前向主动网络拥塞控制算法及其性能分析

王 斌,刘增基,李红滨,张 冰
(西安电子科技大学综合业务网国家重点实验室,西安 710071)

摘 要: 本文提出了一种基于主动式网络(Active Networks)技术的拥塞控制算法 FACC(Forward Active Networks Congestion Control).与传统的 TCP(Transport Control Protocol)相比,FACC 算法通过在网络结点直接提供拥塞检测和拥塞控制机制,大大缩短源端点的拥塞反应时间,从本质上提高了网络拥塞检测和控制的性能,从而提高了终端用户的平均吞吐量.文中还利用计算机仿真研究了 FACC 算法在各种网络条件下的性能,并与传统的 Tahoe,Reno,NewReno 及 SACK TCP 协议做了对比.结果表明无论网络中存不存在非受控数据流时,FACC 控制算法均能明显地提高用户终端的平均吞吐量,并且由于采用 FACC 控制算法而增加的网络结点运算迟延也很小.

关键词: 主动式网络; 拥塞控制算法
中图分类号: TP393 **文献标识码**: A **文章编号**: 0372-2112 (2001) 04-0483-04

Forward Active Networks Congestion Control Algorithm and Its Performance Analysis

WANG Bin, LIU Zeng-ji, LI Hong-bin, ZHANG Bing
(National Key Lab. on Integrated Services Network, Xidian University, Xi'an 710071, China)

Abstract: This paper proposes a congestion control algorithm FACC (Forward Active Networks Congestion Control) based on Active Networks. The comparison with traditional TCP (Transport control protocol) shows that by providing congestion detection and congestion control mechanism in network nodes, FACC greatly reduces the congestion reaction delay of source endpoint and consequently increases the average throughput of each endpoint. Computer simulations are given to analyze the performance of FACC algorithm under various circumstances and compare FACC with Tahoe, Reno, NewReno and SACK TCP protocols. The result of simulation shows that FACC can obviously increase the average throughput of each endpoint with or without the cross traffic in network, resulting in little additional computation delay of the network node.

Key words: active networks; congestion control algorithm

1 引言

主动式网络(Active Networks)[1~3]是一种新兴网络技术,它的中心思想是利用网络结点中提供的标准网络应用编程接口,向用户和网络业务供应者提供一个"开放"的网络控制机制,网络应用编程接口允许应用实体利用网络物理资源构建并管理适合自己所需的业务.通过在网络结点提供标准的网络应用编程接口,它可以使网络具备强大的处理能力并提供丰富的新型业务.

目前 Internet 上广泛使用的拥塞控制协议是 Tahoe TCP[4],基于它的改进协议主要有 Reno TCP[5],NewReno TCP[6] 以及 SACK TCP[7~9]协议.深入研究以上几种协议可以看到:这些协议本质上都是使用诸如确认,超时,重复确认等隐含信号来推断网络状态,并利用反馈来修正数据源的发送窗口,控制注入网络的业务量以缓解网络拥塞.其中一直存在的问题是:网络拥塞的检测和控制不是由发生拥塞的网络结点及时和主动的进行,而是在端到端的基础上由源端通过各种隐含信号推测出来的.这不但延缓了对网络拥塞的检测和控制,还有可能造成更严重的网络拥塞.上面这些改进协议在这一问题上都未能提出较好的解决方案.

针对以上问题并结合借鉴主动式网络的思想,本文提出了一种基于主动式网络的拥塞控制算法,前向主动网络拥塞控制(Forward Active Networks Congestion Control,FACC).它与以上这些 TCP 协议的最大区别在于 FACC 将拥塞检测与拥塞控制机制引入到网络结点.当网络结点发生拥塞而造成数据包丢失时,具备 FACC 控制能力的网络结点可以立即向传送此数据包的上游结点和源端发送指示消息,指示其执行相应的拥塞控制功能.这样就大大缩短了传统 TCP 协议中源端必须等到重发定时器超时或收到三个重复的确认后才能检测到

收稿日期:1999-12-27;修回日期:2000-09-13
基金项目:国家 863 计划通信技术主题,信息网前沿技术研究专项资金(No. 863-317-01-10-99)

网络拥塞的响应时间,从而提高用户终端的平均吞吐量.

