NewReno 拥塞控制方式下路由器缓冲区容量研究

关洪涛, 王 东, 赵有健, 吴建平
(清华大学计算机科学与技术系, 北京 100084)

摘 要: 路由器缓冲区容量的设置问题是近年来路由器研究中的热点课题之一. 已有的研究主要集中在流量模型、网络拓扑及设置、路由器体系结构及设置、网络的动态性以及性能评价指标五个维度在研究. 本文在现有基于评价指标和流量模型所作研究的基础上, 提出了一种新的评价指标——流完成时间比. 该评价指标具有不依赖网络属性的优点. 本文使用这一评价指标进行了基于自相似流量的仿真实验分析, 对 SFCTR、AFCTR 和 FCU 这三个流完成时间比的相关性能指标进行了监测, 得出过大和过小的缓冲区容量都会造成性能下降的结论, 并给出了合理设置路由器缓冲区容量的方法.

关键词: 路由器; 缓冲区; TCP; 拥塞控制
中图分类号: TP93 文献标识码: A 文章编号: 0372-2112 (2009) 07-1440-07

Study on Router Buffer Size with NewReno Congestion Control Method

GUAN Hong-tao, WANG Dong, ZHAO Your-jian, WU Jiarr-ping
(*Department of Computer Science & Technology, Tsinghua University, Beijing 100084, China*)

Abstract: The problem of router buffer size is one of the hotspots in recent router research. Existing research focuses on traffic pattern, network topology and configuration, router architecture and configuration, network dynamic and performance evaluation criterion. Based on the existing study on evaluation criterion and traffic pattern, this paper proposes a new evaluation criterion named Flow-Completion Time Ratio, which does not rely on the network property. Based on this criterion, we perform simulations using self-similar traffic as traffic pattern. Observing SFCTR, AFCTR and FCU, we conclude that too large and too small buffer size will both degrade the network performance. We also provide reasonable method of setting router buffer size.

Key words: router; buffer; TCP; congestion control

1 引言

随着互联网技术的飞速发展, 骨干网传输速度也由最初的每秒几十字节提升到每秒几百亿字节. 路由器作为承载网络运行的重要设备, 其性能也经历了一次次革命性的提升. 当前骨干网核心路由器已经具有超过 40G 的端口速率和超过 1T 的交换能力. 为了能够更好的实现路由查找和分组交换的功能, 路由器内部通常设置一定数量的缓冲区. 作为路由器的重要组成部分, 路由器中的缓冲区技术一直是网络工作者研究的重点. 这其中最受关注的问题之一就是路由器内缓冲区容量的大小.

路由器内缓冲区的作用从不同角度大致可以归纳为以下两个方面. 从网络运营商的角度, 路由器中的缓冲区能够平滑不均匀的网络流量, 使输出链路尽量处于忙碌状态, 提高链路利用率; 而从网络用户的角度, 路由器中的缓冲区能够减少数据流在网络传输过程中由于冲突造成的丢包, 提高传输效率, 提升传输速率. 显然网络运营商更多关注的是网络运营的成本, 而用户更多关注的则是网络的性能. 这是一对相互矛盾却又相互依存的主题.

在网络诞生之初, 链路速率很低, 链路带宽资源非常有限, 因此最大化链路利用率就成为了矛盾的主要方面. 这也就促成了著名的带宽延迟积(BDP)理论[1,2]. 带宽延迟积理论的思想是路由器中缓冲区容量应该不小于链路带宽(C)与数据往返传输时间(RTT)的乘积. 这样就可以在一定条件下保证输出链路的 100% 利用率.

近年来, 随着光通信技术的广泛应用, 链路速率按照超摩尔定律不断增长, 此时链路带宽已不再像互联网诞生之初那样珍贵了. 而随着互联网向家庭的普及, 诸如 VoIP 等多媒体业务大量兴起, 用户对于网络性能的要求不断提高. 运营商为了更好地提供服务, 通常在平均链路利用率达到 30% 时就开始考虑更新设备, 以此

收稿日期: 2007-12-18; 修回日期: 2009-04-01
基金项目: 国家 863 高技术研究发展计划(No. 2007AA01Z219); 国家自然科学基金(No. 60773150)

为用户提供更好的服务. 成本和性能之间的天平发生了新的倾斜.

另一方面, 虽然至今网络设备制造商依然以 BDP 理论作为路由器缓冲区容量设计的重要参考, 但根据 BDP 理论, 即使以 40Gbps 作为链路速率, 以 200ms 作为 RTT 时间进行计算, 路由器缓冲区容量也需要 2GByte. 这样大的存储容量在能耗、成本等方面都为路由器的实现带来了很大的挑战. 而链路速率增长所遵循的超摩尔定律与单个存储芯片容量增长所遵循的摩尔定律之间的差异更将导致这一问题的不断加剧.

