Internet拥塞控制研究*

任立勇** 卢显良

(电子科技大学计算机科学与工程学院 成都 610054)

【摘要】随着互联网业务量的剧增和新业务的层出不穷，单一的TCP协议已经不能胜任所有的拥塞控制任务。因此，论述了Internet拥塞控制研究方面最新的研究进展，分析了IP网拥塞的原因，讨论了网络拥塞控制的方法，包括路由器中拥塞控制策略和对多媒体实时流与组播流的拥塞控制方法。得出了只有采用多种策略，并从多个角度实施拥塞控制，才能更好地保证互联网的正常运行。

关 键 词 互联网; TCP友好; 组播; 公平队列

中图分类号 TP302

近年来，随着网络用户的急剧增加，互联网变得日益繁忙。通信业务的迅猛增长使得主干网越来越拥塞，与此同时，各种各样的新型业务对网络服务质量提出了更高的要求。TCP拥塞控制在保证互联网的鲁棒性和稳定性方面起着至关重要的作用，但随着互联网技术的发展和网络情况的变化，仅靠TCP一种拥塞控制来管理网络业务已经显得越来越力不从心了。

网络的拥塞控制一般有开环和闭环控制两种，对互联网网络业务不断变化的复杂系统，一般采用闭环控制。依据控制论的观点，发送方接收反馈信息，并从反馈信息中推断网络状况，确定控制参数，并依据一定的控制算法调整发送速率，完成对网络拥塞的响应。因此，正确的控制算法和准确及时的控制参数是网络拥塞控制的关键。为此，网络拥塞控制的研究重点就变成研究控制算法和控制参数。

由于许多新的网络应用(如多媒体实时传输、组播应用等)在传输层采用UDP协议及与TCP不同的拥塞控制策略，甚至没有任何拥塞控制，为此，需要研究新的拥塞控制策略，主要从以下两方面入手：1) 在网络设备(主要是路由器)中采取一定的策略来避免和控制网络拥塞；2) 研究新的端到端的拥塞控制算法。

本文介绍了互联网拥塞的原因，分析了造成网络越来越拥塞主要原因是由于许多新的网络应用采用了与TCP不同的拥塞控制策略，并且不采用任何拥塞控制。论述了在路由器中采取拥塞控制策略和对多媒体实时流与组播流的拥塞控制方法。

1 互联网拥塞的原因及TCP拥塞控制

网络产生拥塞的根本原因在于用户提供给网络的负载大于网络资源容量和处理能力，具体表现为报文时延增加、丢失、服务质量降低等。拥塞一旦发生，如果不采取正确控制，拥塞会继续加重最终导致网络崩溃。文献[1]分析了拥塞产生的直接原因，主要表现在三点：1) 路由器存储空间不足；2) 带宽容量相对不足；3) 处理器处理能力弱、速度慢。

TCP拥塞控制是一种自适应、分布式的控制系统[2]，能对网络的拥塞作出及时、准确处理。如果所有的网络流均遵守或兼容TCP控制，网络的拥塞则会得到很好的控制[3]。但许多新的网络应用却采用了与TCP不同的拥塞控制策略，甚至不采用任何拥塞控制，给网络的稳定运行带来了巨大的

2001年10月8日收稿
* 信息产业部生产发展基金资助项目
** 男 30岁 博士生

威胁。新的网络应用主要有以下两类：

1) 多媒体实时应用和部分用户自己编写的网络程序。这类应用为追求高的服务质量，一般采用恒定的发送速率，称为非响应流或恶意流。当拥塞发生时，有拥塞控制反应机制的TCP数据流会按拥塞控制步骤进入拥塞避免阶段，从而主动减小发送入网络的数据量。由于没有端到端的拥塞控制机制，即使网络发出了拥塞指示(如报文丢失，往返时延增加等)，也并不降低发送速率，结果导致遵守拥塞控制的TCP流得到的网络资源越来越少，而没有拥塞控制的数据流会得到越来越多的网络资源，导致了网络资源分配的严重不公平，最终TCP流“饿死”，直至网络崩溃。

2) 组播应用的推广一方面节省了网络资源，降低了网络管理费用，但与此同时，如果组播应用不能对网络拥塞作出正确响应，会给Internet带来比单点投递应用产生的拥塞更为严重的影响。这主要因为：(1) 组播投递流可能沿着其多点投递树广泛分布于整个Internet；(2) 组播流的接收者具有异构性，每个接受者的处理能力不同，其报文投递路径也可能有不同的带宽和差错特性；(3) 组播发送方需处理比单点投递多得多的反馈报文，对报文如果不加以处理，不仅可能淹没发送方，而且其本身也是网络资源的一种极大的开销。