本文仿真实验研究了 FACC 算法在各种网络条件下的性能,并与传统的 Tahoe,Reno,NewReno 及 SACK TCP 协议做了对比.结果表明无论网络中存不存在非受控数据流(Cross traffic)时,FACC 控制算法均能明显地提高用户终端的平均吞吐量,并且由于采用 FACC 控制算法而增加的网络结点运算处理迟延也很小.

本文第 2 节详细描述了 FACC 算法,结合一个简单的仿真模型解释了此协议的各个组成部分及其所实现的功能原理;第 3 节给出了 FACC 控制算法的仿真试验及其性能分析;最后得出结论并介绍进一步的工作.

2 FACC 控制算法

FACC 控制算法包含一个消息单元 FACCIndication 和三个功能模块 FACCQueueSpy,FACCFilter 和 FACCSlowStart.

FACCIndication 消息单元用于在 FACC 各功能模块之间传送有关网络拥塞指示消息[10,11].它包括两个部分:数据包序号和数据流标识.数据包序号用于表示在拥塞结点被中丢弃的数据包在所属数据流中的顺序号,其值通常等于此数据包 TCP 头中的顺序号;数据流标识用于表示在拥塞结点被丢弃数据包所属的数据流,在 FACC 中利用被丢弃 IP 数据包头中源/目的地址和 TCP 报头中源/目的端口号来唯一的标识网络中的数据流.

FACCQueueSpy 和 FACCFilter 都位于网络结点,由用户或网络业务提供商通过主动式网络的业务动态加载服务将这两个功能模块加载至网络结点上[12].FACCQueueSpy 主要功能为网络结点中的队列监视,当通过结点的数据流负载达到一定程度,可能或已经导致结点的拥塞时,FACCQueueSpy 结合结点中的队列管理算法对数据流进行相应的控制.当 FACCQueueSpy 决定对造成网络结点拥塞的数据流进行控制时,它向数据流的上游结点和数据流源端发送 FACCIndication 消息,要求该数据流上游结点和源端分别进行相应的数据流控制.与 ICMP 的源抑制策略不同的是,FACCIndication 将要求上游结点主动进行数据流控制,以尽快的减轻拥塞结点的负载.若 FACCIndication 在传输过程中丢失,则用户终端仍可依据标准的 TCP 协议进行拥塞控制.现在的 Internet 绝大部分的结点采用 FIFO&DropTail 队列管理策略,本文仿真试验中结点队列管理也采用这种策略.

FACCFilter 的主要功能为响应数据流下游结点发出的 FACCIndication,主动地对造成下游结点拥塞的数据流进行控制.FACCFilter 可以采用的控制策略主要包括丢弃、重定向和缓存等方法,本文仿真试验中 FACCFilter 采用丢弃策略.重定向和缓存策略将是进一步的研究内容.

FACCSlowStart 位于产生数据流的源端,它结合用户终端的标准 TCP 拥塞控制协议对 FACCIndication 进行响应.当具备 FACCSlowStart 功能的源端收到 FACCIndication 时立即进行响应,重发在拥塞结点丢失的数据包,并进入 TCP 的 SlowStart 状态.

