Internet 端到端拥塞控制研究综述*)

刘拥民[1,2] 蒋新华[1] 年晓红[1] 鲁五一[1]

(中南大学信息科学与工程学院 长沙 410075)[1] (中南林业科技大学职业技术学院 长沙 410004)

摘 要 本文描述了 Internet 中端对端拥塞控制技术的内容和特点,分析了标准化拥塞控制算法所涉及的技术难点及不足,概括了拥塞控制算法的发展历程,并根据目前最新的研究情况,提出把控制与优化理论引入网络拥塞控制的思路,最后总结出进一步的研究趋势和方向。

关键词 Internet,拥塞控制,TCP 协议,控制理论,端对端

Review on the End-to-End Congestion Control Research in the Internet

LIU Yong-Min[1,2] JIANG Xin-Hua[1] NIAN Xiao-Hong[1] LU Wu-Yi[1]

(College of Information Science and Engineering, Central South University, Changsha 410075)[1]

(College of Vocational Technology, Central South University of Forestry and Technology, Changsha 410004)[2]

Abstract This paper describes the technology and features of congestion control in the Internet. A careful analysis of the implementation of the congestion control algorithm involves technical difficulties and deficiencies were made, and the process of congestion control algorithm development was present. Optimal control theory was used into the network congestion control according to the latest research results. Finally summarize trends and directions for further research.

Keywords Internet, Congestion control, Network protocol, Control theory, End-to-end

随着 Internet 在全球范围内的飞速扩展,端到端网络拥塞控制正成为越来越重要和亟待解决的问题。早期的 TCP 协议是不考虑拥塞控制的。直到 1986 年 Von Jacobson 发现从 LBL 到 UC Berkeley 的数据吞吐量从 32kbps 跌落到 40bps[1]时,网络拥塞问题才正式成为研究课题。

先前认为低速链路、较慢的处理器和较小的缓存是造成网络拥塞的原因,后来发现在高速链路、高性能处理器和大容量缓存相当普遍的情况下,网络拥塞现象不但没有得到消除或缓解,反而进一步恶化[2],甚至导致网络崩溃(congestion collapse)的发生。

此外,随着组播和无线网络的出现,实时多媒体数据传输 QoS 的应用需求,又给传统的 TCP 拥塞控制提出新的要求。为了保证网络畅通并提供一定的服务质量保证,在今天网络拓扑结构、传输链路和应用需求异构化的场合下,必须采取一定的策略来避免和控制网络拥塞。

本文内容组织如下:首先介绍了拥塞控制的基本概念,分析了 Internet 中产生拥塞的原因;描述了拥塞控制算法的分类方法;对现有的拥塞控制算法在实现上的缺陷进行分析,对研究性的拥塞控制算法设计上的困难进行探讨。论述过程当中,分析了组播拥塞控制问题。最后对全文进行总结,提出把控制与优化理论引入网络拥塞控制的思路,给出未来的研究方向。

1 基本概念

1.1 拥塞和拥塞控制

因为存在许多不同的度量来描述拥塞现象,如传输延时、数据吞吐量、队列长度和网络效率等,但是没有哪个度量能在局部和全局意义上完全满足拥塞评判要求,因此人们对拥塞控制并无严格定义,甚至对拥塞的定义都无法完全统一。这里给出一个相对被普遍认可的拥塞的定义。

定义 如果因为网络负载增加而导致用户 I 的满意度降低,用户 I 则认为网络发生拥塞[2](如图 1)。

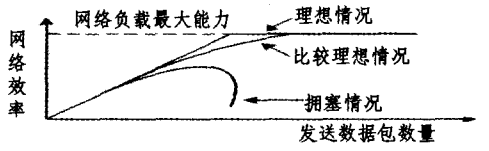

图 1 网络拥塞情况

形象地说,当网络中存在过多的报文时,网络的性能会下降,这种现象称为拥塞[3,4]。拥塞导致的直接结果是分组丢失率提高,端到端时延加大,甚至是整个系统发生崩溃。当网络发生拥塞崩溃时,微小的负载增量都将使网络的有效吞吐量(goodput)急剧下降。

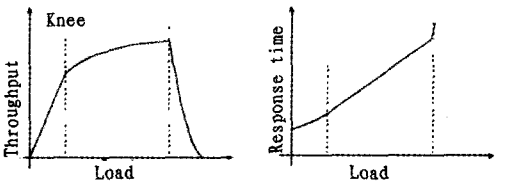

图 2 网络负载情况分析

*)国家自然科学基金资助项目(60474029);中南林业科技大学青年科学研究基金资助项目(101-0582);中南林业科技大学教学研究基金资助项目(中南林学院院改[2004]28-67 号)。**刘拥民** 博士生,讲师,主要研究领域为计算机网络及协议、多媒体通信、宽带 IP 网络体系结构等;**蒋新华** 博士导师,教授,主要研究领域为计算机技术及应用、自动控制、控制理论与工程等;**年晓红** 博士,教授,主要研究领域为控制理论与应用、智能机器人控制;**鲁五一** 硕士导师,教授,主要研究领域为计算机网络及应用、自动控制等。

文[5]使用图2描述拥塞：网络负载较小时，吞吐量与网络负载之间呈线性关系，网络延迟缓慢增加；网络负载超过(Knee)后，网络吞吐量增长缓慢，网络延迟增长变快；网络负载到达(Cliff)后，网络吞吐量急剧下降，网络延迟急剧上升。

网络拥塞控制是确保网络鲁棒性(robustness)的关键因素。拥塞控制算法最基本和最重要的要求就是防止网络出现拥塞崩溃，使网络运行在轻度拥塞的最佳状态。这样，拥塞控制模型或算法既能保证网络效率，又不会出现网络欠载或过载，同时又能保证流量间的公平性。事实上，控制算法包括拥塞避免(congestion avoidance)和拥塞控制(congestion control)。前者的目标是使网络尽可能不发生拥塞，保持网络高吞吐量、低延迟的运行状态；后者的目标是把网络从拥塞中解除出来。

