ISSN 1000-9825, CODEN RUXUEW
Journal of Software, Vol.21, No.7, July 2010, pp.1576−1588
doi: 10.3724/SP.J.1001.2010.03812

E-mail: jos@iscas.ac.cn
http://www.jos.org.cn
Tel/Fax: +86-10-62562563

高速长距离网络传输协议*

任勇毛+, 唐海娜, 李 俊, 钱华林

(中国科学院 计算机网络信息中心,北京 100190)

Transport Protocols for Fast Long Distance Networks

REN Yong-Mao+, TANG Hai-Na, LI Jun, QIAN Hua-Lin

(Computer Network Information Center, The Chinese Academy of Sciences, Beijing 100190, China)

+ Corresponding author: E-mail: renyongmao@cstnet.cn, http://www.cstnet.cn

Ren YM, Tang HN, Li J, Qian HL. Transport protocols for fast long distance networks. *Journal of Software*, 2010,21(7):1576−1588. http://www.jos.org.cn/1000-9825/3812.htm

Abstract: The traditional TCP transport protocol has many drawbacks on Fast Long Distance Networks (FLDnet), and its transfer performance can not satisfy the requirement of increasing bulk data transfer applications. The UDP transport protocol has high transfer speed on FLDnet, but its reliability can not be guaranteed. This paper firstly analyzes the drawbacks of the traditional transport protocols on FLDnet, then, classifies and summarizes the main design principles of all kinds of enhanced transport protocols proposed in recently years and the research work on performance evaluation of transport protocols. Finally, some open issues and further research directions are proposed.

Key words: fast long distance network; transport protocol; congestion control; performance evaluation; bulk data transfer

摘 要: 传统的TCP传输协议在高速长距离网络中存在许多局限,其传输性能不能满足日益增长的海量数据传输应用的需求.UDP 传输协议尽管传输速率很高,却没有可靠性保证.分析了传统的传输协议在高速长距离网络中的局限,分类总结了近年来提出的各种改进的传输协议的主要设计思想以及在传输协议性能评价方面的工作,最后提出了目前研究中仍然存在的开放性问题和进一步的研究方向.

关键词: 高速长距离网络;传输协议;拥塞控制;性能评价;海量数据传输

中图法分类号: TP393 **文献标识码**: A

高性能计算、高速网络、海量存储等信息技术的飞速发展,为海量数据的分析、传输和存储创造了条件,e-Science 科研应用、HDTV 等大数据量应用因此得以发展和普及.同时,这些应用的飞速发展又对网络传输速率等性能提出了越来越高的要求.为了满足应用对于网络性能的需求,一方面需要有高性能的网络基础设施;另一方面也需要有高性能的网络传输协议.因此,高速网络和高性能传输协议已成为网络领域的研究热点.

* Supported by the National High-Tech Research and Development Plan of China under Grant No.2007AA01Z214 (国家高技术研究发展计划(863)); the Knowledge Innovation Program of the Chinese Academy of Sciences under Grant No.CNIC_QN_08004 (中国科学院知识创新工程青年人才领域项目)

Received 2009-06-12; Revised 2009-08-19; Accepted 2009-12-29

大量的e-Science科研应用需要在国际间通过高速长距离网络(fast long distance network,简称FLDnet)进行海量数据传输,这些应用对于网络传输性能有很高的要求.传统的TCP传输协议是针对低速、低延迟的网络而设计的,在FLDnet网络中有着很多的局限,其传输性能不能满足这种需求.而UDP协议尽管传输速率很高,但却没有可靠性保证.

近年来,许多研究人员对此问题进行了研究,在对传输协议的改进和传输协议的性能评价等方面都取得了一定的进展.本文综述了这一领域的主要研究工作以及近年来的研究进展.首先讨论了传统传输协议在FLDnet中的局限,然后介绍了针对FLDnet的传输协议的改进以及性能评价研究进展情况,最后提出了本领域的开放性问题以及进一步的研究方向.

1 传统传输协议在高速长距离网络中的局限

传统IP网络主要使用传输层的TCP和UDP协议传输.

Internet中绝大部分流量都是采用TCP协议传输.TCP协议在低带宽、低时延的LAN网络中运行得很好,但在FLDnet中性能很差.实验发现,在622Mbps带宽、300ms RTT时延的链路中,单个TCP流的传输速率仅有875KB/sec[1].TCP协议在FLDnet上性能差的原因主要有以下几点:

(1) 拥塞避免机制过于保守.TCP采用的AIMD(加性增加、乘性减小)拥塞窗口调整算法过于保守,在FLDnet中,由于往返时延RTT很大,一旦发生拥塞,拥塞窗口减小后,需要很长的时间才能恢复.文献[1]指出,在带宽为622Mbps、RTT为300ms、报文段大小为1460B时,TCP拥塞避免阶段所经历的时间长达41分钟,这严重制约了TCP协议的传输速率;

(2) 流量控制机制过于保守.TCP采用窗口限制发送流量,防止网络拥塞和接收端的缓冲区溢出.默认的最大窗口大小只有64KB,这个值对于低速、低时延网络和早期的低性能终端比较合适,但对于FLDnet,带宽时延积(bandwidth delay production,简称BDP)远大于这个值,使得链路管道容量利用率很低.另一方面,目前终端性能已经大大提高,内存早已达到2GB,4GB等容量,64KB的接收缓冲区大小则显得过于保守.