由于这些原因, 路由器中缓冲区容量问题在近些年成为了一个研究热点, 且至今仍存在争议. 当前路由器缓冲区容量研究主要是沿着五个维度展开的, 分别是: 流量模型、网络拓扑及设置、路由器体系结构及设置、网络的动态性以及性能评价指标[3]. 本文从评价指标和流量模型角度进行探讨, 分别提出了流完成时间比 FCTR(Flow-Completion Time Ratio) 的概念和以自相似流量作为流量模型进行研究的方法. 据我们所知, 本文是最早提出采用自相似流量模型进行路由器内缓冲区容量研究的文章.

本文第二部分首先介绍相关研究工作. 第三部分介绍本文提出的流完成时间比概念, 并给出相关定义. 第四部分根据流完成时间比的评价指标, 采用自相似流量作为网络流量在 ns-2 平台进行仿真实验. 文章的结论及下一步工作在第五部分给出.

2　相关研究工作

作为研究的热点, 路由器中缓冲区容量问题至今仍存在着较大的争议. 这主要是由于缺乏权威的手段进行验证. 要验证路由器缓冲区容量理论, 最具说服力的方法应该是根据理论设计路由器并在大规模真实网络上进行实验. 但设备制造商并不愿意在结论尚不明确的情况下承担减少路由器缓冲区容量所带来的风险, 这就使问题陷入了一个僵局. 另一方面, 互联网流量模型的不确定也为理论研究带来了极大困难. 因此, 当前研究路由器缓冲区容量大都采用近似模型分析结合仿真或小规模实验的方法. 这也就造成了存在多种不同的结论, 且结论之间存在巨大差异.

文献[1] 证明了在链路只存在单一长效 TCP 流的情况下, 如果缓冲区容量设置为带宽延迟积, 则链路利用率可以达到 100%. 同时还通过实验验证了在链路上只存在少量 TCP 流的情况下, 数据流之间将趋于同步, 缓冲区的需求依然是延迟与带宽的乘积. 这就是著名的 BDP 理论.

文献[4] 提出当链路上存在大量 TCP 流时, 这些数据流将趋于非同步. 根据中心极限定理, 数据流叠加的结果将造成锯齿状效应的减弱. 也就是说, 链路上总数据流量将趋于均匀. 要保证链路利用率 100%, 则缓冲区容量可以减小为 BDP 理论值的 $1/\sqrt{N}$, 其中 N 表示链路上长效 TCP 流的个数. 作者随后在文献[5, 6] 中通过实验对这一结论进行了验证. 然而, 文献[7] 通过实验得出了与文献[4] 相悖的结论, 即锯齿效应的存在和网络的不稳定性. 造成这一分歧的原因有可能来自于对流量和路由器设置的不同. 但对这一问题至今尚无理论分析结果.

文献[4] 以链路利用率作为评价指标得出的结论同样受到了文献[8] 的质疑. 后者根据在不同网络拓扑和网络动态性的实验环境下得出结论: 路由器缓冲区容量不止不应小于 BDP 理论的指导值, 更应该大于这一值, 这样才能避免发生大量的丢包.

此外, 光交换技术的发展对极小缓冲区的需求越来越强烈. 文献[7] 和[9] 认为在接入网速率很低的条件下, 如果可以容忍 10% 到 15% 的链路带宽损失, 那么路由器缓冲区仅需具有缓存几十个数据包的能力, 就可以获得较小的丢包率. 文献[10] 则提出, 如果使用 Paced TCP 协议, 即使不是在接入网速率很低的条件下, 结论同样成立.

上述研究从不同维度出发进行不同的配置和组合得出了迥异的结论, 各种结论的建议缓冲区容量甚至相差了好几个数量级. 可以预见, 关于路由器中缓冲区容量问题在今后相当长一段时间还将继续争论下去.

3　流完成时间比的提出及相关定义

性能评价指标是缓冲区容量研究中非常重要的一个维度. 当用户浏览 web 网页、传输文件或发送邮件的时候, 他们最为关心的问题是网页能否快速打开, 文件能否尽快下载完[11], 多媒体数据流传输速率是否稳定[12]. 因此, 有学者建议使用流完成时间 FCT(Flow Completion Time) [13] 或文件传输时间 FDT(File Delivery Time) [14] 作为评价指标. 这两个评价指标明显是从用户角度出发的, 而且具有直观、清楚的特点. 以文件传输时间为例, 相同的链路, 相同的文件, 相同的链路拥塞情况, 越小的传输完成时间当然意味着越好的性能. 但作为以时间为度量的标准, 它们对于度量的前提条件有特别的要求. 在不同条件下的比较将失去意义. 例如: 传输一个 100MB 的文件和传输一个 1MB 的文件, 其文件传输时间显然没有可比性; 或者, 同样传输一个 100M 的文件, 使用 10G 的链路和使用 56K 的链路也没有可比性; 此外, 链路只有少量数据流的情况与链路上存在大量数据流的情况也同样不具可比性.