2 路由器的拥塞控制策略

随着网络规模越来越大，结构日趋复杂，为了更有效地管理网络拥塞，部分中间网络设备，如路由器必须参与到拥塞控制的工作中来。路由器的拥塞控制策略主要可分为以下几种。

2.1 缓冲管理策略

路由器中传统的缓冲管理采用了“先来先服务”(FCFS)策略，报文到达时如果缓冲已满，路由器则丢弃该报文，因此FCFS又称为“丢尾”。由于FCFS实现简单，故仍被大多数路由器所采用。但FCFS将拥塞控制的所有责任都推给端用户，既不考虑丢包优先级，也不能处罚恶意流。

为了改进路由器的缓冲管理，研究者提出了各种缓冲管理策略，其中最为典型的是随机早期检测算法RED[4]。RED算法在路由器中检测队列长度，当平均队列长度Q_{avg}小于预先设定的阈值$\min_{th}$时，转发所有报文；当Q_{avg}大于$\min_{th}$并小于$\max_{th}$时，按照一定的概率P_{drop}丢弃新到数据报；当Q_{avg}大于$\max_{th}$时，RED退化为丢尾算法。算法中

$$Q_{avg} = (1-w)Q_{avg} + wQ_{sample} \tag{1}$$

$$P_{drop} = P_{\max}\frac{Q_{avg} - \min_{th}}{\max_{th} - \min_{th}}\frac{PacketSize}{PacketSize_{\max}} \tag{2}$$

RED算法优点在于路由器在缓冲溢出前就按一定的概率丢弃进入路由器的数据报，可以提前通知源端减小拥塞窗口，避免网络拥塞。另外，RED按照连接占用的网络带宽成比例丢弃该连接的报文，在一定程度上惩罚了恶意流和高带宽流，保证了一定的公平性。但缺点是对于恶意攻击的用户或对分组丢失不敏感的传输层协议，该方法并不能独立地做到避免拥塞的发生，当拥塞发生时也无法保证对各个流的公平性，它要依赖用户终端协作与配合才能真正发挥作用。另外，选择合适的配置参数也不是件容易的事。近年来，研究者提出了许多RED的改进算法[5, 6]，在一定程度上改善了RED的性能。

2.2 公平排队算法FQ

RED的缺陷在于路由器只维护一个队列，无法区分不同的数据流，当出现拥塞时，不能隔离不良行为流，使公平性得不到保证。公平排队算法为每个连接建立一个输出队列[7]，路由器按“轮询”的方式处理每个队列，依次将每个队列的第一个报文发送出去。由于不同连接具有不同长度的报文，因此为保证所有连接公平分配带宽，路由器为每个队列维护一个时间戳，每转发一个报文，则给该队列的时间戳加上发送该报文用去的时间，可以保证每个连接完全公平占用网络带宽。

加权公平排队算法(WFQ)及其改进算法是FQ的改进[8]，给每个连接分配一个权值，该权值决定了路由器每次发往该队列的比特数量，从而控制数据流得到的带宽。

2.3 显示拥塞指示(ECN)[9]

大多数的拥塞控制算法都是用包丢失作为告诉端系统网络发生拥塞的指示，这种方式对一次性大批量、时延要求不高的数据传输效果较好，但需重传报文，从而造成网络资源浪费，报文时延增加。在ECN算法中，路由器采用RED算法管理缓冲区，当平均队列长度处于两个阈值之间时，路由器按照一定概率给报文设置CE使能位，而不是简单地丢弃该报文。当下端路由器发生拥塞时，首先选择有CE使能位的报文丢弃。ECN的优势在于不需要重传超时，有效地提高网络带宽的使用效率。

3 多媒体实时流拥塞控制

为保证用户要求的服务质量，大多数多媒体实时应用采用UDP作为其传输层协议。多媒体实时流一般对时延敏感，但不要求可靠传输，在连接期间要求有稳定(平滑)的接收速率和一定的带宽下界，超时和失序的分组并不需要重传。这些特殊的要求使TCP的基于窗口的和式增加积式减少的拥塞控制已经不能满足要求。为此，提出了方程的拥塞控制算法EBCC[10]。方程的拥塞控制的原理是通过一个以丢失事件率等为参数的方程来计算发送方的发送速率上限，发送方以此计算结果为依据来对自身的发送速率进行调整，并保证发送速率不会高于这个值

$$T \leqslant \frac{1.5\sqrt{2/3}M}{RTT\sqrt{p}} \approx \frac{1.22M}{RTT\sqrt{p}} \tag{3}$$