另外,TCP采用重复ACK和RTO超时来检测包丢失,并将所有包丢失事件(包括非拥塞引起的包丢失)默认为拥塞发生,导致了不必要的或幅度过大的拥塞反应产生.TCP采用肯定应答ACK确认的差错控制机制,发送方依赖于重复ACK来间接判断哪些报文需要重传,在大时延的FLDnet上效率不高.

由于以上TCP协议在FLDnet的各种局限性的存在,使得TCP协议在FLDnet上的传输性能很差.

UDP协议是非可靠的无连接传输协议.UDP比TCP简单,它只是增加了端口用来标识应用进程和校验以及用来检测错误数据包并简单丢弃错误包.UDP没有像TCP那样具有序列号、ACK和重传等可靠传输机制.此外,UDP没有像TCP那样进行拥塞控制和流量控制.因此,UDP在FLDnet上传输得很快.但是,UDP因为没有可靠的传输保证,不能满足很多应用可靠数据传输的要求.另一方面,由于UDP没有拥塞控制机制,当在分组共享的IP路由网络中使用时,如果进行大规模的数据传输,则有可能导致网络产生严重的拥塞.

2 传输协议改进

由于传统TCP协议和UDP协议在FLDnet中的性能存在很大的局限,因此,研究适用FLDnet的传输协议成为网络研究中的一个热点问题.研究人员在对标准TCP协议和已有改进协议的性能评价的基础上,不断地提出新的改进协议或设计新的传输协议.从文献调研的情况来看,绝大部分改进协议都是在TCP协议或UDP协议的基础上作了改进.因此,这些协议大致可以分为两类:一类是TCP改进协议;另一类是基于UDP的改进协议.图1对各种改进的传输协议进行了分类总结.

另外,还有其他改进的或新的传输协议,如SCTP[2],RTP[3],RTSP[4],DCCP[5]等,它们是针对Internet上实时多媒体流传输等应用而提出来的,强调的是在互联网中数据传输的实时性.由于不是专门针对高速长距离网络,因此本文不对它们进行详细讨论.

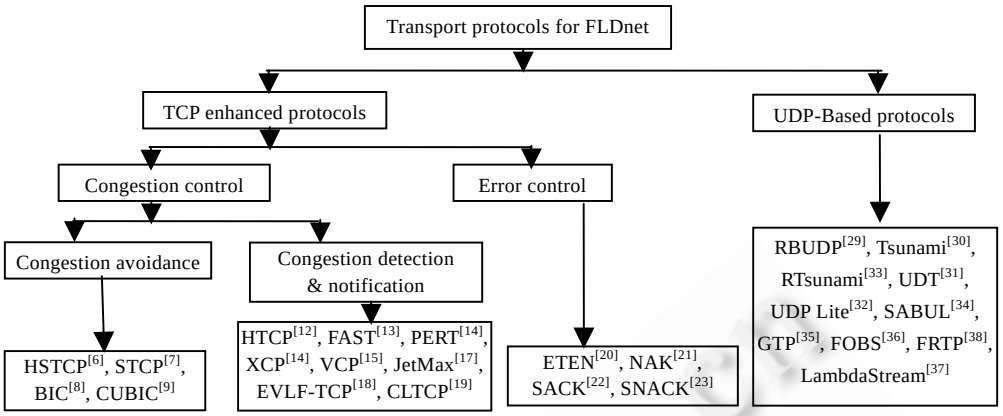

Fig.1 FLDnet transport protocols: A taxonomy

图 1 高速长距离网络传输协议分类

2.1 TCP协议改进

对传输协议的改进,主要包括拥塞控制机制的改进和差错控制机制的改进两个方面.大多数协议改进研究工作是对 TCP 拥塞控制机制的改进,而对 TCP 差错控制机制改进的研究相对较少.在对 TCP 拥塞控制机制的研究中,大量的研究集中在对 TCP 的 AIMD 窗口调整算法的改进上,而对于拥塞检测和拥塞通告等方面的研究相对较少.

2.1.1 TCP 拥塞控制机制的改进

高速网络 TCP 改进协议在拥塞控制机制方面的改进主要包括两类:一类是对 AIMD(additive increase and multiplicative decrease,加性增加倍乘减小)窗口调整算法的改进.已有的改进协议主要集中在对拥塞窗口 cwnd 的调整算法方面.另一类是对拥塞检测和通告机制的改进.