为减轻上述评价指标对客观条件的过度依赖, 本文提出流完成时间比的概念. 与流完成时间和文件传

输时间不同的是, 流完成时间比表示的是一种比较关系. 比较的对象分别是: 流完成时间和在相同客观条件下数据流在某种虚拟假想网络完成所需要的时间. 这一思想与公平队列调度研究中引入 Shadow GPS 的思想类似, 目的在于消除由于客观条件不同所带来的不可比性.

为了定义流完成时间比, 这里首先对用于比较的虚拟假想网络进行说明. 在上一节中, 我们提到了路由器缓冲区研究的五个维度, 其中之一是路由器体系结构及设置. 当前绝大部分研究都是基于 OQ 交换结构路由器的, 当然也有少量研究[15]涉及了诸如 CIOQ 等其它交换结构路由器. 在虚拟假想网络中, 路由器设定为 OQ 交换结构并采用 GPS 调度算法[16]. 虽然 GPS 本身是流体流模型下的调度算法, 不能直接用于分组调度, 但其分组化的近似算法 WF^2Q[16]和 SRR[17]等都能够获得与 GPS 接近的性能. 因此, 从长期统计的角度, 我们可以将由分组化的近似算法所实现的路由器调度近似看作由 GPS 实现的路由器调度.

在全部由 GPS 调度路由器构成的虚拟假想网络中, 如果路由器对其调度的每个流分配相同的带宽, 那么结合网络拓扑等其它条件, 每个数据流将得到一个确定的带宽. 这里假设主机按照每条流获得的带宽以均匀速率发送, 也就是说忽略 TCP 造成的流传输速率的波动性, 那么对于一个确定的数据流, 它在虚拟假想网络的完成时间也就确定了.

下面给出单流完成时间比 SFCRT(Single Flow Completion Time Ratio), 平均流完成时间比 AFCTR(Average Flow Completion Time Ratio)和链路流完成效率 FCU(Flow Completion Utilization)的定义.

定义 **1(SFCTR)** 如果数据流 f_i 的总数据发送量为 B_i, 其在虚拟假想网络所获得的带宽为 w_i. 其在虚拟假想网络的流完成时间为 t_i^*, 而其在真实网络中的流完成时间为 t_i. 将这条流的单流完成时间比SFCTR 简记作 F_i, 则有:

$$F_i = \frac{t_i}{t_i^*} = \frac{t_i}{B_i / w_i}$$

从定义中可以看出 SFCTR 表示的是数据流的真实流完成时间与其在虚拟假想网络的完成时间的比. 因此, 它是一个不具有单位的值. 而且由于我们无法确定数据流的真实流完成时间与其在虚拟假想网络的完成时间的大小关系, 因此, SFCTR 的值既有可能大于 1, 也有可能小于 1. 当然, 值越大表示流发送的真实时间相对越长, 也就是说实际传输性能相对越差, 反之亦然.

定义 **2(AFCTR)** 对于需要关注的 m 条特定数据流, 如果每条数据流的 SFCTR 值分别为 F_i, 定义 m 条数据流的平均流完成时间比 AFCTR 为这些流的单流完成时间比的算术平均值, 简记为$\overline{F}$. 则有

$$\overline{F} = \frac{\sum_{i=1}^{m} F_i}{m}$$

AFCTR 表示的是网络中某些用于观测而被选定的特定数据流的单流完成时间比的平均值, 它是一个统计意义的量, 反映被观测的一类数据流的统计性能.

定义 **3(FCU)** 对于网络中链路容量为 C 的某个链路L, 假设在 t 时间段内传输了 n 条数据流, 每条数据流已经完成的数据传输量为 B_i', 则链路 L 的链路流完成效率FCU 为:

$$FCU_L = \frac{\sum_{i=1}^{n} B_i'}{C \times t}$$

从定义中可以看出, 这里定义的链路流完成效率与传统的关于链路利用率的定义是不同的. 在 FCU 的定义中, 那些因为超时等原因而产生的丢弃数据没有算作有效流量, 取而代之的是数据流真正完成的数据传输量.

这三个定义是就同一个问题从不同的方面进行展开. 其中 SFCTR 是从单个流的角度做出的定义, AFCTR 是从统计平均的角度做出的定义. 它们的出发点都是用户性能. 而 FCU 则是从链路和网络的角度出发所做出的定义. 通过这三个指标可以从用户和网络两方面进行性能评价.

4 自相似流量下的仿真实验及分析

前文提到不同的结论源自不同的性能评价指标和流量模型, 文献[4]等研究都是基于泊松到达的流量模型基础上的, 主要考虑长效流的聚合特性. 而 Crovella 等人经过研究大量的实测流量数据后认为, 网络上的 web 流和 FTP 流具有明显的自相似性[18]. 其流的大小和流量的到达都符合长相关性和重尾特性, 可以用 Pareto 模型来模拟, 并计算了自相似参数. 这为输入流量模型的设计提供了重要参考. 本文的仿真实验采用 ns-2 仿真平台, 以 Pareto 模型构建输入流量, 使用FCTR 评价指标分析 TCP NewReno[19]下路由器缓冲区容量与网络和用户性能的关系.