$$T = \frac{M}{RTT\sqrt{2p/3} + T_{RTO}(3\sqrt{3p/8})\,p(1+32p^2)} \tag{4}$$

实现方程的拥塞控制关键是选择一个合适的控制方程。由于目前Internet中95%流都是TCP，因此新的拥塞控制算法必须是TCP友好的。为此，一般选择稳定状态下的TCP流量模型作为控制方程。式(3)、(4)分别为两种情况下的TCP流量模型，其中式(3)没有考虑超时重传，当丢失率大于16%时，该公式则过高地估计连接占用的带宽，式(4)在推导过程中不仅考虑了快速重传对吞吐量的影响，同时还考虑了由于超时导致的吞吐量下降问题，因此在刻画TCP连接吞吐量时较为准确。

EBCC在计算丢失率时作了特殊的平滑处理。假设过去n个丢失间隔分别为s_1，s_2,…,s_n，其中s_1为当前丢失间隔，s_n为离当前第n个丢失间隔。平均丢失间隔定义为

$$\tilde{s} = \sum_{i=1}^{n} w_i s_i \Big/ \sum_{i=1}^{n} w_i \tag{5}$$

则平均丢失率$p=1/\tilde{s}$。经过平滑处理后的丢失率既能及时反应最近丢失状况，又能一定程度上反应历史丢失状况。另外，EBCC采用指数加权移动平均(EWMA)估算往返时延RTT和超时值T_{RTO}。

分析和试验证明，EBCC流不仅能与TCP流公平共享链路带宽，同时能平滑调整源端发送速率，较好地满足了多媒体实时流的要求。

4 组播拥塞控制

组播是一种为优化使用网络资源而产生的技术，它通常使用在点对多工作方式下的应用中。组播技术虽然节省了网络资源，但也给网络带来了许多潜在的问题，其中组播拥塞控制是其中最重要，也是最难解决的问题之一。

近年来，组播拥塞控制已成为一个研究热点[11]。设计组播拥塞控制最大的挑战主要有两个：1) 可扩展性，组播用户往往数量巨大，而且接收链路和处理能力也千差万别，这些都是影响组播扩展性关键因素；2) 公平性，组播拥塞控制的公平性表现在两个方面：一方面是组播流与TCP流公平共

享瓶颈链路带宽，即TCP友好；另一方面是指组播用户间的公平性。目前处理组播拥塞控制的方法主要有以下4种：

(1) 组播拥塞控制模型。组播拥塞控制模型主要分为发送方与接受方。在发送方模型中，所有组播用户向发送方发送反馈报文，发送方能及时了解网络状况并做出相应处理，但随着组播用户的增加，发送方的处理开销将急剧增加，最终崩溃，解决的方法有反馈聚集和反馈抑制。接收方模型中，每个组播用户计算自身链路状况和接收能力并向发送方反馈信息。现有的接收方的组播拥塞控制一般应用在连续媒体流和大批量数据分发的层次传输中。

(2) 组播用户结构关系。组播用户的逻辑关系主要有平面型和结构型。平面型的组播用户将反馈信息以单播或组播的方式发送至发送方和其他组播用户。这种类型的好处有协议简单、不需中间设备支持等好处。但其致命缺陷是容易导致反馈风暴，可伸缩性较差。解决的方法是反馈抑制技术，如定时器方案、轮询方案、概率方案和基于代表的方案等。结构型将组播用户构建成一定的逻辑结构，利用中间设备(路由器或部分组播用户)作为中间节点，对反馈报文进行聚集和完成本地丢失报文的恢复任务等，主要包括树的组播拥塞控制和环的组播拥塞控制方案。结构的组播拥塞控制虽然在一定程度上减轻了反馈爆炸问题，但同时也带来了其他问题，如反馈延迟增加、自适应能力弱等。

(3) 窗口和速率调节。组播拥塞控制可以分为窗口的类TCP和速率的TCP友好。其中窗口的组播拥塞控制一般遵守

$$w_j=\begin{cases} w_j+A_j(w_j,t_j) & sucessful\quad packet\quad delivery\quad to\quad j \\ w_j-B_j(w_j,t_j) & packet\quad loss\quad ovserved\quad by\quad j \end{cases} \tag{6}$$