(1) 对 AIMD 窗口调整算法的改进

由于传统 TCP 的 AIMD 拥塞控制算法在 FLDnet 中表现出很多缺陷,尤其是拥塞恢复时间太长,因此对 AIMD 算法的改进成了一个研究热点,许多研究人员提出了各种改进算法和协议.下面介绍几种比较著名的改进协议机制:

Floyd 等人提出的 HSTCP(high speed TCP)[6]和 Kelly 等人提出的 STCP(scalable TCP)[7]采用了高速/低速模式切换.当拥塞窗口小于阈值窗口(legacy window size,简称 lwnd)时(即 *cwnd*<*lwnd*),HSTCP 和 STCP 采用传统 TCP 协议的 AIMD 窗口调整算法.当拥塞窗口大于阈值窗口(即 *cwnd*>*lwnd*)时,采用更为积极的窗口增长和更为缓和的窗口减少算法.这样,HSTCP 和 STCP 既与传统 TCP 兼容,又能更有效地利用带宽.HSTCP 的默认阈值窗口大小为 38 个分组,STCP 的默认阈值窗口大小为 16 个分组.

当拥塞窗口大于阈值窗口时,HSTCP 和 STCP 对标准 TCP 协议的算法进行了修改,具体为:

在拥塞避免阶段(congestion avoidance),标准 TCP 的 AIMD 算法是:

ACK: $w=w+1/w$ (收到 ACK 确认)

Drop: $w=0.5w$ (收到包丢失信号)

AIMD 算法的增加和减少参数分别固定为 1 和 0.5.

HSTCP 对此 AIMD 算法进行了修改:

ACK: $w=w+a(w)/w$ (收到 ACK 确认)

Drop: $w=(1-b(w))\times w$ (收到包丢失信号)

修改后的 AIMD 算法的增加和减少参数基于当前的拥塞窗口值而变化.

STCP 对标准 TCP 的 AIMD 算法的修改为:

ACK: $w=w+0.01$ (收到 ACK 确认)

Drop: $w=w-0.125w$ (收到包丢失信号)

HSTCP 和 STCP 都在很大程度上减少了拥塞恢复时间.比如在链路带宽为 10Gbps,RTT 为 100ms,分组大小为 1500B 且发生拥塞时,标准 TCP 的恢复时间长达 1.6 小时,而 HSTCP 只需要 12 秒[6].标准 TCP 协议对拥塞反应时将拥塞窗口减小一半,而 STCP 仅减少 0.125 倍.很明显,STCP 这种机制,具有很强的带宽抢占能力,会对使用标准 TCP 协议的背景流量产生影响.具体影响仍需进一步研究[7].文献[8]中的测试结果表明,HSTCP 和 STCP 都不是 RTT 公平的.

Rhee 等人提出的 BIC TCP(binary increase congestion TCP)协议[8]将拥塞控制视为一个搜索问题,通过包丢失给出当前发送速率(窗口)是否大于网络容量的反馈.当它得知一个分组丢失事件,BIC 与 STCP 一样,仅将拥塞窗口减少 0.125 倍.减小前的窗口大小设为 maximum,刚减小后的窗口大小设为 minimum.然后,BIC 采用这两个参数执行“二值搜索(binary search)”,寻找目标窗口.这种方法的原理是,因为网络丢包发生在新的最大窗口值附近,而不是在新的最小窗口值附近,目标窗口大小必然是在这两个值的中间.

BIC TCP 在高速网络中实现了较好的可扩展性(scalability)、公平性和稳定性.但是,BIC TCP 的拥塞窗口增长函数对于其他 TCP 协议流来说仍然太“激进(aggressive)”,尤其是在低速、低延迟网络环境中.而且,窗口控制的好几个不同的阶段(binary search increase,max probing,Smax 和 Smin)增加了协议实现和性能分析的复杂性[9].

CUBIC[9]是 BIC TCP 的一个改进版本,试图在保留 BIC TCP 的优点(尤其是稳定性和可扩展性)的基础上,简化窗口控制并增强它的 TCP 友好性.正如 CUBIC 这个名称所代表的意思(cubic,三次函数),CUBIC 将 TCP 的线性窗口增长函数修改为一个三次函数,以提高 TCP 在 FLDnet 中的可扩展性.CUBIC 的拥塞窗口增长函数如公式(1)所示:

$$W(t)=C(t-K)^3+W_{\max} \tag{1}$$

其中,C 是一个扩展因子,t 是自上次窗口减小(即上次丢包事件发生)以来消逝的时间,$W_{\max}$ 是上次窗口减少之前的窗口大小,K 是以上函数中 W 增加到 $W_{\max}$ 的时间周期.如果在此期间不再丢包,则 K 由公式(2)计算:

$$K=\sqrt[3]{\frac{W_{\max}\beta}{C}} \tag{2}$$

其中,β是丢包事件发生后的乘性减少因子.

CUBIC 只有一个窗口增长函数,简化了 BIC 的窗口增长函数.当窗口值与饱和点 $W_{\max}$ 值差距很大时,CUBIC 快速增长窗口值;当窗口值与饱和点 $W_{\max}$ 值差距很小时,CUBIC 缓慢增长窗口值.这一特点使得 CUBIC 在高带宽时延积网络中具有很好的扩展性,同时具有很好的稳定性,对标准 TCP 流也比较公平.

此外,针对拥塞窗口调整算法的改进还有很多方案.比如 Floyd 等人最近提出了新的 Quick Start 机制[10],该机制旨在当路径上有很大未用带宽时,可以允许发送端使用更高的发送速率,该机制需要中间路径上路由器的支持.Hacker 等人提出的 Stochastic TCP[11],采用统计方法来管理拥塞窗口.