图 1 为一个简单的具备 FACC 控制算法的网络拓扑图.图 2 给出了在此拓扑结构中 FACC 控制数据包发送的全部过程.在图 1 中,结点 0 和 1 为用户端点,其具备了 FACCSlowStart 功能.结点 2～5 均为网络结点,数据流如图中实线箭头所示,数据源为 Ftp 流.其中结点 2,3 具备 FACCFilter 功能,结点 4 具备 FACCQueueSpy 功能.值得指出的是,在实际应用中网络结点应同时具备 FACCFilter 和 FACCQueueSpy 功能.不失一般性,设拓扑结构中各结点之间链路带宽为 1Mb/s,传输迟延为 10ms;端点 0,1 传输层协议采用标准 Tahoe TCP 协议,其初始 cwnd 为 1,awnd 为 20 个数据包,数据包长为 1000bytes;结点 4 中队列管理采用 FIFO & DropTail,队列长度为 20.从用户端点 0 流向结点 5 的数据流经过结点 4,(以下所用到的数据流如无特殊说明均指此数据流,数据包指此数据流中的 TCP 数据包).

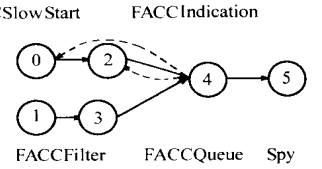

图 1 具备 FACC 控制算法的简单网络拓扑

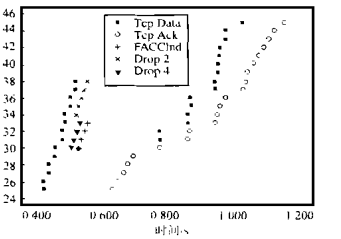

图 2 具有 FACC 控制的数据包发送

图 2 中,数据流的第 30 包数据到达结点 4 时,结点 4 中队列已满并开始丢包,此时 FACCQueueSpy 立即向此数据流的上游结点 2 和源端 0 发送 FACCIndication,告知其数据流在网络结点 4 由于拥塞而产生丢包.

上游结点 2 中的 FACCFilter 首先收到来自结点 4 的 FACCIndication,得知其下游结点产生了拥塞,并且由 FACCIndication 消息中得知数据流的第 30 包已被丢失,FACCFilter 立即在结点 2 的队列管理中设置过滤器,将第 30 号以后的数据包主动丢弃,以防止后续的数据包造成结点 4 拥塞的进一步恶化,如图 2 第 34 - 38 包数据在结点 2 被直接过滤.在 FACCIndication 由结点 4 向结点 2 传输的过程中,第 31 - 33 包数据到达结点 4,由于此时结点 4 的拥塞没有得到缓解,这几包数据也被丢弃.

终端用户 0 中的 FACCSlowStart 收到来自结点 4 的 FACCIndication 后得知第 30 包数据被丢弃,因此它立即重发第 30 包数据,然后进入 SlowStart 状态.由于结点 4 连续丢弃了第 30 - 33 包数据,因此会连续收到 4 个 FACCIndication,为防止 SlowStart 状态重入,FACCSlowStart 规定在 SRTT/2 的时间片内最多只响应一个 FACCIndication.通过比较不难发现,对于网络拥塞,FACC 比 Tahoe TCP 的反应要快得多.

3 FACC 控制算法的仿真与性能分析

图 3 给出本文仿真实验所使用的网络拓扑图,为了便于比较 FACC 算法和各种传统 TCP 算法在性能上的差异,它们均使用相同的网络拓扑结构.

图 3 中 m 个用户终端产生 Ftp 数据流业务,数据流经中

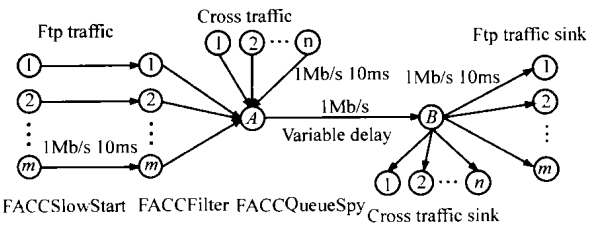

图 3　仿真试验网络拓扑图

间网络结点到达对应的 Ftp traffic sink. n 个用户终端产生 Cross traffic 数据流业务,数据流经中间网络结点到达对应的 Cross traffic sink(本实验中 m, n 均取 10).