窗口维护的任务应分布在各个组播用户方，这一方面减轻了发送方的负担，另一方面也避免发送方利用最大往返时延和最大丢失率计算出窗口超低的情况[12]。

速率的组播拥塞控制一般采用稳定状态下的TCP响应函数作为发送方的速率调节函数，如式(3)、(4)分别为两种TCP响应函数，实现这种类型的拥塞控制关键是精确计算所有组播用户的RTT和丢失率。一种通用的速率调节公式为

$$r_j=\begin{cases} r_j+a_j(r_j,t_j) & sucessful\quad packet\quad delivery\quad to\ j \\ r_j-b_j(r_j,t_j) & packet\quad loss\quad ovserved\quad by\ j \end{cases} \tag{7}$$

除上述方法外，还有一种测量瓶颈链路带宽的方法，如

$$R=\frac{probe\,packet\quad size}{gap\,between\quad 2\quad probe\quad packets} \tag{8}$$

发送方定时向组播用户发送两个连续的探测报文，接收者根据两个探测报文到达的时间差计算瓶颈链路的带宽，发送方根据“每个”接收者反馈报文中的瓶颈带宽确定下一步的发送速率。

(4) 组播拥塞控制的公平性。公平性是组播拥塞控制机制能否得到广泛应用的基础，也是每个可靠组播拥塞控制机制所追求的设计目标之一。一般而言，组播拥塞控制采用以下两种方法达到协议间公平即TCP友好：(1) 采用窗口的类TCP来模拟TCP对拥塞的响应；(2) 采用稳定状态下的TCP的响应函数作为速率调节公式。由于大多数组播协议以最慢速接收者速率作为发送速率，虽然在不同数据流之间有一定的公平性，但会在同一个组播组中不同的接受者之间导致不公平，特别是对并没有拥塞或拥塞程度较低的接收者来说不公平，即所谓的协议内公平较差。解决这个问题的思路主要在于对源数据采取合理的分层，每层分别对应于不同的组播组，有不同服务质量要求和不同链路状况的用户可加入不同的组播组，以此实现协议内相对公平。

5 结 束 语

随着互联网的飞速发展，其鲁棒性也越来越依赖于网络的拥塞控制。单一的TCP拥塞控制机制

尽管在网络的正常运行中发挥重要的作用，但随着业务的膨胀和新应用的增加，它已经不能胜任所有的拥塞控制任务，必须采用多种策略，从网络的各个部位、多角度全方位对拥塞加以控制，才能保证网络正常、稳定地运行。

参 考 文 献

1 罗万明，林 闯，阎保平. TCP/IP拥塞控制研究. 计算机学报，2001，24(1): 1-18

2 Allman M, Paxson V, Stevens W. TCP Congestion Control. Request for Comments: 2581, April 1999

3 王 晟，李乐民. 一种线性闭环拥塞控制方案的渐近性能分析. 电子科技大学学报，2000, 29(4): 450-456

4 Floyd S, Jacobson V. Random early detection gateways for congestion avoidance. IEEE/ACM Trans. On Network, 1993, 1(4): 397-413

5 Athuraliya S, Low S, Lapsley D, Random early marking. in Proceedings of the First International Workshop on Quality of future Internet Services (QofIS'2000), Berlin, Germany, September 2000

6 Lin D, Morris R. Dynamics of random early detection. Proc. of ACM SIGCOMM '97, 1997: 127-137

7 涂晓东，李乐民. 分组公平排队算法的仿真研究. 电子科技大学学报, 2000, 29(4): 440-444

8 任立勇，卢显良. EWFQ: 一种新的高速网络分组调度算法. 计算机科学, 2001, 28(11)

9 Ramakrishna K, Floyd S. A Proposal to add Explicit Congestion Notification (ECN) to IP. Request for Comments:2481, Jan. 1999

10 任立勇，卢显良. 一种基于方程的多媒体实时流拥塞控制机制. 计算机科学, 2001, 28(8): 60-63

11 任立勇，卢显良. 可靠组播拥塞控制最新研究进展. 计算机应用, 2001, 21(10)

12 Golestani S J. Fundamental Observations on Multicast Congestion Control in the Internet. In Proc. IEEE INFOCOM, March,1999: 990-1 000

A Survey of Congestion Control in the Internet

Ren Liyong Lu Xianliang

(College of Computer Science and Engineering, UEST of China Chengdu 610054)

Abstract With increasing drastically of Internet traffics and new application, only TCP protocol cannot be competent for all tasks of congestion control. This is review paper on recent word about congestion control of Internet. The reason of congestion of Internet is analyzed, then several methods of network congestion control are presented, they include congestion control strategies for routers and congestion control methods for multimedia real-time streams and multicast streams. Particularly, a reasonable category about multicast congestion control mechanism is presented, with analyzing of their advantage and shortcoming. At last we make conclusion that multiform strategies must be used for congestion control of network.

Key words internet; TCP-friendly; multicast; fair queuing