图 1 和图 2 表示了 New Reno 协议数据流发送过程中拥塞窗口的变化过程. 图 1 表示的是数据流开始传输时经历的慢启动和拥塞避免的过程. 图 2 表示的是数据流进入相对稳定传输后的过程.

实验中网络拓扑结构采用图 3 所示的哑铃型结构, 这也是相关研究广泛采用的一种拓扑结构. 在这一拓扑结构中, 每个输入输出链路容量是 1G, 同时具有很大

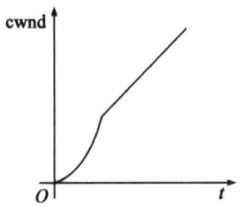

图1　NewReno慢启动及拥塞避免过程

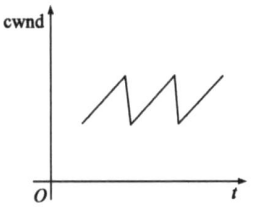

图2　NewReno持续传输过程

的缓冲区和很小的 RTT 时间, 以保证不会在这些链路上产生丢包. 节点 S' 和 D' 之间的链路容量也为 1Gb/s, RTT 时间 200ms, 于是就形成了单一的瓶颈链路.

首先设置一组过载情况下的网络流量模型进行实验. 每对源目的节点对(Si-Di) 间的流传输长度符合均值为 50pkt(包大小固定为 1000Byte) 的 Pareto 分布(考虑到现有网络的实际, 比 Crovella 实测数据略大) , α 值为 1.4. 流量到达符合均值为 0.12s 的 Pareto 分布, α 值为 1.2. n 为 400. 计算可得平均流量为 1.33Gbps, 为过载流量.

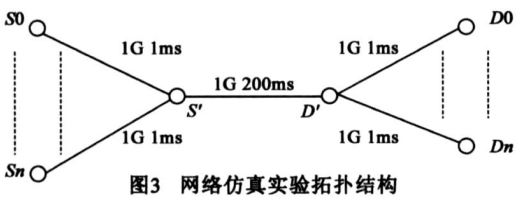

图3　网络仿真实验拓扑结构

根据 BDP 理论可以算得缓冲区容量应该为 25000 个数据包的大小. 于是这里首先以 25000 个数据包大小设置缓冲区容量, 进行 SFCTR 的性能分析.

图 4 中, 横轴为流的总数据量的大小, 纵轴为 SFCTR 值. 如图所示, SFCTR 随着传输流的长度增加而变小, 而且越小的流的 SFCTR 值的变化幅度越大. 当流足够长时, 它的 SFCTR 值分布在 1 附近, 其中一定比例的值小于 1. 这说明了在 NewReno 拥塞控制协议下实现的网络对于长流和短流的传输存在着不公平性, 对短流的传输效率远低于长流.

这主要是受到慢启动过程的影响. NewReno 拥塞控制协议的第一阶段是慢启动过程, 所有的数据流都要经历这一过程之后才有可能进入后面的相对稳态过程. 慢启动过程的传输速度是远小于后面的稳态过程的. 以一个 400pkt/RTT 带宽的链路为例, 如果有四个长效数据流存在, 则其在稳态下各获得 100pkt/RTT 的带宽. 而如果其中的一个数据流是只包含 15 个数据包的短流, 那么它也需要至少 4 个 RTT 时间, 实际发送速率仅为 3.75pkt/RTT. 很显然前者与后者存在数量级上的差距.

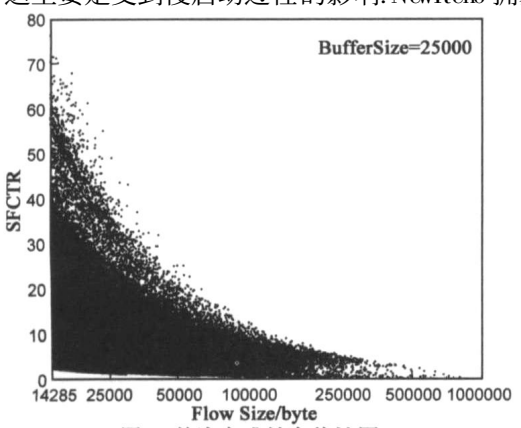

图4　单流完成效率统计图

此外, 通过图 5 可以看出, 随着数据流发送速率的提升, 缓冲区很快就被充满. 对于 DropTail 的队列管理策略, 其丢包率对于长流和短流是一样的. 但是, 由于丢包率很小和随机涨落, 不同的短流之间的性能差别会更大, 因此流传输延迟抖动增加.