设此拓扑结构中所有链路带宽为 1Mb/s,传输迟延为 10ms. 所有网络结点采用 FIFO&DropTail 队列管理算法,队列长度为 20 个数据包. 网络结点 A 和 B 之间链路传输迟延为可变量,其取值介于 10ms～300ms,此范围对应于从最短的局域网平均迟延到卫星通信上的链路平均迟延. 在仿真中通过改变此链路迟延就可改变链路的带宽迟延积. 有一点值得指出,在网络结点 A 和 B 之间链路带宽取 1Mb/s 是为了减少仿真试验的计算量和缩短仿真时间,并不会失去结果的一般性. 实际上,可以将此带宽设置为很宽,此时通过增加 Ftp 和 Cross traffic 的个数也可得到同样的仿真结果.

所有协议中设初始 cwnd 为 1,awnd 为 20,通过检测网络中各种 IP 数据包可知,长度为 64 - 128 和 1024 - 1532 字节的数据包所占比例最大,为了方便起见,在本仿真实验中设数据包长为 1000bytes. 在 SACK TCP 中 SACK Option 的个数为 3.

当采用 FACC 算法时,用户终端传输层采用标准 Tahoe Tcp 协议并增添了 FACCSlowStart 功能. 图中每一端点均通过一个具备 FACCFilter 功能的网络结点与网络结点 A 相连. 网络结点 A 具备 FACCQueueSpy 能力,当拥塞发生时,FACCQueueSpy 将通知数据流的上游结点和源端点,指示它们立即对网络拥塞进行处理.

这里利用 Cross traffic 模拟网络中的非受控数据流,Cross traffic 为服从指数分布的通断型数据流,在其通断状态平均维持时间为 500ms,通状态时,其数据速率为 200Kb/s. 在仿真中通过加大 Cross traffic 的个数可以增大通过网络结点 A 的非受控数据流.

3.1　不存在 Cross traffic 时 FACC 的性能

图 4 中,当网络中不存在 Cross traffic 时,FACC 控制算法对提高每个用户终端的平均吞吐量有较明显的作用. 随着链路迟延的增加,也即带宽迟延积的增加,FACC 表现出来的性能越来越好. 这主要是因为当链路迟延增加后,数据包的 RTT 时间增加,Tahoe,Reno,NewReno,SACK TCP 协议的定时器长度也随之增加,造成了用户终端对网络拥塞进行检测和控制时间的延长. 另外,由于这些 TCP 协议需要经过较长时间才能发现网络拥塞,在此之前终端将继续发送大量的数据包,这样就有可能造成拥塞结点情况的进一步恶化,造成数据包的大量丢失,从而也影响了平均吞吐量. 同时,从图中还可以看到,采用 Reno,NewReno,SACK TCP 协议时,每个用户终端的平均吞吐量都要好于 Tahoe TCP,这也证实了这些改进算法的性能优势.

