# 显式流量控制协议 XCP 研究*)

宋 军 曹建秋 杨 林
(重庆交通学院 重庆 400074)

**摘 要** 显式流量控制协议(XCP)能提高网络带宽利用率并改善带宽分配的公平性,本文分析了 TCP 在带宽时延乘积较大时所存在的性能问题,说明了 XCP 体系结构和流量控制规则,并对 XCP 与 TCP 的带宽时延乘积方面进行了性能比较。

**关键词** 带宽利用率,带宽分配公平性,XCP,带宽时延乘积

## Study on the Explicit Control Protocol

SONG Jun CAO Jian-Qiu YANG Lin
(Chongqing Jiaotong University, Chongqing400074)

**Abstract** Explicit Control Protocol(XCP) improves the utilization rate of network bandwidth and fairness of bandwidth allocation through an explicit control on network traffic. This paper analyzes the performance problems of TCP as product of bandwidth and latency increase, specifies the architecture and traffic control policy of XCP, and compares its performance to TCP in the networks with high bandwidth-delay product.

**Keywords** Bandwidth utilization rate, Fairness of bandwidth allocation, XCP, Bandwidth-Delay product

TCP 是互联网最基本的端到端数据传输控制协议,其拥塞控制机制为主机之间高效、可靠的数据传输提供了保证[1]。随着有线电视 HFC 网络、ADSL、地面和卫星无线网络等宽带接入技术的广泛应用,下一代互联网中传输链路的类型日趋多样化,既有高带宽的链路,也有时延较大、丢包率较高的链路,整个网络环境呈现带宽时延乘积(bandwidth-delay product,BDP)较大的特点。理论和实践证明[2~6],当网络 BDP 增加时,TCP 协议的性能将严重下降且不稳定,难以发挥其原有的效能,成为实施各种新型网络应用的巨大障碍。

显式流量控制协议(eXplicit Control Protocol,XCP)是一种基于滑动窗口机制的网络流量控制协议。XCP 在 TCP 执行传统拥塞控制的基础上对网络流量进行显式控制,其目标是在各种网络环境中(特别是 BDP 较大的网络),为各种数据流公平地分配网络带宽资源,使其保持高效稳定的吞吐量,提高网络带宽资源的利用率。

## 1 TCP 拥塞控制存在的问题

TCP 是一种可靠传输协议,为各主机之间提供可靠按序的数据传输服务。在拥塞避免阶段,TCP 采用和式增加积式减少(AIMD)的拥塞控制算法,即发送端每接收到一个确认包就将拥塞窗口长度增加 1,当发送端检测到数据包丢失则认为网络发生拥塞,将当前的拥塞窗口长度减半甚至将拥塞窗口长度置 1,以降低发送速率。然而在有线电视 HFC 网、无线局域网等 BDP 较大的网络中,TCP 的这种拥塞控制机制存在着以下三个主要问题:

· 带宽利用率极低[7,8]。随着网络带宽增加,发送端将不断增大其数据发送速度,导致节点路由器的队列长度较大且频繁振荡,增大了数据包的转发时延和丢包概率,造成发送端数据吞吐量振荡频繁;随着 RTT 不断增加,发送端拥塞窗口增长缓慢导致数据吞吐量下降,当 RTT 远大于超时重传时间,由于确认信息超时,发送端认为网络发生拥塞而成倍地降低数据吞吐量。因此,在 BDP 较大的网络环境中,TCP 数据流难以保持较高吞吐量和充分利用网络带宽资源。如图 1 所示,当网络带宽超过 2Gbps 后,TCP 的平均带宽利用率不足 50%;当 RTT 大于 1 秒后,TCP 的平均带宽利用率急剧下降至不足 40%。

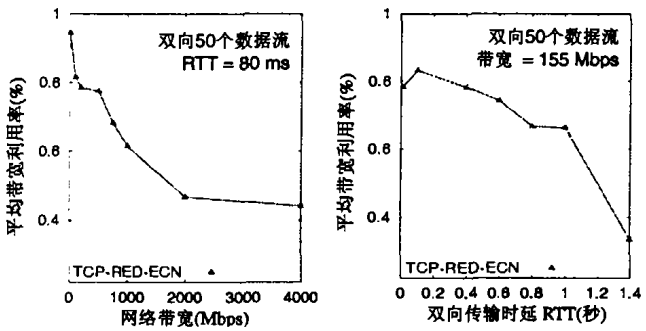



图 1 TCP 的网络带宽利用率

· 带宽分配不公平[9]。当多个数据流共享同一网络带宽资源时,RTT 较大的数据流其拥塞窗口增长速度非常慢,RTT 较小的数据流将迅速占用绝大部分可用带宽资源。因此,有线电视 HFC 网、无线局域网中 BDP 较大的数据流在经过骨干网时,无法与高速以太网中 BDP 较小的数据流公平地共享骨干网带宽资源。