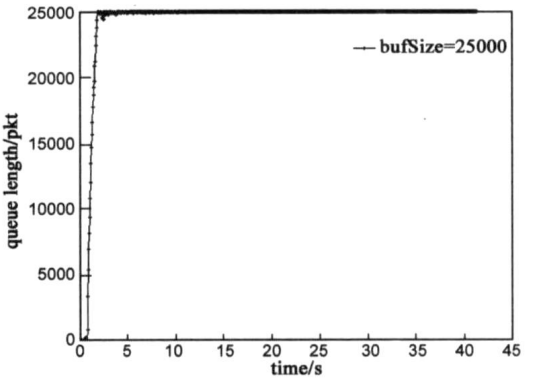

图5　单流情况下, 瓶颈链路缓冲区排队长度

上面的实验是为了说明根据 BDP 理论设置缓冲区容量情况下的 SFCTR 性能. 下面将研究不同缓冲区容量下的情况. 实验中的拓扑结构和流量模型不变, 缓冲区容量设置从 20—2C* RTT 之间均匀选取了几十个测量点. 在目标评价上, 采用了具有统计特性的 AFCTR. 从仿真最小流长 14285 字节(由 Pareto 模型参数计算出最小值) 开始, 按照每 1000 字节(1pkt) 递增划分区间, 在每个区间内求其 AFCTR. 实验结果如图 6, 7 所示. 为使结论显示更清晰, 图示选取了具有代表性的几个测试点数据.

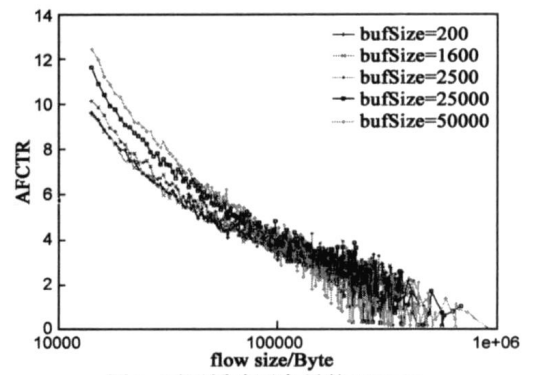

图6　不同缓冲区容量的AFCTR

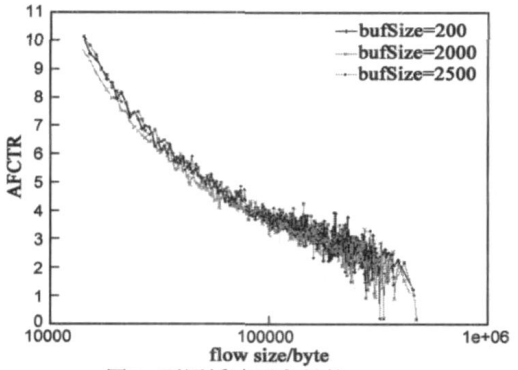

图7 不同缓冲区容量的AFCTR

图 6 描述了 20—2C* RTT 之间五个不同缓冲区容量设置(20pkt, 200pkt, 1600pkt, 2500pkt, 25000pkt, 50000pkt)的仿真结果. 从图中可以观察到两个现象: 第一, 所有缓冲区容量下的 AFCTR 都随着流的长度增大而减小. 这一点验证了上文中关于 SFCTR 的实验结果. 第二, 大的缓冲区更加偏爱长流, 而小的缓冲区对短流的服务效率更高. 为了能更加清楚的显示这个现象, 将上图以流长度等于 100000 Byte 为界左右分开, 分别表示为图 8(*a*) 和图 8(*b*).

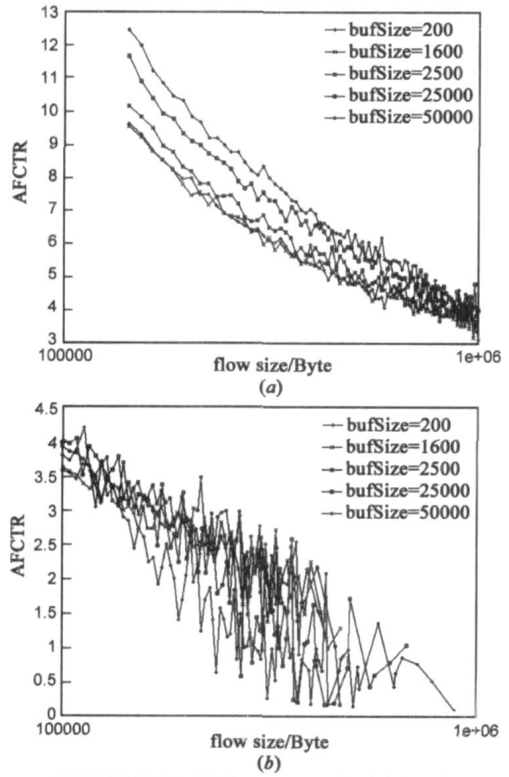

图8 (*a*) 不同缓冲区容量的FCU;(*b*) 不同缓冲区容量的FCU

图 8(*a*) 中显示, 对于较小长度的数据流, 当缓冲区容量大于 200pkt 时, 其容量越小, AFCTR 值越小, 即传输的效率越高. 图 8(*b*) 显示出, 对于长效流则是缓冲区容量越大, AFCTR 值越小, 传输效率越高.