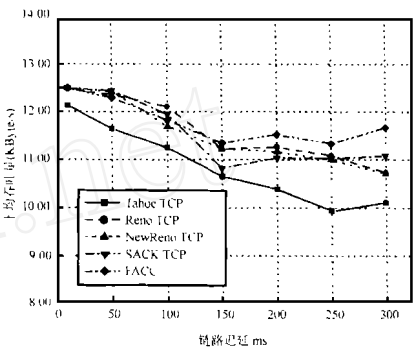

图 4　网络中不存在 Cross traffic 时的平均吞吐量

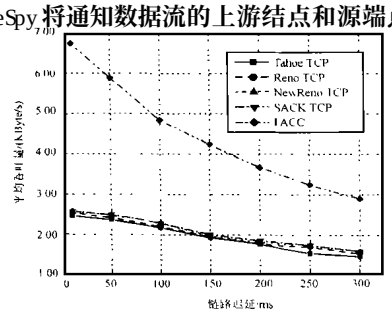

图 5　网络中存在 Cross traffic 时的平均吞吐量

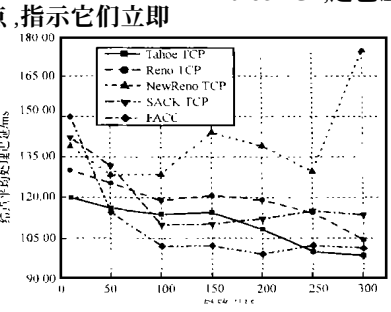

图 6　不存在 Cross traffic 时的网络结点处理时间

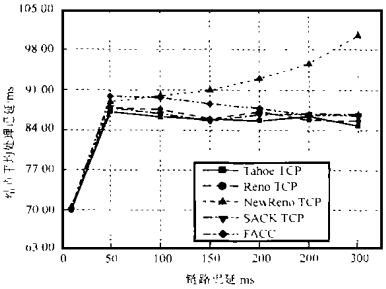

图 7　存在 Cross traffic 时的网络结点处理时间

3.2　存在 Cross traffic 时 FACC 的性能

图 5 显示了网络中存在 10 个 Cross traffic 时 FACC 对用户终端平均吞吐量的影响. 从中可以看到,Tahoe,Reno,NewReno,SACK TCP 协议反映出的性能基本相近,此时 FACC 显示出很好的性能,这主要是因为当网络结点发生拥塞时,由于 Cross traffic 对拥塞无反应,不会降低其发送速率,此时 FACC 能够确保用户终端以最快的速度对拥塞做出反应,进入 SlowStart 状态并立即重发丢失的数据包. 由于 FACCFilter 已经主动丢弃了可能进一步造成网络结点拥塞的数据包,因此网络结点的拥塞情况会得到一定的缓解,此时用户终端重发的数据包就能尽快通过拥塞结点,提高了吞吐量. 即使重发的数据包在拥塞结点又被丢失,FACC 会坚持重发数据包,这在总体效果上也提高了终端的平均吞吐量. 从仿真中还可以看出,随着网络中 Cross traffic 的增加,网络拥塞变得越来越严重,此时在采用所有协议情况下用户终端平均吞吐量均成下降趋势,但是 FACC 对这种重负载的适应性要比那些传统的 TCP 协议好,能

够为用户在网络重负载的情况下提供相对较高的平均吞吐量.

3.3 FACC 控制算法对网络结点处理能力的影响

图 6,7 分别显示了在网络中不存在和存在 Cross traffic 时,随着链路迟延的增加,数据包在其所经过的所有结点的平均处理迟延.从图中可以看出,FACC 和各种 TCP 协议的平均处理迟延很接近,随着链路迟延的增加,由于网络结点增加 FACC 控制算法所造成的延迟对总的线路迟延的影响越来越小.表 1 给出了网络中存在 10 个 Cross traffic 时,FACC 和各种传统 TCP 协议对用户终端平均吞吐量和网络结点平均处理迟延影响的一些仿真试验数据比较.考虑到 FACC 对平均吞吐量的提高,由于 FACC 控制算法所造成的处理延迟是可以接受的.

表 1 平均吞吐量和网络结点平均处理迟延仿真试验数据

	Tahoe	Reno	NewReno	SACK	FACC
平均吞吐量(KByte/s)	1.96057	2.019	2.06771	2.082	4.49886
性能提高(以 Tahoe 为准)		2.98 %	5.46 %	6.19 %	129.47 %
平均处理迟延(ms)	83.79014	84.33586	89.92014	84.42914	85.65343
性能提高(以 Tahoe 为准)		-0.65 %	-7.32 %	-0.76 %	-2.22 %

4 结束语

本文提出了一种基于主动式网络的拥塞控制算法 FACC,并利用仿真实验证实了它在及时检测与控制网络拥塞,提高终端用户平均吞吐量上的优势.FACC 拥塞控制算法在具体实现过程中还存在有待进一步研究改进的地方,如利用 RED 队列管理算法提高对各用户终端服务的公平性[13];利用定时器解决 FACCIndication 在传输过程中丢失而造成 FACC 算法失效;利用后向拥塞指示算法减轻 FACCIndication 给网络带来的额外负担等,这都将是下一步的研究重点.