· 拥塞控制混乱[10]。有线电视 HFC 网、无线局域网等网络较高的信道误码率将造成数据包丢失,TCP 发送端会将其错误地解释为网络发生拥塞,从而导致发送端进行不必要的拥塞控制响应。非拥塞数据包丢失对相关数据流进行的流量惩罚将进一步降低网络带宽资源的利用率,并使数据流之间

*)重庆交通学院博士基金项目(2004-2-06)。宋 军 博士,主要从事 MAC 协议、网络 QoS 和智能交通系统研究。

的带宽分配不公平现象加剧。

## 2 XCP 协议体系结构[11,12]

XCP 的基本思想是根据各个数据流申请的数据流量，以预留方式将数据流通路中瓶颈链路的剩余带宽资源公平地分配给每个数据流，并通过数据包的 XCP 报文头将流量控制信息显示地反馈给数据发送端，从而达到有效控制网络数据流量、防止网络发生拥塞、提高网络带宽利用率之目的。为提高可扩展性，各个数据流的状态信息也由每个数据包的 XCP 报文头携带，节点路由器不需要保存和维护这些信息。

### 2.1 XCP 报文头格式

每个数据包都携带了一个 XCP 流量控制报文头，用于发送端与节点路由器之间交换数据流状态信息和流量控制信息，其格式如图 2 所示。


图 2 XCP 报文头格式

XCP 报文头由 9 个字段组成，其中用于网络流量控制与信息交换的字段分别为：RTT 字段表示发送端测量到的数据流双向传输时延(最小值为 1 毫秒)，如果该字段的值为 0，则表示数据传送刚开始，发送端还未测量出数据流的双向传输时延；Throughput 字段表示发送端当前的数据吞吐量；Delta_Throughput 字段表示发送端期望增加的数据吞吐量；Reverse_Feedback 字段表示节点路由器反馈给发送端的数据流量调整值。

### 2.2 XCP 工作过程

XCP 流量控制由数据发送主机、数据接收主机和数据传输通路中的各个节点路由器协同完成，其工作过程如图 3 所示。

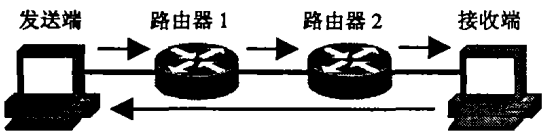



图 3 XCP 工作过程

发送端负责计算和维护数据流的双向传输时延、当前数据吞吐量、期望的数据吞吐量增加值等参数，并将它们分别填入 XCP 报文头的 RTT 、Throughput 和 Delta_Throughput 字段，通过发送的每个数据包将这些数据流状态信息携带给数据流所经过的各个节点路由器。

节点路由器则根据输入接口数据流量、输出接口队列长度、链路剩余带宽资源和每个数据包携带的数据流状态信息，在充分考虑优化链路带宽资源利用率和带宽资源分配公平性前提下，计算出每个数据流的数据流量调整值。如果路由器计算出的数据流量调整值小于 Delta_Throughput 字段中的当前数据流量调整值，路由器将更新该字段的内容，否则其内容将保持不变。因此，发送端数据吞吐量的增减最终由网络中的瓶颈节点路由器控制。数据流量调整值可以为正或负，正值表示发送端可以增加其数据吞吐量，负值则表示发送端需减少其数据吞吐量。

接收端接收到数据包后，将 Delta_Throughput 字段中的数据流量调整值拷贝到 Reverse_Feedback 字段中，并通过确认包将其反馈给发送端。发送端将根据每个确认包中的数据流量调整值增加或减少注入网络中的数据流量，从而完成网络对各个数据流的流量控制。

## 3 XCP 流量控制规则

各个数据流的数据流量调整值通过两个计算步骤获得。首先，节点路由器以带宽利用效率最大化为目标，计算出整个链路的数据流量调整值；在此基础上，路由器再将其公平地分摊给各个数据流。这两部份工作分别由节点路由器的效率控制模块和公平性控制模块完成。

### 3.1 效率控制

在带宽较高的情况下，为使各个数据流迅速利用链路的空闲带宽资源，提高链路的带宽资源利用率，XCP 采用了“积式增加积式减少”(MIMD)规则控制数据流的数据流量。在每个流量控制周期中，节点路由器的效率制模块对整个链路的会聚数据流量按比例 $\Phi$ 进行调整：

$$\Phi=\alpha\cdot d\cdot S-\beta\cdot Q \tag{1}$$

其中，$\alpha$、$\beta$ 为常数 0.4 和 0.226，$d$ 为各个数据流 $RTT$ 的平均值，$S$ 为空闲带宽资源，$Q$ 为瞬间队列长度。