由图 9 可知, 链路传输处于稳态时, 缓冲区不论容量大小, 一直是充满的, 也就是说任何数据包进入路由器总是要经过一个排队过程才能被转发. 因此, 随着缓冲区容量大小的变化, 链路延迟和丢包率之间存在着此消彼长的关系. 缓冲区越大延迟越大, 但丢包率越小. 此外, 丢包率的适度增加对短流影响较小. 这是因为在持续传输过程中, 一个流丢包后, 要损失一半的带宽, 短流由于本身占用的带宽比较小、传输时间短, 因此所受影响较小; 而长效流由于占用了很大的带宽, 损失一半带宽后再恢复到原来的值需要很长的时间. 而且, 在 Pareto 分布中长流所占的比重很小, 丢包后可以为短流让出更多的带宽, 所以在较高的丢包率情况下, 短流的效率要高一些. 这也就造成大缓冲区更加偏爱长流, 而短流在小缓冲区条件下效率更高的现象.

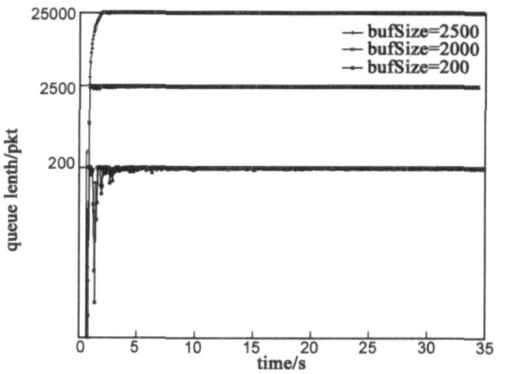

图9 不同缓冲区容量的链路排队长度

通过对图 7 和图 8 的分析可以看出, 缓冲区容量为 200pkt 和 1600pkt 的曲线基本重合, 这说明缓冲区容量在这个区间内的变动对链路性能的影响较小. 而对于更小的缓冲区如 20pkt, 它和缓冲区容量为 2500pkt 的 AFCTR 值曲线基本重合, 这是由于 20pkt 的缓冲区容量太小造成丢包率过大, 效率反而下降.

至此, 已经对从用户角度定义的 SFCTR 和 AFCTR 的性能进行了分析. 图 10 反映的是从网络和链路角度定义的 FCU 的性能.

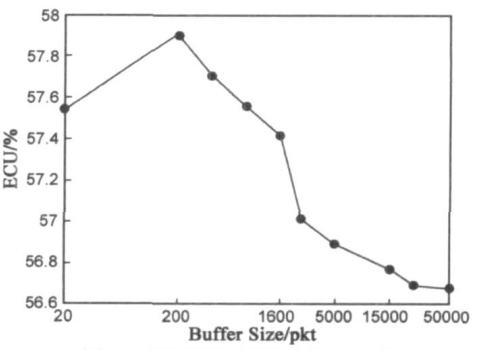

图10 不同缓冲区容量的FCU效率

由图 10 可知: 第一, FCU 值并不随着缓冲区容量的增加而一直增加. 当缓冲区容量大于 200pkt 时, 它随着缓冲区容量的增加而减小. 而当容量小于 200pkt 时,

FCU 也会随缓冲区容量减少而下降. 第二, 当缓冲区容量大于20pkt 时, FCU 值在 56. 6% – 57. 9% 范围内, 变化并不明显.

这是因为较小的缓存区容量虽然丢包率较大, 但链路的延迟小. 而过小的缓冲区则会由于丢包率过大导致网络性能下降. 实测 FCU 值较低主要是因为链路过载丢包严重和TCP 流出现了同步.

下面设置一组欠载情况下的网络流量模型进行实验. 这里采用相同的网络拓扑结构, 修改流量模型参数, 使得流的到达间隔加大. 具体设计如下: 相对前面的过载情况, 这里仅将流量到达改为符合均值为 0. 2s 的Pareto 分布, 其他参数均不改变. 计算可得平均流量为 0. 8Gbps, 为欠载流量, 链路负载为 80% .

首先我们对比一下欠载与过载流量下 SFCTR 值. 观察比较欠载和过载流的 SFCTR 图 4 和图 11 可知, SFCTR 值与传输流的长度的趋势关系没有改变, 只是在负载减轻后, 总体的流完成效率提高了. 少量的比较大的 SFCTR 值是自相似流模型的突发性造成的.