拥塞控制与Best-Effort服务模型有紧密的联系。流控制主要考虑接收端，目的是使发送端的发送速率不超过接收端的接收能力；而拥塞控制主要考虑端节点之间的网络环境，目的是使负载不超过网络的传送能力。传统的网络使用Best-Effort服务模型，只有流控制(flow control)而没有拥塞控制，接收端利用TCP报头将接收能力通知发送端。这个服务模型只考虑了接收端的接收能力，而没有考虑网络的传输能力。

1.2 拥塞的原因

网络拥塞是Best-Effort服务模型的一个固有属性。用户间无法相互协作共享资源，多个用户对同一网络资源提出请求时，就可能发生拥塞。产生网络拥塞的原因有很多，直接原因主要有三个方面：

(1)存储空间不足。多个分组同时到达路由器，并期望经同一个输出端口转发时，等待处理服务的分组序列自动进入中间节点上的缓存等待接受处理。如果这种情况持续发生，当缓存空间被耗尽时，路由器只有丢弃分组。从表面上看，增大缓存可以防止由拥塞引起的分组丢弃，但如果缓存太大，端到端的时延也会增大。因为分组的持续时间(lifetime)是有限的，超时的分组同样需要重传，反而会加剧网络拥塞。

(2)带宽容量不足。任何信道带宽最大值(即信道容量)为 $C = B\log2(1 + S/N)$(N为信道白噪声的平均功率，S为信源的平均功率，B为信道带宽)[6]。要求所有信源发送的速率R必须小于或等于信道容量C。如果R大于C，则在网络低速链路处就会形成带宽瓶颈，一旦当其满足不了所有通过它的源端带宽的需求，网络就会发生拥塞。

(3)处理器间处理能力和速度不一致也可能造成拥塞[7]。如果路由器的CPU在执行排队缓存，更新路由表等功能时，处理速度跟不上高速链路，也会产生拥塞。同样，低速链路对高速CPU也会产生拥塞。比如：S端以1Mb/s的速率向D端发送数据，在数据经过网关R时，就会发生拥塞(如图3)。

图3 网络拥塞示意图

网络产生拥塞的根本原因在于用户(即端系统)提供给网络的负载(Load)大于网络资源容量和处理能力(Overload)[8]。拥塞是一种持续的网络超负荷状态。其典型表现就是数据包时延增加、丢弃概率增大、上层应用系统性能显著下降等。

2 拥塞控制算法的分类

拥塞控制算法存在多种分类方法[9]。本节将讨论几种常用的分类方法。

从推断网络状态的反馈信息的类型上，可以分为显式拥塞控制(explicit)和隐式拥塞控制(implicit)[10]。前者网络具有独立的拥塞控制过程，系统使用显式信号向执行流量控制的端点通告其状态(有效带宽、缓存容量等)，可再细分为定性和定量控制两种；后者是通过对数据传输的观察以获取当前网络状态，如控制端使用流量测量或者通过诸如超时、重复ACK等隐含信号来推断网络状态。

从控制论角度出发，按作用方式分为：开环控制方式(或前动式/预防)和闭环反馈控制方式(或被动式)两类。当流量特征可以准确规定、性能要求可以事先获得时，适合使用开环控制。这是一种预防式的控制方法，其主要缺点是需要一个精确模型且网络资源利用率低。显然，对网络这样不断变化的复杂系统，开环控制并不是理想的选择。当流量特征不能准确描述或者当系统不提供资源预留时，适用闭环控制。闭环控制是根据各网络连接的发送端根据网络内部反馈信息调节信源发送速率的，不管模型是否精确，都能有效地补偿干扰[11,12]，Internet中主要采用闭环控制方式。

从实施控制的位置，可以分为在端系统上使用的源算法(Source algorithm)和在网络设备上使用的链路算法(Link algorithm)[13]。源算法是指端对端的拥塞控制(end-to-end)，它仅在网络边缘(即在TCP层)端系统或者主机处执行，此时中间节点仅负责向端系统产生和转发必要的反馈信息，源算法中使用最广泛的是TCP协议中的拥塞控制算法。而链路算法指路由器拥塞控制(router-based)，是在网络设备(路由器或交换机)等中间节点处(即在IP层)执行的，多采用的是“弃尾”算法。

TCP是目前Internet上使用最广泛的一种传输协议。今天Internet的迅猛发展，证明TCP协议中的源算法的引入是非常成功的。据MCI的统计：在Internet上总字节数的95%以及总数据包数的90%使用TCP协议传输[25]。源算法已经成为保证Internet稳定性、继续健康发展的重要因素[14]。

主动队列管理AQM(active queue management)是在网络设备的缓冲溢出之前就丢弃或标记报文，所以平均队列较短，具有以下优点：增强网关容纳突发流量的能力，从而减少网关的报文丢失；减小报文在网络设备中的排队延迟；避免发生lock-out行为[15]。

链路算法的研究目前集中在AQM，AQM的典型代表就是RED(random early detection)[16]。研究表明，RED比Drop Tail具有更好的性能。但是由于RED的性能对算法的参数设置十分敏感，至今没有在Internet中得到广泛的使用。

源算法的作用是在根据反馈信息调整发送速率；链路算法的作用主要是检测网络拥塞的发生，产生拥塞反馈信息。由于实际中网络内部的路由器和网络边缘的端系统都参与了拥塞控制，因此拥塞控制算法设计的关键问题是如何生成反馈信息和如何对反馈信息进行响应。

3 源拥塞控制算法AIMD研究现状

对于任何一个实际运用的算法，都是标准控制算法组成部分的组合、变形或改进而来的，或者说TCP的4个状态是相辅相成、相互转换的。TCP协议在Internet体系结构中的