(2) 对拥塞检测和通告机制的改进

传统 TCP 协议主要依靠重传定时器 RTO 超时和重复 ACK 来进行拥塞检测和通告.这种拥塞检测和通告方式在 FLDnet 中不够准确,有些研究人员提出了一些在终端进行拥塞检测的改进方法以及通过中间路由器进行拥塞通告的方法,下面对这些代表性协议进行介绍.

HTCP[12]主要采用上次拥塞事件以来消逝的时间Δ来检测网络拥塞的程度.Δ越大,说明网络拥塞程度越轻;反之,Δ越小,说明网络拥塞程度越严重.HTCP 主要根据由此方法检测到的拥塞程度相应地调整发送速率,其 AIMD 的增长因子α为Δ的函数,且α随 RTT 的变化而变化.AIMD 减小因子β的调整根据 RTT 的变化来决定.

FAST[13]根据队列时延和分组丢失来检测网络拥塞程度.当拥塞程度较轻时,队列时延是主要的拥塞信号;拥塞程度较重时,分组丢失是主要的拥塞信号.在正常的网络条件下,FAST 根据下面的公式(3)周期性地基于平均 RTT 更新拥塞窗口:

$$w \leftarrow \min\left\{2w,(1-\gamma)w+\gamma\left(\frac{RTT_{base}}{RTT}w+\alpha\right)\right\} \quad (3)$$

其中,w 为拥塞窗口大小,$\gamma\in(0,1)$,RTT_{base} 是至今观测到的最小的 RTT,RTT 是观测到的平均 RTT 值,α是一个正的协议参数,表示沿流量路径中路由器在平衡状态时总的队列分组数.在 FAST 的原型系统中,窗口更新周期是20ms.

从公式(3)可以看出,RTT 值反映出队列时延的变化,从而检测到网络拥塞状况的变化.当 RTT 变大时,说明拥塞程度更加严重,从而相应地减小拥塞窗口 w;反之,当 RTT 变小时,说明拥塞程度减轻,从而相应地增加拥塞窗口 w.

FAST 的提出者认为,基于时延的拥塞检测方法比基于丢包概率的检测方法更准确.但是在 FLDnet 中,RTT 的测量估计也是一个难题.

PERT[14]在终端主机上根据 RTT 的平滑估计值 $srtt_{0.99}$ 的变化,采用类似于 RED 的算法计算拥塞预测概率值.理论分析和仿真实验结果表明,PERT 协议具有很好的稳定性,但其预测精确度还需进一步加以研究.

以上协议主要在终端上对 TCP 的拥塞检测机制试图进行改进,在传输速率、稳定性和公平性等方面取得了一定效果,但其拥塞检测精确度还需进一步研究.另外,有些协议利用路由器配合进行显式拥塞反馈,由路由器向通信终端通告网络的拥塞状况,终端据此调整发送速率.比较典型的主要有:

XCP[15]为数据包增加了拥塞报头,由发送端写入当前的窗口值和 RTT 估计值,为路由器计算可分配带宽提供信息.路由器将吞吐率反馈信息写入报头反馈字段,发送端据此反馈信息来调整拥塞窗口.VCP[16]采用 IP 头中的 ECN 字段的两位作为负载因子,表征负载状况,在低负载区(低于 80%)采用乘性增加策略,在高负载区(80%~100%)采用加性增加策略,在过负载区(>100%)采用乘性减少策略.JetMax[17]的主要思想是,由路由器平均分配剩余带宽给各个流,以求在稳态下实现 Max-Min 公平性.EVLF-TCP[18]和 CLTCP[19]的主要原理是,每个路由器基于路由器容量、总输入流量及队列长度维护一个预分配速率因子 r,链路低载时逐渐增加 r,过载时减少 r.路由器将 r 值通告给终端,终端将 r 值视为链路容量上限,并据此调整发送速率.

以上这些协议由路由器进行拥塞检测和通告,需要路由器的支持,难以在实际网络中进行部署.

目前,大部分 TCP 改进协议局限于对 TCP 的拥塞窗口的 AIMD 调整算法做出改进.大部分改进协议的主要改进其实只是对 TCP 拥塞窗口调整函数的一个修正,并没有从整个拥塞控制,包括拥塞检测、拥塞通告和拥塞反应等整个过程来考虑.另外,这些改进协议是针对传统的 IP 共享网络,所以都强调公平性(fairness),并不是专门为了提高传输速率而作的改进.

2.1.2 TCP 差错控制机制的改进

除了对拥塞控制机制的改进以外,最近也有人提出对差错控制机制的改进.比如,Rajesh 提出显式传输错误通告(explicit transport error notification,简称 ETEN)机制[20].但是正如该文作者所指出的,ETEN 仍然还有许多问题有待继续研究,还不能真正在实际中使用.

对 ACK 确认机制的改进有 NAK(negative acknowledgement)[21],SACK(selective acknowledgement)[22]和 SNACK(selective negative acknowledgement)[23,24]等机制.