将拥塞控制这一传统 Internet 中的经典算法放到主动式网络中进行研究,是我们在主动式网络领域的初步探索.从中可以看出,通过在网络结点提供灵活而丰富的控制处理功能,主动式网络确实能够提供比传统 Internet 更高性能和更丰富的业务,它的指导思想将很大程度上影响到下一代的 Internet.

参考文献:

[1] Y. Yemini and S. da Silva. Towards programmable networks [A]. IFIP/IEEE Int'l. Wksp. Dist. Sys.: Ops. and Mgmt. [C], L'Aquila, Italy Oct. 1996:132 - 137.

[2] Kenneth L. Calvert, Samrat Bhattacharjee and Ellen Zegura. Directions in active networks [J]. IEEE Communication Magazine, Oct. 1998, 36(10):72 - 79.

[3] D. L. Tennenhouse et al. A survey of active networks research [J]. IEEE Communication Magazine Jan, 1997, 35(1):80 - 86.

[4] V. Jacobson. Congestion avoidance and control [A]. Proc. SIGCOMM Symposium on Communication Architectures and Protocols, ACM SIGCOMM Stanford [C], CA, Aug. 1988:314 - 329.

[5] V. Jacobson. Modified TCP congestion avoidance algorithm [Z]. end2end-interest mailing list, April 30, 1990.

[6] S. Floyd, ACIRI, and T. Henderson. The NewReno Modification to TCP's Fast Recovery Algorithm [S]. RFC2582 Apr. 1999.

[7] S. Floyd. Issues of TCP with SACK [R]. Technical report, Mar. 1996. URL ftp://ftp.ee.lbl.gov/papers/issues sa.ps.Z.

[8] S. Floyd. SACK TCP: The sender's congestion control algorithms for the implementation "sack1" in LBNL's "ns" simulator (viewgraphs) [R]. Technical report, Mar. 1996. URL ftp://ftp.ee.lbl.gov/talks/sacks.ps.

[9] Matthew Mathis, Jamshid Mahdavi, Sally Floyd, and Allyn Romanow. TCP selective acknowledgment Options [S]. RFC 2018, 1996.

[10] D. Murphy. Building an active node on the internet [A]. M. Eng. thesis, MIT, June 1997.

[11] D. J. Wetherall and D. L. Tennengouse. The ACTIVE IP option [A]. Proc. 7th SIGOPS Euro. Wksp [C]. 1996:86 - 95.

[12] David Wetherall, Ulana Legedza and John Guttag. ANTS: A toolkit for building and dynamically deploying network protocols [J]. IEEE Network, May/June 1998, 12(3):12 - 20.

[13] S. Floyd and V. Jacobson. Random early detection gateways for congestion avoidance [J]. IEEE/ACM Trans. On Networking, August 1993, 1(4):397 - 413.

作者简介:

王 斌 生于 1973 年 6 月.分别于 1995 年和 1998 年在西安电子科技大学获通信与信息系统专业学士,硕士学位,研究方向为综合业务数字网与 IP 网络.从 1998 年至今,于西安电子科技大学 ISN 国家重点实验室继续攻读博士学位,研究方向包括主动式网络,QoS,网络管理和 ISDN.

刘增基 生于 1937 年 12 月,现为西安电子科技大学教授、博士生导师、综合业务网国家重点实验室主任、中国通信学会会士.当前主要从事宽带通信网络技术的研究.