位置相当于OSI(open systems interconnection)模型中的第4层[17],在不可靠的IP层基础上为应用层提供一个可靠的、按序传输、端到端的数据包传输服务。目前主要的网络拥塞控制算法主要有:TCP Tahoe、TCP Reno、New-Reno TCP、TCP SACK、TCP Vegas。

3.1 TCP Tahoe

Jacobson在1988年改进了原来的TCP,提出了Tahoe。在早期实现的基础上,Tahoe加入了慢启动、拥塞避免(加性增加部分)、快速重传机制三个部分。重传计时器的超时值得到了改进。Tahoe的基本思想是探测网络的可用带宽,在拥塞时急剧地降低数据发送速率[5,6]。Tahoe具备RTT估计和差错恢复的功能。

3.2 TCP Reno

1990年,Jacobson提出了Reno。Reno的快速重传中包括快速恢复,在三次重复的ACK后不进行慢启动,这使得在仅仅一个分组丢失时的TCP性能得到了提高[7],但有多个分组从同一数据窗口丢失时,依旧存在性能问题。Reno的基本思想是用快速恢复取代了慢启动,在收到三个和以上相同的ACK时进入快速重传和快速恢复,在超时的时候进入慢启动。

3.3 New Reno

1996年,Fall和Floy在Reno的基础上提出了New Reno。在TCP Reno的基础上进行了小小的改变,使得在一个窗口中有多个分组丢失时的TCP性能有所提高。在Reno中,处于快速恢复状态的发送方在收到第一个正常的确认(Partial ACK)后,就会立即离开快速重传/快速恢复。当存在多个分组丢失时,这一算法很难将它们全部恢复。New Reno中的发送方只有在收到所有丢失分组的确认后,才会离开快速恢复状态,否则就继续进行快速重传/快速恢复。每一个RTT发送一个丢失的分组,直到原来窗口中丢失的分组已经全部被重传过。

3.4 TCP SACK

1996年,Mathis、Mahdavi、Floy和Romanow还提出了Reno的另一变形:SACK,采用选择性确认,而不是GO BACK N机制,进一步提高TCP在拥塞较严重且一个窗口中有多个分组丢失时的性能[9]。Reno和New Reno在一个RTT内最多只能重传一个丢失的分组。如果一个窗口中有多个分组丢失的话,这就有可能导致网络中没有分组。SACK的基本思想是接受方TCP发送SACK分组来通知发送方接受数据的情况,这样发送方只重传丢失的分组。SACK在进入快速重传状态时,如果网络中的所有分组已经得到确认,那就会退出快速重传状态。

3.5 TCP Vegas

1994年,L. S. Brakmo等[8]提出了一种新的拥塞控制策略:TCP Vegas。由于RTT值与网络运行情况有密切关系,因此TCP Vegas通过观察TCP连接中RTT值改变来感知网络是否发生拥塞,从而控制拥塞窗口大小。Vegas对Reno进行了三项重要的技术改进:(1)采用了新的重传触发机制,即用一个重复ACK(而非Reno中的3个)来启动超时判定规程,这样可以更及时地检测到拥塞的发生;(2)在慢启动阶段采用了更加谨慎的方式来增加窗口大小,减少了不必要的分组丢失;(3)改进"拥塞避免"阶段的窗口调整算法。

从上面的分析可以看出目前TCP协议主要版本的一些特性:TCP Tahoe包括了3个最基本的拥塞控制阶段:"慢启动"、"拥塞避免"和"快速重传"[11,12];TCP Reno在TCP Tahoe基础上增加了"快速恢复"方式;TCP New Reno对TCP Reno中的"快速恢复"方式进行了修正,它考虑了一个发送窗口内多个数据包丢失的情况。在Reno版中,发送端收到一个新的ACK后,就退出"快速恢复"阶段;而在New Reno版中,只有当所有的数据包都被确认后才退出"快速恢复"阶段[14]。TCP Vegas在Internet上是不大可能被广泛使用的,因为这里有一个致命的问题没有解决:在竞争带宽时,使用TCP Vegas的数据流在带宽竞争能力方面极差,远不及未使用TCP Vegas的数据流,因此会导致网络资源分配的严重不公平[16]。

当然,类似于标准AIMD(Additive Increase Multiplicative Decrease)参数控制方案,使用增加/减小常数来控制发送窗口或者发送速率的大小,除了上面论述到的各种典型的TCP变形之外,还有Rejaie等人提出的基于速率的RAP(Rate Adaptation Protocol)[17]等。由于RAP不具备自同步机制,因此表现出与标准TCP拥塞控制完全不同的瞬态响应状态。

4 源拥塞控制算法的最新研究进展

扩大"慢启动"的初始阈值:为了更好地传输短数据流(比如:HTTP流),文[18]直接将初始拥塞窗口的值设置为4MSS(maximum segment size),而不是标准值1MSS。逐步减小窗口增长的速度,能够实现从"慢启动"到"拥塞避免"的平稳过渡[19]。相关的研究性算法有SPAND[20]和"TCP Fast Start"[21],它们是根据网络当前的拥塞状况来确定"慢启动"的初始阈值,也不是标准值1MSS。

TCP拥塞控制算法使Internet在全球的迅速蔓延成为可能,但网络拓扑、协议、应用和实现的异构化,原有的算法变得不再合适。为了适应不断出现的新业务应用需求(比如多媒体数据传输),研究者提出基于TCP-Friendly的拥塞控制算法,这种算法是建立在TCP吞吐量模型[19]上的基于速率的控制策略。

典型算法有Floyd等人提出的TFRC[22]。TFRC的工作原理是将速率的控制设为分组丢失率和RTT的函数,它只是响应固定间隔时间上测得的分组丢失率,而不同于标准拥塞控制方案中的对每一个分组丢失事件都产生响应。