NAK 是否定应答机制,明确告诉发送端哪个包没有收到.这样,TCP 发送端可以确切知道重传哪个包,而不需要等待重复 ACK 或重传定时器 RTO 超时(timeout).NAK 每次只能通告一个需重传的包,在大延迟的 FLDnet 网络上,ACK 传输延迟较大,这种一次通知一个重传包的效率不高.

SACK 是肯定应答,在选项中明确收到了数据块.发送端根据接收到 SACK 指示可以知道这些块中间缺少的块,每个 SACK 可以指示 3 个洞(hole,即缺失块).这样,发送端也能确切知道需要重传哪些块,而无须等待很长时间.但是,SACK 仍然依赖快速重传算法来检测包丢失和触发重传.对于 FLDnet 来说,由于 RTT 时延很大,在接收到 4 个重复 ACK 之前,RTO 可能超时,触发快速重传算法,清除 SACK 信息.

SNACK 结合了 NAK 和 SACK 的优点.SNACK 选项中采用位向量(bit victor)标识接收到的报文段,用 1 表示正常报文段,0 表示需要重传的报文段.这样建立起位向量与报文段之间的映射关系,仅用一个位就能标识一

个洞,其效率有了很大的提高.SNACK 最初是针对大延迟的卫星网络而提出来的[23,24],后来也有研究人员将SNACK 应用于无线网络以提高 TCP 在无线网络中的性能[25−27].文献[28]提出,将 SNACK 应用于 FLDnet 并做出改进,提出了 SNACK-A TCP 协议,并采用模拟实验证明了 SNACK-A 能够提高 TCP 在 FLDnet 中的性能.

2.2 基于UDP的改进协议

由于 UDP 协议无可靠传输保证,一般不采用,只有对数据传输可靠性要求不高而又需要有很高传输速率的时候才采用.由于采用 UDP 传输速率很快,因此研究人员在 UDP 协议的基础上增加可靠性保证机制,提出了一些基于 UDP 的改进协议.这些改进协议通常是将数据信息和控制信息分开传输,采用 UDP 传输数据,用 TCP 传输控制信息.

基于 UDP 的改进协议主要有 RBUDP(reliable blast UDP)[29],Tsunami[30],UDT(UDP-based data transfer protocol)[31],UDP Lite(lightweight UDP)[32]等.

He 等人提出的 RBUDP 协议,基于 UDP,简单增加了 ACK 和重传机制以保证可靠性.RBUDP 首先采用 UDP 持续传输所有数据,接收方对收到的数据块进行记录,但并不发送 ACK.直到所有数据发送完成后,接收方收到 DONE 信号,它才发送一个 ACK,该 ACK 包含了已经成功收到的数据块的记录.发送方据此重新发送缺失的数据块.这个过程不断重复,直到所有数据成功被收到为止.

RBUDP 协议主要存在 3 个问题:第一,RBUDP 需要用户手动设定发送速率,用户在使用 RBUDP 传输数据之前必须测量链路可用带宽.但是,由于应用流量在变化,链路可用带宽通常是动态变化的.RBUDP 采用固定的发送速率,不能动态适应可用带宽的变化,无法充分利用链路带宽.而且由用户来设定发送速率,大大减少了 RBUDP 的可用性;第二,RBUDP 完全没有拥塞控制机制,如果发生拥塞将导致大量丢包,尤其是当设定的发送速率较大时;第三,对传输文件大小有限制.因为只有在发送完所有数据之后才能收到接收方的确认,发送端必须保留所有已经发送的数据以备重传之需.这样,如果文件太大而内存又不足以存储,就无法传输.

Meiss 等人提出的 Tsunami 协议,其基本原理与 RBUDP 类似.Tsunami 主要对 RBUDP 作了两点改进:首先,Tsunami 接收方不是等待所有数据发送完成,而是周期性地反馈发送方一个重传请求,并计算当前的错误率发送给发送方;第二,Tsunami 增加了基于丢包率的拥塞控制机制.发送方根据接收方反馈的错误率,通过调整包间延迟来控制发送速率,实现拥塞控制.另外,如果需要重传的块太多,发送方将从指定的块号处重新发送.

Tsunami 消除了 RBUDP 的文件传输大小限制,增加了简单的拥塞控制机制.但是,Tsunami 基于丢包率的拥塞控制机制过于简单,丢包率不能准确反映拥塞状况.而且,接收端计算出丢包率再反馈给发送端进行拥塞反应时,这个过程在 FLDnet 上需要较长时间,当发送端采取拥塞控制措施时,拥塞很可能已经解除.这种拥塞后再做反应的机制没有拥塞预防,有效性差.针对 Tsunami 协议拥塞控制算法过于简单、鲁棒性差的问题,文献[33]做出了改进,提出了 RTsunami(robust Tsunami)改进协议.实验结果表明,RTsunami 具有比 Tsunami 更高的传输速率和更好的鲁棒性.