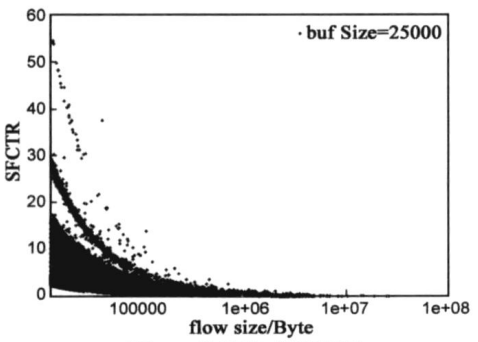

图11　单流完成效率图

进一步的实验观测 AFCTR 性能. 选取 20—2C * RTT 之间几十个测量点, 并选取了如图 12 所示的五个代表性的测量点进行分析. 可以看到, AFCTR 随流长度变化的总的趋势与前述过载情况相同, 只是值更小.

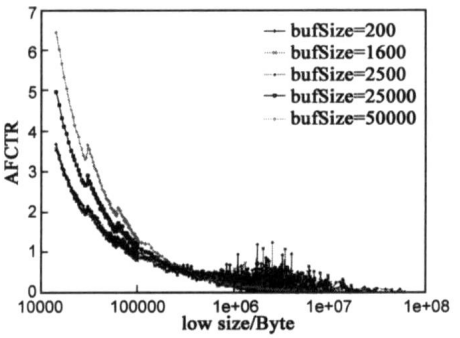

图12　不同缓冲区容量的AFCTR

图 13 显示了不同缓冲区容量下的 FCU 效率. 从图中可以看到, 在 80% 负载的情况下, 缓冲区容量对于 FCU 效率的影响和过载线路的情况下基本一致. 其效率的变化也不是非常的明显. 而且当缓冲区容量很小时, FCU 随缓冲区容量的减小下降得非常快.

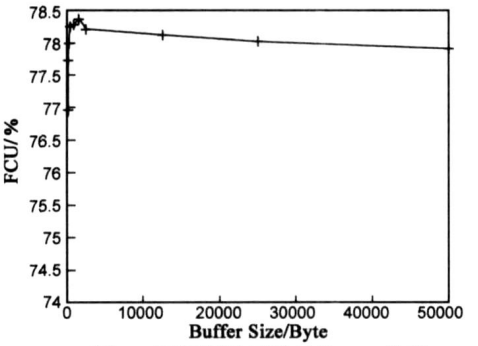

图13　不同缓冲区容量的FCU效率

通过观察发现, NewReno 拥塞算法在链路过载和欠载两种情况下, 实验现象基本一致. 在多流聚合情况下, 较小的缓冲区就可以达到较高的链路实际利用率, 而且对于短流用户有更高的效率. 但是过小的缓冲区反而会使性能下降. 根据实验观测, 本文建议对于 1G-10G 的链路, 可以选择 1MByte(相当于本实验 1000pkt) 左右的缓冲区容量.

5　结论和下一步工作

本文提出了一个研究路由器缓冲区容量的新的评价指标 ——流完成时间比, 并给出了流完成时间比的三个定义 ——单流完成时间比 SFCTR、平均流完成时间比 AFCTR 和链路流完成效率 FCU. 流完成时间比相对于之前的流完成时间和文件传输时间具有不依赖网络属性的优点. 通过以自相似流作为输入流量的仿真实验, 以及对 SFCTR、AFCTR 和 FCU 的性能分析, 文章认为过大或过小的路由器缓冲区容量都会造成网络性能的下降. 并基于实验结果得出结论: 对于普通 1-10G 带宽容量的链路, 大约 1MB 左右的缓冲区容量可以达到较好的网络效率, 同时使用户获得更快的响应时间.

本文中的流完成时间比的概念是为了解决流完成时间对于网络属性的依赖而提出的. 为了能够消除网络属性的影响, 文中提出采用比较的方法, 即对客观网络通过某种形式建立模型, 并通过比例的方式消除客观网络的影响. 在具体做法上, 文中采用了建立参考虚拟假想网络的方法, 这是借鉴了公平流调度研究中引入 Shadow GPS 的思想. 但是, 本文中提出的虚拟假想网络的方法本身是近似简化的, 还应该在此基础上进一步考虑其他因素, 如对应的 GPS 的带宽分配等, 这是我们下一步研究工作的重点. 此外, 我们在下一步的工作中还将分析更加一般化的拓扑结构和拥塞控制算法, 以得到更加普遍的结论.

参考文献:

[1] C Villamizar, C Song. High performance TCP in the ANSNET [J]. ACM SIGCOMM Compute Communication Review,

1994, 24(5): 45- 60.