相关的研究还有TCP仿真机制(TCP Emulation at Receiver,TEAR)。它是在响应分组丢失时,不像TCP拥塞控制那样,将拥塞窗口(或传输速率)减半,而是采用较为缓慢的速率调节算法。这不同于标准的AIMD参数控制算法[23]。TEAR并没有改变TCP的拥塞窗口计算方法,只是在接收端对拥塞窗口进行了必要的修正,使其同时具备TCP-Friendly和慢响应特性。

4.1 组播拥塞控制

随着新型分布多媒体和分布式处理应用的发展,单点到多点有效的数据传输方式成为非常急迫的通信需求。但是现在组播仍然没有得到ISP(Internet service provider)的广泛应用。网络中传统的TCP拥塞控制算法受本身协议机制约束无法支持多点投递,组播(Multicast)拥塞控制面临的最大挑战是无法同时满足扩展性(scalability)和TCP友好性(TCP-friendly)。

4.2 单速率组播拥塞控制

TRAM(tree-based reliable multicast protocol)是基于单

速率组播拥塞控制协议的典型代表。为了保证协议的 TCP-friendly，TRAM 通过最小、最大速率和 ACK 窗口等参数的设定来实现；为了保证发送速率有一个平滑的变化范围，协议预设两个变量：最小速率和最大速率。TRAM 使用修复树(repair tree)收集接收端到发送端的反馈来限制数据重传范围。修复树避免了反馈爆炸问题，但仍然存在 LPM(loss path multiplicity)问题。另外，TRAM 并不能确保协议的友好性，因为最小、最大速率和 ACK 窗口等参数预设值可能是不正确的。

TFMCC(TCP-friendly multicast congestion control)[24]是 TFRC 协议[25]的一种改进算法。同 TFRC 一样，TFMCC 是基于 TCP 流量模型[41]来调整速率的，接收端使用平均丢失间隔算法(average loss interval method)计算分组丢失率。为了防止反馈爆炸，协议使用指数加权随机时钟来抑制接收端的反馈数量。为了保证协议的 TCP-friendly，TFMCC 发送端在数据分组中携带时间标记(timestamp)，接收端根据时间标记计算 RTT，这样减轻了 LPM 的影响。但是 TFMCC 初始化过程比较费时。TFMCC 也不能确保协议的友好性，因为在接受端计算出速率后，发送端所选的瓶颈接收端代表 CLR(current limiting receiver)可能是不合适的。

RLA(random listening algorithm)是基于窗口调整的单速率组播拥塞协议的一个典型代表。RLA 是 TCP SACK 在组播应用中的功能扩展。由于 RLA 使用随机监听技术，发送端随机地对来自接收端的拥塞信号产生反应，所以不会发生 LPM 问题。基于窗口的调整机制的 RLA 具有较好的公平性，但是 RLA 没有反馈抑制机制，因此协议的可扩展性不理想。

MTCP(multicast TCP)[26]是一个针对可靠组播的拥塞控制协议。基于窗口调整的单速率 MTCP 具有较好的 TCP-Friendly。由于 MTCP 中的每个节点都会转发其子节点瓶颈链路信息到自己的父节点上，因此接收端总是可以收到全部瓶颈链路信息。为此 MTCP 使用逻辑树进行丢失重传和反馈聚合，从而避免了 LPM 和反馈爆炸问题。但是 MTCP 协议实现复杂，需要建立逻辑树，每个节点都必须具备缓存、修复和拥塞监控等功能。另外，MTCP 中的接受端对每一个分组进行确认的工作方式限制了协议的可扩展性。

PGMCC(pragmatic general multicast congestion control)[27]也是一个基于窗口调整的单速率组播拥塞控制协议，在使用"代表"的工作方式上，与 TFMCC 很相似。协议从组成员中选择一个瓶颈接收端作为代表，负责对每个接收到的分组进行确认。PGMCC 的"代表"技术避免了 LPM 问题和反馈爆炸问题。PGMCC 也不能确保协议的 TCP-Friendly，因为在每个接收端使用 TCP 流量模型计算出期望的速率后，发送端所选出的最低速率接收端可能是不正确的。

4.3 分层组播拥塞控制

RLM(receiver-driven layered multicast)[28]是为传输视频数据设计的早期分层组播协议之一。RLM 将视频数据分为多个层，发送端在每层使用独立的组播组发送，接收端订阅第 1 层，开始接收数据。一段时间后，如果没有分组丢失，它就周期性地加入试验(join experiment)订阅下一层；如果有分组丢失，接收端取消最新订阅的层，加入试验失败可能增加其他接收端的拥塞。RLM 使用接收端驱动(receiver-driven)机制提高了协议的可扩展性，但是 RLM 的友好性不好，也没有考虑接收端之间的加入/离开同步问题。

Vicisano 改进 RLM 之后，提出了 RLC(receiver-driven layered congestion control)[29]。RLC 在数据分层时，模仿 TCP 的 AIMD 行为：按指数递增分配每层的带宽，加入层的等待时间也呈指数增长。一旦发生分组丢失，接收端立即取消最新订阅的层，从而使接收速率减半；在没有分组丢失的情况下，接收速率随加入新层的等待时间呈比例增加。为了减轻加入试验带来的额外拥塞，RLC 减少了加入试验的次数。RLC 的接收端只在同步点 SP(synchronization point)处加入新层，每层的 SP 数量呈指数递减，改善了接收端间的同步加入/离开问题。但是 RLC 仍然存在许多缺陷，比如在高 RTT 情况下，可能出现 TCP-unfriendly 的情况；RLC 要求数据支持分层；RLC 无法充分利用带宽，同时各 RLC 流量间不能公平地使用带宽。