Gu 等人提出的 UDT 协议比 RBUDP 和 Tsunami 更复杂,它与 TCP 比较相似.基于 UDP,除了增加可靠性以外,UDT 还增加了拥塞控制和流量控制机制.在可靠性方面,UDT 接收方以固定的时间段发送 SACK,并且只要检测到包丢失就发生否定式确认(NAK),显式反馈给发送方.在拥塞控制方面,UDT 采用所谓的 DAIMD(AIMD with decreasing increase)算法来调整发生速率(但并不是像 TCP 那样的拥塞控制窗口).为了区分拥塞和错误,UDT 对第 1 个丢包不作反应,只有在一个拥塞事件中有多个包丢失时才减少发生速率.此外,UDT 使用基于接收方的包对(packet pair)来估计链路容量.类似于 TCP,UDT 也使用流量控制窗口并限制发送方发送未确认的包.UDT 采用了 TCP 的许多机制或类似机制,比如拥塞控制和流量控制,并且像 TCP 一样采取了保证“公平性”的措施.这些机制使得 UDT 协议本身变得更加复杂,而且很大程度上限制了其传输速率.

由 Larzon 等人提出的 UDP Lite 协议是对 UDP 协议的改进.某些应用,比如视频,可以处理或允许很少量的位错误.但是,标准 UDP 协议由于 UDP 的校验和或者覆盖整个数据包,或者不作校验,如果检测到数据包中有错误,就简单地将数据包丢弃,而不递交给应用层程序.UDP Lite 对此进行了修改,提供一个可选的部分覆盖的校验和.当使用这个选项时,数据包被分为敏感部分(被校验和所覆盖)和非敏感部分(不被校验和覆盖).在非敏感

部分的错误将不会导致数据包被丢弃.当校验和覆盖整个数据包时,UDP-Lite 就相当于 UDP.与 UDP 相比,UDP-Lite 的部分校验功能提高了灵活性,有错误的数据包仍然可以被递交给应用程序.但是,这种功能必须得到链路层的支持,链路层不能将这种在非敏感部分出错的数据包丢弃,而这样的链路层很少.一般链路层都有 CRC 校验.

其他基于 UDP 的改进协议还有 SABUL(simple available bandwidth utilization lirbrary)[34],GTP(group transport protocol)[35],FOBS[36],LambdaStream[37],FRTP(fixed rate transport protocol)[38]等.SABUL 是 UDT 的前身版本.GTP 主要关注 Lambda Grids 上多点到一点的通信模式下的性能.FOBS 采用 UDP 传输数据,通过应用层的确认和重传机制来保证可靠性.LambdaStream 主要采用并行流传输方法.FRTP 是对 SABUL 的一种改进版本.

3 传输协议性能评价

传输协议的性能如何、有哪些优点和缺陷,这都需要对传输协议进行性能评价.对传输协议进行性能评价的目的,一是评价各种传输协议的性能优劣,二是发现各种传输协议的性能缺陷,为协议改进奠定基础.

评价标准是影响性能评价结果的一个关键因素.到目前为止,对于传输协议的性能评价标准仍在研究当中,存在很多争议,尚无统一标准.一般的性能评价标准包括吞吐率、延迟、丢包率等基本指标,其中最重要的是吞吐率指标.更全面的性能评价也包括公平性、友好性、鲁棒性等指标.

评价方法也是影响性能评价结果的重要因素.一般来说,对传输协议性能评价的方法主要有 3 种:一种是采用理论模型分析法,进行性能建模和分析;一种是采用模拟和仿真实验的方法;还有一种就是在实际网络中进行实际传输测试.从已有文献来看,NS2 是传输协议研究中使用最为广泛的模拟和仿真平台软件.采用 NS2 进行性能评价的方法使用最普遍,NS2 的模拟实验结果受到专业领域的普遍认可.

3.1 TCP改进协议性能评价

对 TCP 改进协议的性能评价,大多数研究通常是针对某个改进协议与标准 TCP 协议或其他协议进行比较来进行性能评价,比如,针对 FAST TCP 协议的性能评价工作非常多,单就 FAST TCP 协议的稳定性方面,就有许多采用数学模型或 NS2 模拟实验的方法进行的性能评价研究.但是,较为全面地比较多个协议的传输性能的工作还相对较少.下面具体介绍这方面的一些典型的已有研究工作.

Lopez-Pacheco 等人[39]采用 NS2 模拟实验方法对 TCP NewReno[40],HSTCP 和 XCP 这 3 种 TCP 改进协议在不同带宽环境下的性能进行了比较评价.他们建立了两种带宽变化模型,带宽变化范围从 30Mbps~200Mbps.一种是基于正弦(sine-based)变化,另一种是基于阶跃(step-based)变化.主要比较的性能指标是吞吐率.他们采用的带宽变化模型主要是模拟低速的分组共享网络,这种带宽变化范围对高速网络来说意义不大,而且他们建立的两种带宽变化模型对于实际网络意义也不大.