由式(1)可知，会聚数据流量的调整比例与链路超载($S\geqslant 0$)或欠载($S\leqslant 0$)程度、节点路由器当前队列长度($Q$)相关，而与链路中数据流的数量无关。

### 3.2 公平性控制

XCP 采用 AIMD 规则和带宽混洗来调整各个数据流的带宽占用率，以逐渐逼近带宽资源分配的公平性要求。

在 AMID 规则下，当 $\Phi>0$ 时，无论单个数据流当前数据流量为多少，各个数据流的流量增加值相同；当 $\Phi<0$ 时，各个数据流量均成倍减少。因此，当 $\Phi\neq 0$ 时，AIMD 规则可以对各个数据流的流量进行连续的公平性调整。

当 $\Phi\approx 0$ 时，链路带宽利用率已接近优化水平，AIMD 规则将停止流量调整。因此，节点路由器还通过带宽混洗对各个数据流的数据流量进行微调：

$$h=\max(0,\gamma\cdot y-|\Phi|) \tag{2}$$

式(2)中，$y$ 为一个流量控制周期中输入节点路由器的数据流量；$\gamma$ 为常数，通常取 0.1，表示在每个流量控制周期中至少有 10%的数据流量根据 AMID 规则重新进行调整。

## 4 XCP 性能分析

XCP 具有较好的链路带宽利用率。如图 4 所示，随着网络的带宽和时延不断增加，TCP 数据流对链路带宽资源的利用率急剧下降，而 XCP 数据流则始终保持接近 100%的链路带宽资源利用率。因此，在 BDP 较大的网络环境中，XCP 比 TCP 具有更优的网络流量控制性能。

XCP 对网络带宽资源的分配具有较好公平性。如图 5a 所示，当各个数据流 RTT 不同时，TCP 数据流的数据吞吐量存在巨大差异，1 至 5 号 TCP 数据流的数据吞吐量维持 1 以上，26、27 号 TCP 数据流的数据吞吐量则几乎接近为 0，显然各个数据流的网络带宽资源分配极不公平；如图 5b 所示，当各个数据流 RTT 相同时，TCP 数据流的数据吞吐量依然存在较大差异，5、6、15、17 号 TCP 数据流保持较高数据吞吐

量，9、19 号 TCP 数据流的则维持较低数据吞吐量；然而，无论各个数据流的 RTT 是否相同，XCP 数据流均保持较高的数据吞吐量。因此，在 BDP 较大的网络环境中，XCP 数据流能够公平地获得网络带宽资源。

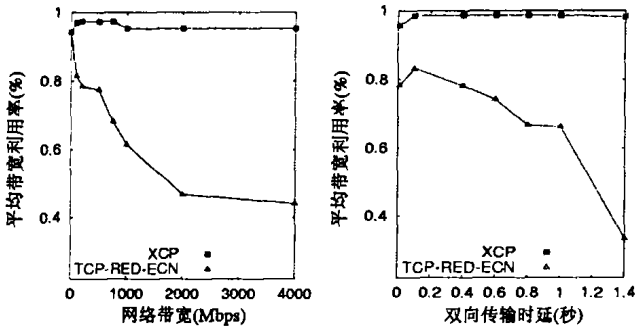
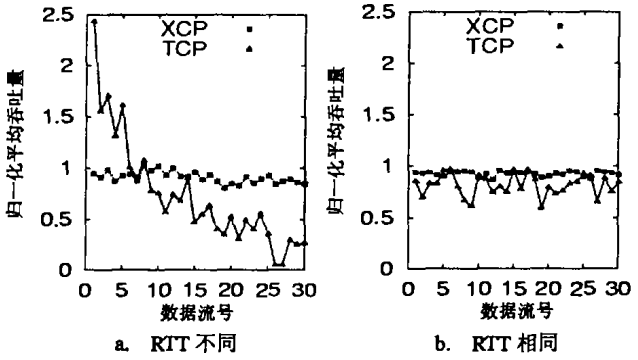



图 4　XCP 与 TCP 的带宽利用率比较




a. RIT 不同　　b. RIT 相同

图 5　XCP 与 TCP 数据吞吐量比较

**结束语**　XCP 通过数据包携带数据流状态信息和流量控制信息，提高了可扩展性；带宽利用率控制和带宽分配公平性控制的分别实施，保证了网络带宽资源的高效利用和公平分配，同时还增加了灵活性，可以很容易地支持区分服务，为网络运营商为不同业务提供区分服务而获取更多利润提供了技术手段；效率控制模块和公平控模块只需要少量代码即可实现，非常适合应用在核心路由器中。因此，XCP 极有希望成为下一代互联网的主流网络流量控制协议。

## 参 考 文 献


1　Jacobson V. Congestion Avoidance and Control. In: Proc. ACM SigComm'88, Aug. 988

2　Lakshman T V, et al. The Performance of TCP/IP for Networks with High Bandwidth-Delay Products and Random Loss. IEEE Trans. on Network, June 1997