FLID-DL(fair layered increase/decrease with dynamic layering)[30]是 Byers 等人为了改善 RLC 的缺陷而提出的一种协议。主要的改进有：协议在发送端使用 Digital Fountain[31]编码，使得分层方案更加灵活。为了减少加入和离开延迟，FLID-DL 引入了动态分层(dynamic layering)方案；为了保证 FLID-DL 的 TCP-Friendly，FLID-DL 使用基于 TCP 流量模型的公平分层增加/减少(fair Layered increase/decrease)方案对动态分层进行补充。由于没有考虑 RTT 对于网络性能的影响，FLID-DL 在高 RTT 情况时，也可能出现 TCP-unfriendly 的情况。另外，FLID-DL 的加入/离开操作比较频繁，因此给路由协议造成的开销很大。

4.4 组播拥塞控制的未来研究方向

现有的 TCP 流量模型多数是建立在一定的假设条件之上，只能符合 Internet 或者共享瓶颈链路部分情形。事实上，这个模型可能与真实情况不相符合，改善现有的 TCP 流量模型是一个可行的研究方向。

IP 组播技术要求路由器的支持，因此在现有的 Internet 中一直无法得到广泛的应用，传统的 IP 组播模型趋于理想化，不适合商业应用。于是研究者们开始试图建立新的组播模型(比如 EXPRESS[31])来解决组播拥塞。

由应用层负责组播路由是一个较好的研究方向。因为这种技术符合 Internet 的设计思想[28]，将组播的复杂性从网络转移到端系统，在网关上只进行少量的操作。而且这种技术直接使用传统的单播拥塞控制，避开了组播拥塞控制的难题。应用层组播最主要的问题是如何合理地建立逻辑树，这类似于传统的组播 QoS 路由问题。

拥塞控制属于传输层协议。对于使用基于树的技术的拥塞控制协议，在很多情况下无法直接获得网络层的路由信息。是否应将组播拥塞控制放到网络层，与路由结合，这是一个正在争论的问题。

5 控制理论

Van Jacobson 奠定了 TCP 拥塞控制的基本框架，但由于拥塞控制对 Internet 的健康发展起着至关重要的作用，此后许多学者对它进行了更为细致而深入的研究，发现了许多新问题，因此需要引入新的策略与算法来完善和解决它们。

从控制的角度看，Internet 是一个典型的复杂自适应系统，可以看成一个分散控制与决策问题：大量用户分享资源(路由器、物理线路和服务器等)，资源中各个部分一直处在动态环境中相互影响、相互作用，并且在延迟的、冲突的信息基础上做出决策。这使得分析各种控制策略的相互影响以及它

们对整个网络稳定性的影响变得更为困难。尽管如此，最近几年中，人们还是在利用控制与优化理论分析现有拥塞控制的稳态与动态性能以及设计新的拥塞控制算法方面做了大量的工作，取得了良好的开端。

开环控制典型的例子就是资源预留(如 RSVP 协议)。这类控制机制较适用于音频和活动图像业务。这种方法简单、直接，但由于要事先精确确定业务特性几乎是不可能的；同时为保证服务质量、避免拥塞，往往需要预留多余的网络资源(Over allocate resources)，很容易造成网络资源利用率低下。

闭环控制对于大带宽延时(BwD)的高速网络往往是无效的，因为如果 BwD 乘积大就意味着在信源承认网络拥塞时，已有大量未受控的分组存在于网络中[32]。

Low 等[33]提出了基于优化理论的 TCP/AQM 对偶性模型。该模型把发送速率当原始变量，把拥塞度量当对偶变量。TCP 拥塞控制就转化为求解具有适当效用函数的最优速率分配问题的分布式算法，在理论上可以分析网络在平衡状态的吞吐量、数据丢失率等性能。

文[33]给出了常见的 TCP/AQM 策略，如 Reno/Drop Tail，Reno/RED 和 Vegas/Drop Tail 等的 3 元组(F，G，U)的具体形式，并分析了它们的稳态性质。最近，Paganini[34, 35]在上述模型基础上，利用反馈控制理论，进一步分析了基于优化的拥塞控制的稳定性和鲁棒性。此外，Kelly[36]基于优化理论提出了另一类拥塞控制框架，并利用 Lyapunov 稳定性理论分析了拥塞控制系统的稳定性。

在不考虑定时机制的情况下，Misra 等[34]提出 TCP/AQM 微分方程模型，利用这种微分方程，Hollot 等[34]通过理论分析和实验证明了经典 PI 控制律优于 RED 算法，基于经典控制理论和 Smith 原理的拥塞控制思想。Mascolo[35]利用 Smith 的优点，将时延系统转换为无时延系统，设计了一个相对稳定的拥塞控制算法。事实上，通过分析发现[74]，现有的端对端 TCP 拥塞控制也是一个 Smith 预测器。

大量以往的闭环控制系统的研究中，所提出的速率调节算法，如基于前向或后向显式拥塞标识系统以及基于显式速率 ER 控制系统等，往往是直观的，在它们的速率调节算法推导过程中，很少系统地利用控制理论方法。此外，在以往的基于速率的 ABR 控制系统的研究中，很少考虑 VBR 流的影响，这是现有的基于速率的拥塞控制方法引起不稳定性的重要原因之一。在闭环拥塞控制系统中，加入更加系统的控制理论方法，是未来的一个重要研究方向。

基于传统控制理论的网络拥塞控制是通过建立一个严格的系统数学模型，根据网络拥塞控制的目标来进行端系统的速率调节器的设计和参数配置，从而提高网络运行的鲁棒性，以保证不同用户所要求的 QoS，有效地避免网络拥塞。常用的这类调节器有：P 调节器[36]、PD 调节器[37]和 LQ 调节器[38]等。

然而，这些数学模型并不能真实地模拟事实上属于非线性随机系统的 Internet。同时，在网络建模的过程中，本身就对参数的选择设定了限制条件，因此网络性能得不到期望的提高效果。随着智能控制理论的发展，人们把智能控制理论引入到网络拥塞控制当中。目前此类研究热点主要集中在模糊逻辑控制和神经网络控制。