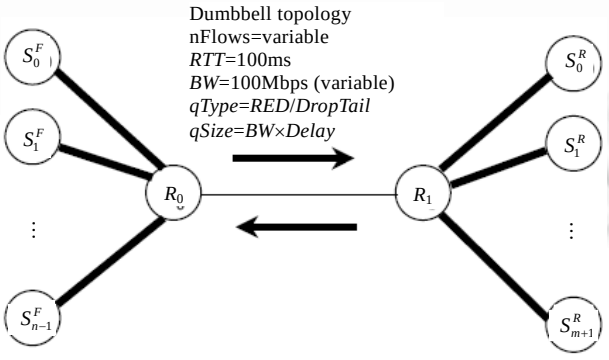

Fig.2 Simple dumbbell topology for simulation[41]

图 2 单哑铃模拟实验网络拓扑[41]

Chuvpilo 等人[41]采用 NS2 模拟实验方法对 High-Speed TCP 和 XCP 两个协议的部署问题进行了性能评价,讨论了如何在当前的 Internet 上安全地部署这两个协议、逐步部署路线以及缓冲区大小对部署的影响等等.模拟实验中建立了广泛使用的单拥塞链路哑铃模型,如图 2 所示,瓶颈带宽为 100Mbps,*RTT* 为 100ms.

张福杰等人[42]同样采用 NS2 模拟实验方法对 TCP Reno[43],High-Speed TCP 和 XCP 这 3 种协议对大带宽延迟网络环境中的性能进行了比较.模拟实验中,作者同样建立了单拥塞链路的哑铃模型,瓶颈带宽为 100Mbps, *RTT* 为 50ms.比较了 3 种协议在瓶颈链路利用率、公平性、TCP-Friendly 等指标上的性能.

上述研究主要是对 HSTCP 和 XCP 协议与标准 TCP 协议进行比较来进行性能评价.所比较的改进协议很少,只有 High-Speed TCP 和 XCP 两个协议.实验中所建立的网络模型也不是 FLDnet,其带宽和 *RTT* 时延都比较小,不能反映这些协议在 FLDnet 中的性能.

Bullot 等人[44]在实际的高速网络中对几种 TCP 改进协议性能进行了测量,包括 HSTCP,FAST,STCP,HSTCP-LP[45],HTCP 和 BIC TCP.作者对吞吐率、*RTT* 影响、公平性、稳定性等性能指标进行了实验评价.实验结果显示,大多数 TCP 改进协议在高速网络中比标准 TCP 协议的性能有所提高,但在公平性等方面也存在一些问题.实验在 3 条链路上进行,瓶颈带宽均为 622Mbps(OC12),*RTT* 时延分别为 10ms,70ms 和 170ms.但是,这些实验没有考虑各种背景流量所产生的影响.由于背景流量总在变化,因此对每次传输所产生的影响也不一样,难以公平比较各种协议的行为.

Li 等人[46]采用 Dummynet 网络仿真器等设备建立了网络实验床,采用仿真实验方法对 STCP,HS-TCP,BIC-TCP,FAST TCP 和 H-TCP 协议进行了性能比较评价.实验所用的网络实验床拓扑如图 3 所示,实验在终端中采用了 Web100[47](对 Linux 内核作了扩展)中实现的各种 TCP 改进协议.该文主要是对公平性方面进行性能评价.另外,评价了后向兼容性、效率、收敛时间等方面的性能.该文实验所设置的瓶颈带宽范围为 10Mbps~250Mbps,这个带宽范围太小,最高才 250Mbps,远远小于目前的实际高速网络带宽.Ha 等人[48]同样采用 Dummynet 网络仿真器等设备建立了网络实验床,采用仿真实验方法对 BIC-TCP,CUBIC,STCP,HSTCP, FAST TCP 和 HTCP 协议进行了性能比较评价.他们的工作主要是评估网络背景流量对各种 TCP 改进协议性能所产生的影响.

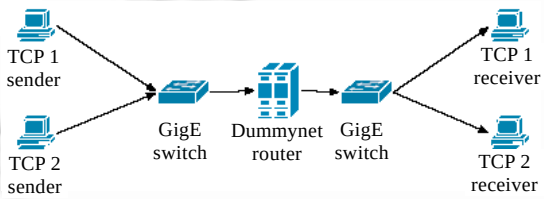

Fig.3 Testbed for performance evaluation of transport protocols[46]

图 3 传输协议性能评价实验床[46]

杨征等人[49]采用 NS2 模拟实验方法 BIC-TCP,STCP,HS-TCP,FAST TCP 和 H-TCP 协议进行了性能比较评价,他们主要考察了队列大小和传播时延对性能的影响.文献[50]同样采用了 NS2 模拟实验方法,对目前主要的 TCP 改进协议进行了比较,侧重于比较不同带宽、时延及丢包率等条件下各种改进 TCP 协议的性能.

现有的对 TCP 改进协议的性能评价研究,评价的协议不够充足,也没有充分考虑各种不同的网络条件,因此评价不够全面.而且大都针对分组共享的路由网络,强调公平性、友好性等指标,但却没有强调吞吐率指标.

3.2 基于UDP的改进协议性能评价

由于大多数改进协议都是对 TCP 协议的改进,相对而言,基于 UDP 的改进协议还很少,并且出现得较晚.因此,针对基于 UDP 的改进协议的性能评价研究到目前为止还相对较少.

