmetic of BACnet Routers// ICIEA2008. Singapore, 2008

[2] Tae Jin P, Won Seok S, Seung Ho H. Experimental performance evaluation of BACnet MS/TP protocol. International Journal of Control, Automation and Systems, 2007, 5(5): 584-593

[3] Wu Shubin, Liu Xiande, Wang Zhongming. Congestion Control in BACnet Networks Based on RPRED Algorithm. SPIE-The International Society for Optical Engineering, 2005, 6022 I: 602218

[4] 李方敏，周祖德，彭小兵，等. 区分服务环境下 TCP 拥塞控制机制研究. 系统仿真学报，2003，15(6)：832-836

[5] Analoui M, Jamali S. Congestion control in the internet: Inspiration from balanced food chains in the nature. Journal of Network and Systems Management, 2008, 16(1): 1-10

[6] 唐伟，郭伟，苏俭. 一种适用于 Ad Hoc 网络的拥塞控制算法. 计算机科学，2005，32(10)：41-43

[7] 王庆辉，潘学松，王光兴. 基于带宽估计的 ad hoc 网络拥塞控制机制. 通信学报，2006，27(4)：42-48

[8] 徐跃东，关治洪，王华. 基于仿真的 TCP 拥塞控制研究. 计算机工程，2004，30(23)：85-86