模糊逻辑 FCC 调节器(Fuzzy congest controller)[39]是一种基于显示速率调节算法的非线性调节器，分布运行于每个交换机中。它采用两输入、单输出的模糊控制器结构，主要由四部分组成：模糊器、规则库、推理机和解模糊器。最终控制形式简单，易于实现。FCC 瞬变响应快，具有较低的端-端时延且链路利用率高。但是这类控制方案很难制定出一套全面准确、行之有效的语言控制规则，而且其控制精度不是很好。

神经网络 NNC 调节器(Neural network controller)[40]是采用具有三层 4/4/1 结构的基于 BP 神经网络的速率调节器，可用于解决 ATM 网络用户接口 UNI 的拥塞控制问题，主要是面向 CBR 和 VBR 服务的。NNC 学习过程可看作一个专业化增强学习形式，它利用性能指标函数 J 不断地动态调节神经网络的权值，以达到控制网络拥塞的目的。NNC 可实时测量网络参数，具有在线学习和并行处理功能，并且自适应能力强。但是它的学习时间较长，瞬态响应迟钝。NNC 的学习和控制算法的收敛性还需要进一步研究。

Hespanha 等[41]提出一个采用“去尾”策略的 TCP 拥塞控制的混杂系统模型。Li 等[42]利用混杂控制系统模型探讨了多个 TCP 共享带宽的暂态行为。混杂系统的分析与控制是近年来控制理论中一个备受关注的研究方向。

研究者把 El Faro l 酒吧问题模拟为 Internet 网络拥塞控制的一个简化模型[43, 44]。这样网络拥塞控制的关键就是如何协调多个“参与者”的行为，使“网络瓶颈”(酒吧)中人数接近最优值。Zam Brano 证明了酒吧人数的经验结果收敛于一组相关均衡，事实上这就相当于在适当的信号装置下博弈“条件”Nash 均衡[44]。这是目前的一个研究热点。

另外，由 El Faro l 酒吧问题而产生了一个活跃的研究方向是由 Challet 和 Zhang 提出的“少数博弈”问题[45, 46]。或者考虑更加复杂的博弈(存在更多的酒吧)，对于网络拥塞瓶颈的多个资源有效利用的研究也是重要课题之一。

6 研究趋势和未来发展方向

高速发展的 Internet 已成为人类进入信息社会的一个主要标志。但随着其复杂性的不断增加，Internet 能否持续稳定地发展便成为一个令人关注的问题。作为下一代 Internet 应用的一个关键性支撑技术，经过几十年的发展，拥塞控制技术日臻成熟。

本文分析了拥塞控制基本机制，总结了 Internet 端到端拥塞控制技术的研究现状，介绍了该领域的最新研究进展。讨论了拥塞控制领域的研究热点，这些研究热点可以总结为以下几大类：

对于 Internet 拥塞控制的研究，大多采用依赖于知觉的启发式设计，然后通过仿真试验来进行验证之后根据经验改进算法。这种设计模式确实为拥塞控制发掘了不少新的方法与策略。但是，对于这种设计模式在理论上没有一个准确的理解和把握，一旦应用的背景发生变化，性能将无法保障。这是目前的研究难点。

TCP 端到端的反馈控制机制中的自相似问题研究得还不太清楚。自相似问题对网络拥塞控制算法性能非常重要。把价格机制概念引入到拥塞控制中研究当中来，是目前的一个研究热点。

多播中的拥塞控制虽已提出了许多解决方案，但是目前还没有一套完备的组播拥塞控制标准。目前的多播拥塞控制协议大多是在传输层实现的，网络层的多播拥塞控制则显得不成熟。

从控制理论观点出发介绍了基于传统控制理论方法和智能控制方法的一些典型算法，并对这些算法进行了详细的性

能比较，阐述了 Internet 中拥塞控制系统的研究方面所作的努力。近些年来，非线性规划理论[44]和系统控制理论[46]被引入到拥塞控制的研究中来，在算法的性能分析和评价方面给出许多有价值的结论，这将进一步推动拥塞控制的研究。

在今后一段时间内，针对现有的 TCP 协议本身的改进，包括对 TCP 中各种机制的改进和对 TCP 在各种网络环境下优化的研究。另外，公平性理论模型及支持移动计算的相关机制等方面，还存在许多有待解决的问题。

尽管端到端拥塞控制在目前是有效的，但路由器支持的拥塞控制，特别是“主动队列管理”更有希望成为以后的主流方式。由于 Internet 用户在快速地增长，仅仅靠端节点使用端到端的拥塞控制是不够的，网络层本身必须具有拥塞控制的机制。如何充分发挥 IP 层在拥塞控制中的作用问题，是目前研究的热点之一。

最后需要指出的是：由于拥塞控制算法的分布性、网络的复杂性和对拥塞控制算法的性能要求，在算法策略上涉及到数据流的调度算法、缓存管理技术等网络资源的分配，使得拥塞控制算法的设计具有很高的难度，只有通过通信、控制和数学等多学科的共同努力，才有望获得突破性的成果。

参 考 文 献

1 Jacobson V. Congestion avoidance and control. ACM Computer Communication Review, 1988,18(4): 314～329

2 Jacobson V. Congestion Avoidance and Control. IEEE/ACM Transaction Networking, 1998, 6 (3): 314～ 329

3 Gevros P, Crowcroft J, Kirstein P, et al. Congestion control mechanisms and the best effort service model. IEEE Network, 2001, 15(3): 16～26