[2] 谢高岗, 汤艳霞. 带宽测量实验研究及其算法改进[J]. 电子学报, 2002, 30(12A): 2142- 2145.
XIE Gao-gang, TANG Yan-xia, ZHANG Da-fang, LI Zhong-cheng. Experimental research on bandwidth measurement technology and method improvement[J]. Acta Electronica Sinica, 2002, 30(12A): 2142- 2145. (in Chinese)

[3] Y Ganjali, N McKeown. Update on buffer sizing in internet routers[J]. ACM SIGCOMM Computer Communication Review, 2006, 36(5): 67- 70.

[4] G Appenzeller, I Keslassy, N McKeown. Sizing router buffers [A]. Proceedings of the 2004 conference on Applications, technologies, architectures, and protocols for computer communications[C]. New York, USA: ACM Press, 2004. 281- 292.

[5] Y Ganjali, N McKeown. Experimental study of router buffer sizing[A]. Proceedings of the 8th ACM SIGCOMM conference on Internet measurement[C]. New York, USA: ACM Press, 2008. 197- 210.

[6] M Wang, Y Ganjali. Unifying buffer sizing results through fairness[R]. Technical Report, HR06-HPNG-060606, Stanford University, June 2006.

[7] G Raina, D Wischik. Buffer sizes for large multiplexers: Tcp queueing theory and instability analysis[A]. Proceedings of Next Generation Internet Networks 2005[C]. Washington, DC, USA: IEEE Computer Society Press, 2005. 173- 180.

[8] A Dhamdhere, C Dovrolis. Open issues in router buffer sizing [J]. ACM SIGCOMM Computer Communication Review, 2006, 36(1): 87- 92.

[9] M Enachescu, Y Ganjali, A Goel, N McKeown, T Roughgarden. Routers with very small buffers[J]. ACM SIGCOMM Computer Communication Review, 2005, 35(3): 83- 90.

[10] A Aggarwal, S Savage, T Anderson. Understanding the performance of TCP pacing[A]. Proceedings of IEEE INFOCOM 2000. Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies[C]. USA: IEEE Computer and Communications Societies, 2000. 1157- 1165.

[11] A Odlyzko. The Internet and other networks: Utilization rates and their implications[J]. Information Economics and Policy, 2000, 12(4): 341- 365.

[12] S He, S Sun, H Guan, Q Zheng, Y Zhao, W Gao. On guaranteed smooth switching for buffered crossbar switches[J]. IEEE/ACM Transactions on networking, 2008, 16(3): 718- 731.

[13] S Gorinsky, A Kantawala, J Turner. Link buffer sizing: A new look at the old problem[A]. Proceedings of ISCC 2005. 10th IEEE Symposium on Computers and Communications[C]. USA: IEEE Computer and Communications Societies, 2005. 507- 514.

[14] N Dukkipati, N McKeown. Why flow-completion time is the right metric for congestion control[J]. ACM SIGCOMM Computer Communication Review, 2006, 36(1): 59- 62.

[15] N Beheshti, Y Ganjali, R Rajaduray, D Blumenthal, N McKeown. Buffer sizing in all-optical packet switches[A]. Proceedings of OFC/NFOEC2006. Optical Fiber Communication Conference, 2006 and the 2006 National Fiber Optic Engineers Conference[C]. USA: IEEE, 2006. 5- 10.

[16] J Bennett, H Zhang. WF^2Q: worst-case fair weighted fair queueing[A]. Proceedings of IEEE INFOCOM'96. Fifteenth Annual Joint Conference of the IEEE Computer and Communications Societies[C]. USA: IEEE Computer and Communications Societies, 1996. 120- 128.

[17] S Ramabhadran, J Pasquale. Stratified round robin: A low complexity packet scheduler with bandwidth fairness and bounded delay[A]. Proceedings of the 2003 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications[C]. New York, NY, USA: ACM 2003. 239- 250.

[18] M Crovella, A Bestavros. Self-similarity in world wide web traffic: evidence and possiblecauses[J]. IEEE/ACM Transactions on networking, 1997, 5(6): 835- 846.

[19] S Floyd, T Henderson. The NewReno Modification to TCP's Fast Recovery Algorithm[S]. IETF RFC 2582, 1999.

作者简介:

关洪涛 男, 1980 年生于北京, 2003 年在清华大学计算机科学与技术系获得学士学位, 现在清华大学计算机科学与技术系网络所攻读博士. 研究方向为可扩展路由器交换结构.
E-mail: ght@csnet1.cs.tsinghua.edu.cn

王 东 男, 1978 年生于山东临清, 清华大学计算机科学与技术系硕士研究生. 研究方向为路由器缓存技术.
E-mail: wangmqi@csnet1.cs.tsinghua.edu.cn

赵有健 男, 1969 年生于甘肃会宁, 博士, 教授, 博士生导师. 主要研究领域为路由器体系结构, 交换与调度算法.

吴建平 男, 1953 年生于山东巨野, 博士, 教授, 博士生导师. 主要研究领域为计算机网络体系结构、协议工程学、互联网络.