3　Allman M, et al. TCP Performance over Satellite Links. In: Proc. 5th ICTS, Mar. 1997

4　Caceres R, et al. Improving the Performance of Reliable Transport Protocols in Mobile Computing Environments. IEEE Journal of Selected Areas in Communications, 1995, 13(5)

5　Low S H, et al. Dynamics of TCP/ARM and A Scalable Control. In: Proc. InfoCom'02, Jun. 2002

6　Katabi D, et al. Internet Congestion Control for Future High Bandwidth-Delay Product Environments. In: Proc. ACM SigComm'02, August 2002

7　Partridge C. Gigabit Networking, Reading, Massachusetts. Addison-Wesley, 1994

8　Low S H, et al. Internet Congestion Control: An Analytical Perspective. IEEE Control Systems Magazine, Feb. 2002

9　Henderson T, et al. On Improving the Fairness of TCP Congestion Avoidance. In: Proc. IEEE Globecom'98, Nov. 1998

10　Dawkins S, et al. End-to-end Performance Implications of Links with Errors. IETF RFC3155, Aug. 2001

11　Falk A, et al. Specification for the Explicit Control Protocol (XCP). Internet-draft, Oct. 2004

12　Katabi D, et al. Congestion Control for High Bandwidth-Delay Products Networks. In: Proc. ACM SigComm'02, Aug. 2002


(上接第 25 页)

布的原因。针对已有模型不能全面体现网络的特征属性，提出了基于组增长的小世界 Scale-free 模型。该模型既考虑到了网络节点具有很强的本地连接，又考虑到了网络成长的动态性和新增加连接边与节点度的关联性，把小世界网络和 Scale-free 网络联系了起来。使用该模型生成的网络图既具有小世界特征，又具有 Scale-free 特征，能较全面地反映实际网络的特征属性。该模型有助于建立更接近实际的网络拓扑结构和网络的动态成长过程，从而能更准确地分析、设计和评测与网络结构和行为相关的工作。下一步的工作是对 GGSS 模型结构特征和性质作进一步的分析，如容错、抗攻击能力等，并和其他网络模型的相应内容进行比较。

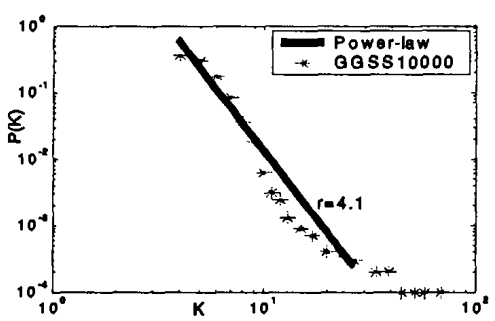



图 2　节点度分布 $P(k)$对数曲线

## 参 考 文 献


1　Adamic L. The Small-world web. In: Proc. Eur. Conf. on Digital Libraries(ECDL), Sept. 1999. 443～452

2　Albert R, Jeong H, Barabasi A. Diameter of the world wide web. Nature, 1999, 401: 130～131

3　Albert R, Barabasi A. Topology of evolving networks: Local events and universality. Physical Review Letters, 2000, 85: 5234～5237

4　Faloutsos M, Faloutsos P, Faloutsos C. On power-law relationships of the Internet topology. In: Proc. of ACM SIGCOMM, 1999

5　Erdos P, Renyi A. Publ. Math. Inst. Hung. Acad. 1960, Sci. Ser. A 5, 17

6　Watts D, Strogatz S. Collective dynamics of 'Small-world' networks. Nature, 1998, 393: 440～442

7　Newman M E J, Watts D J. Renormalization group analysis of the Small-world network model. Phys. Lett, 1999, 263: 341～346

8　Kleinberg J. The Small-world phenomenon: An algorithm perspective. In: ACM Symposium on Theory of Computing (STOC), 2000, 5: 163～170

9　Barab'asi A, Albert R. Emergence of scaling in random networks. Science, 1999, 286: 509～512

10　Aiello W, Chung F, Lu L. A random graph model for massive graphs. In: proc. of the 32rd Anmu/ ACM Synpo. sium on Theory of Computing. 2000

11　Jm C, Chen Q, Jamm S. Inet: Internet Topology Generator: [Technique Report CSE TR 433 00]. University of Michigan, EECS dept. 2000

12　Medina A, Matta I, Byers J. On the origin of power laws in Internet topologies. ACM SIGCOMM Computer Communication Review, April 2000

13　Bu T, Towsley D. On distinguishing between Internet power law topology generators. In: Proc. of IEEE INFOCOM, 2002

14　Krapivsky P L, Redner S. Theory Probab. and its Appl. Phys. Rev. E, 2000, arXiv: cond-mat/0011094