4 Steves W. TCP Slows Start, Congestion Avoidance, Fast Retransmit, and Fast Recovery Algorithms. RFC2001, 1997

5 Meom C. A new approach to model the stationary behavior of TCP Connections. IEEE Computer Society, 2000

6 Peterson L L, Davie B S. Computer Networks: a System Approach. Morgan Kaufmann Publishers, 2000

7 Jain R. Congestion control in computer networks: issues and trends. IEEE Network Magazine, 1990,4(3): 24～30

8 Veres A, BodaM. The chaotic nature of TCP congestion control. In: Proc. IEEE INFOCOM 2000, Tel Aviv, Israel, CA: IEEE Computer Society, 2000

9 Chaintreau A, Baccelli F, Diot C. Impact of network delay variation on multicast sessions with TCP-like congestion control. In: Ammar M. ed. Proceedings of the IEEE INFOCOM. Anchorage: IEEE Communications Society, 2001. 1133～1142

10 Floyd S, Handley M, Padhye J. A comparison of equation-based and AIMD congestion control. 2000. http://www. aciri. org/floyd/ papers. htmls

11 Widmer J, Denda R, Mauve M. A survey on TCP-friendly congestion control. IEEE Network, 2001,15(3):28～37

12 Thompson K, Miller G J, Wilder R. Wide-Area Internet traffic patterns and characteristics. IEEE Network, 1997,11(6): 10～23

13 Floyd S, Fall K. Promoting the use of end-to-end congestion control in the Internet. IEEE/ACM Transactions on Networking, 1999,7(4): 458～472

14 Floyd S, Jacobson V. On traffic phase effects in packet-switched gateways. Internetworking: Research and Experience, 1992, 3 (3): 115～156

15 Legout A, Biersack E W. PLM: Fast convergence for cumulative layered multicast transmission schemes. In: Drushel P. ed. Proceedings of the ACM Sigmetrics. Santa Clara, CA: ACM Press, 2000. 13～22

16 Floyd S, Henderson T. The New Reno Modification to TCP's Fast Recovery Algorithm. RFC 2582, February 1999

17 Stevens W. TCP Slow Start, Congestion Avoidance, Fast Retransmit, and Fast Recovery Algorithms. RFC 2001, 1997

18 Rejaie R, Handley M, Estrin D. RAP: An end to end rates based congestion control mechanism for realtime streams in the Internet. In: Proceedings of IEEE INFOCOM 1999 , New York , 1999. 1337～1345

19 Seshan S, Stemm M, Katz R H. Network measurement architecture for adaptive applications. In: Sidi M. ed. Proceedings of the IEEE INFOCOM. Tel Avlv: IEEE Communications Society, 2000. 285～294

20 Padmanabhan V N. Addressing the challenges of Web data transport :[Ph D Thesis]. Berkeley, CA: University of California, Berkeley, 1998

21 Widmer J, Denda R, Mauve M. A survey on TCP-friendly congestion control. IEEE Network, 2001,15(3):28～37

22 Rhee J, Balaguru N, Rouskas G. MTCP: scalable TCP-like congestion control for reliable multicast. In: Doshi B. ed. Proceedings of the IEEE INFOCOM, New York: IEEE Communications Society, 1999. 1265～1273

23 Rhee I, Ozdemir M, Yi Y. TEAR: TCP emulation at reciver flow control for multimedia streaming:[Technical Report]. NCSU,2000. http://www. csc. ncsu. edu/

24 Widmer J, Handley M. Extending equation-based congestion control to multicast applications. In: Floyd S. ed. Proceedings of the ACM SIGCOMM, San Diego: ACM Press, 2001. 275～286

25 Floyd S, Handley M, Padhye J, et al. Equation-Based congestion control for unicast applications. In: Floyd S. ed. Proceedings of the ACM SIGCOMM, Stockholm: ACM Press, 2000. 43～56

26 Rizzo L. PGMCC: a TCP- friendly single-rate multicast congestion control scheme. In: Floyd S. ed. Proceedings of the ACM SIGCOMM, Stockholm: ACM Press, 2000. 17～28

27 Vicisano L, Rizzo L, Crowcroft J. TCP-Like congestion control for layered multicast data transfer. In: Charny A. ed. Proceedings of the IEEE INFOCOM, San Francisco: IEEE Communications Society, 1998. 996～1003

28 Holbrook H W, Cheriton D R. IP multicast channels: EXPRESS support for large-scale single-source applications. In: Chapin L. ed. Proceedings of the ACM Sigcomm, Cambridge: ACM Press, 1999

29 Low S H. A duality model of TCP and queue management algorithms [DB/OL]. http://net. lab. caltech. edu, 2001-09-15/2001-10-25

30 Low S H, Lapsley D E. Optimization flow control (I): Basic algorithm and convergence [J]. IEEE/ACM Trans on Networking, 1999, 7 (6): 861～875

31 Athuraliya S, Low S H. Optimization flow control (II): Implementation [DB/OL]. Http://netlab. Caltech. edu, 2001-09-15/2001-10-25

32 Paganini F. Flow control via pricing: A feedback perspective [DB/OL]. Http://www. ee. ucla. eduö～paganini, 2001-09-20/2001-10-25

33 Paganini F. On the stability of optimization based flow control

[DB/OL]. http://www.ee.ucla.edu? ~paganini, 2001-09-20/2001-10-25

34 Kelly F P, Maulloo A , Tan D. Rate control for communication networks: Shadow prices, proportional fairness and stability [J]. J of Operations Research Society, 1998, 49 (3) : 237~252

35 Waldvogel M, Rinaldi R. Efficient topology-aware overlay network. ACM Communications Review, 2003,33(1):101~106

36 Hollot C, Misra V, Towsley D, et al. On designing improved controllers for AQM routers supporting TCP flows. In: Proceeding of INFOCOM 2001, Anchorage, Alaska,2001. 1726~1734

37 Christiansen M, Jeffay K, Ott D, et al. Tuning RED for Web traffic. In: Proceedings of ACM SIGCOMM 2000, Stockholm, Sweden ,2000. 139~150

38 Newman P. Backward Explicit Congest Notification for ATM Local Area Networks. In: Proc. of IEEE GLOBECOM '93, Houston, TX, 1993. 719~ 723

39 Rohrs C E, Berry R A. A Linear Control Approach to Explicit Rate Feedback in ATM Networks. In: Proc. of IEEE IN FOCOM '97, Kobe, Japan, 1997. 277~ 282

40 Zhao B Y, Huang L, Stribling J, et al. A resilient global-scale overlay for service deployment. IEEE Journal on Selected Areas in Communications, 2004,22(1):41~53

41 Habib I, Tarraf A , Saadawi T. A Neural Network Controller for Congest Control in ATM Multiplexers. Computer Networks and ISDN Systems, 1997, 29 (3) : 325~334

42 Hespanha J P, Bohacek S, Obraczka K, et al. Hybrid modeling of TCP congestion control [A]. Hybrid Systems: Computation and Control [C],Berlin: Springer Verlag, 2001. 291~304

43 Shenker S. Making greed work in networks: A game-theretic analysis of switch service disciplines [J]. IEEE/ACM Trans on Networking , 1995, 3 (6): 819~831

44 Kelly F P, Maulloo A , Tan D. Rate control for communication networks: shadow prices, proportional fairness and stability. Journal of Operations Research Society, 1998,49(3):237~252

45 Zhang X Y, Liu J C, Li B,et al. CoolStreaming/DONet: A data-driven overlay network for live media streaming. In: Znati T. ed. Proc of the IEEE INFOCOM, Miami: IEEE Press, 2005. 2102~2111

46 Hollot C V, Misra V, Towsley D, et al. On designing improved controllers for AQM routers supporting TCP flows. In: Sengupta B. ed. Proceedings of the IEEE INFOCOM, Anchorage, Alaska: IEEE Communications Society, 2001. 1726~1734

(上接第 5 页)

51 Christen P,Churches T, Zhu J X. Probabilistic name and address cleaning and standardization. The Australian Data Mining Workshop, November 2002. available at:http://datamining.anu.edu.au/projects/linkage.html

52 邱越峰,田增平.一种高效的识别相似重复记录的方法[J].计算机学报,2001

53 Wang R Y. A product perspective on total data quality management [J]. Communica-tions of the ACM, 1998

54 Neal R. Probabilistic inference using Markov chain Monte Carlo methods. CRGTR -93-1. Department of Computer Science, University of Toronto,1993

55 Pon R K, Cardenas A F. Ddata quality inference. In: IQIS Workshop,2005

56 Rahm E,Do H H. Data Cleaning: Problems and Current Approaches. IEEE Data Engineering Bulletin, 2000

57 Little R,Rubin D B. Statistical analysis with missing data. New York: John Wiley and Sons,1987

58 Wang R Y, Kon H B,Madnick S E. Data quality requirements analysis and modeling [C]. In:Proc. of Ninth ICDE, 1993

59 Ananthakrishna R,Chaudhuri S,Ganti V. Eliminating Fuzzy Duplicates in Data Warehouses [C]. In:Proceedings of VLDB,2002. 586~597

60 Sarawagi S, Bhamidipaty A. Interactive deduplication using active learning [C]. In: The Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining,2002. 269~278

61 Chaudhuri S,Ganjam K,Ganti V,et al. Robust and efficient fuzzy match for online data cleaning [C]. In: Proceedings of the 2003 ACM SIGMOD International Conference on on Management of Data,2003. 313~324

62 Chaudhuri S, Ganti V, Motwani R. Robust identification of fuzzy duplicates [C]. ICDE, 2005

63 Churches T, Christen P, Lim K,et al. Preparation of name and address data for record linkage using hidden Markov models [C]. 2002

64 de Waal T, Quere R. A fast and simple algorithm for automatic editing of mixed data [J]. Journal of Official Statistics, 2003,19 (4): 383~402

65 Dasu T, Johnson T. Exploratory data mining and data cleaning [M]. John wiley, 2003

66 Dasu T, Johnson T. Hunting of the snark: finding data glitches using data mining methods [R]. AT&T lab, 1999

67 Dasu T, Johnson T, Muthukrishnan S, et al. Mining database structure; or, how to build a data quality Browser [C]. ACM SIGMOD, 2002

68 Verykios V S, Moustakides G V. A Bayesian decision model for cost optimal record matching [J]. VLDB Journal ,2003, 12(1): 28~40

69 Winkler W E. Methods for evaluating and creating data quality [J]. Information System,2004,29: 531~550

70 Winkler W E. Set-covering and Editing discrete data [J]. In: Proc. of the Section on Survery Research Methods. American statistical association, 1997

71 Winkler W E, Draper L A. The SPEER edit system. Statistical data editing (volume 2); methods and techniques, united nations. 1997

72 Winkler W E,Petkunas T F. The DISCRETE edit system. Statistical data editing (volume 2); methods and techniques, United Nations. 1997

73 Fan W, Lu Hongjun, Madnick S E, et al. Discovering and reconciling value conflicts for numerical data integration. Information Systems,2001,(26) :635~656b

74 Low Wai Lup, Lee Mong Li, Ling Tok Wang. A knowledge-based approach for duplicate elimination in data cleaning [J]. Information Systems,2001,26(8):585~606

75 Wu Xintao, Barbara D. Learning missing values from summary constraints [R]. ACM SIGKDD Explorations Newsletter, 2002

76 Wu Xintao, Barbara D. Modeling and Imputation of Large Incomplete Multidimensional Datasets. In: Proceedings of the International Conference on Data Warehousing and Knowledge Discovery. Aix-en-Provence, France, Septemeber 2002

77 Zhu Xingquan, Wu Xindong,Chen Qijun. Eliminating class noise in large data sets [C]. In:Proc. of 20th International Conference on Machine Learning ,2003

78 Yair Wand, Richard Y W. Anchoring data quality dimensions in ontological foundations [C]. Communication of ACM, 1996