Kumazoe 等人[51]在实际的 FLDnet 中,对 HSTCP,STCP 和基于 UDP 的改进协议 SABUL 进行了性能比较评价.实验在日本的一个开放实验床 JGN 网络上进行.该实验床是一个 ATM 网络,实验采用了两条 UBR PVC 链路.实验中,实际瓶颈链路带宽为 305Mbps,两条 PVC 链路 *RTT* 延迟分别为 100ms 和 460ms.实验结果显示,3 个改进协议的吞吐率都比标准 TCP 协议要高,其性能提高顺序为 SABUL>STCP>HSTCP,但丢包率也是同样的顺序.他们的实验结果也显示出,TCP 协议采用的 ACK 速率控制和差错控制机制可能不适合 FLDnet,因为接收 ACK 分组的时间依赖于距离(*RTT*)和接收方的 TCP 实现.Kumazoe 等人建议将差错控制机制与速率控制机制分离,以使速率控制更平滑和稳定.

以上实验是在 ATM 网络中进行的,不同于本文讨论的 FLDnet,其结果与 IP 分组交换网络或电路交换光网

络中的结果是不一样的.

Huang 等人[52]在 TeraGrid 等实验网络环境中对 PSockets(parallel socket)[53],RBUDP,UDT 和 TCP 协议的性能进行了比较评价.实验在 3 段链路上进行了传输测试,带宽为 1Gbps,*RTT* 分别为 0.2ms,0.38ms 和 61ms.其中,在 *RTT* 为 61ms 的链路上,Iperf 测得的最大可用带宽为 578Mbps.Huang 等人主要对吞吐率指标进行性能评价,实验结果显示,PSockets,RBUDP 和 UDT 协议的吞吐率远高于 TCP 协议.在延时较小时,其吞吐率接近于可用带宽,而 TCP 只有 4.57Mbps 吞吐率.

Anglano 等人[54]在 PlanetLab[55]实验床上对 bbFTP[56],RBUDP,FOBS 和 UDT 协议的性能进行了比较评价.实验链路带宽为 100Mbps,*RTT* 最小为 23ms,最高为 300ms.主要评价的指标是吞吐率.实验结果显示,在低带宽时延积网络中,TCP 改进协议和基于 UDP 的改进协议性能相似;而在高 BDP 网络中,基于 UDP 的改进协议性能明显高于 TCP 改进协议.

Wu 等人[57]采用 Dummynet 网络仿真器等设备建立了网络实验床进行仿真实验,以及在 TeraGrid 实验网络环境中进行实验,对 RBUDP,SABUL/UDT 和 GTP 这 3 个基于 UDP 的改进协议进行了性能比较.Wu 等人在多种通信模式(包括单个流、多个并行流、汇聚流等)下对吞吐率、丢包率、公平性以及协议开销等性能指标进行了实验评价.实验对 3 种基于 UDP 的改进协议流对 TCP 协议流的影响进行了测试,结果见表 1 和表 2.

Table 1 Influence of UDP-based protocols on TCP flow (LAN)[57]

表 1 基于 UDP 的改进协议流对 TCP 流的影响(LAN)[57]

	Rate-Based and TCP (Mbps)		Single TCP throughput (Mbps)	Influence ratio (%)
	Rate-Based	TCP		
RBUDP	467	450	912	49.3
UDT	552	380	912	41.6
GTP	612	328	912	35.9

Table 2 Influence of UDP-based protocols on TCP flow (WAN)[57]

表 2 基于 UDP 的改进协议流对 TCP 流的影响(WAN)[57]

	Rate-Based and TCP (Mbps)		Single TCP throughput (Mbps)	Influence ratio (%)
	Rate-Based	TCP		
RBUDP	771	2.1	24.3	8.6
UDT	751	23.6	24.3	97.2
GTP	760	9.7	24.3	40.0

表 1 显示的是单个基于 UDP 的协议流和单个 TCP 流在 LAN 中的点对点 1Gbps 链路上的吞吐率.从表中可以看出,基于 UDP 的协议可与 TCP 协议很好地共享带宽资源.表 2 显示的是单个基于 UDP 的协议流和单个 TCP 流在有 Dummynet 仿真的链路上点对点 800Mbps 链路(*RTT*=30ms)上的吞吐率.从表中可以看出,当存在 RBUDP 和 GTP 协议流时,TCP 流吞吐率不能达到 TCP 流单独运行时的吞吐率,而当 UDT 协议流存在时,TCP 流的吞吐率接近 TCP 流单独运行时的吞吐率.此实验结果说明,RBUDP 和 GTP 协议比较“激进(aggressive)”,而 UDT 比较“温和(gentle)”.

上述实验所做的性能评价工作得到了许多有价值的结论,但是这些实验所采用的仿真或实际网络并不是典型的 FLDnet,主要表现在带宽不够高(有的只有 100Mbps)、*RTT* 延迟不够大(大多数只有几十毫秒).这些实验中都没有考虑到链路丢包率对传输协议性能的影响.而且,采用实际网络测试的方法受到背景流量的影响很大,其实验结果准确性难以保证.

还有一些研究人员侧重于公平性、稳定性等方面的性能评价.比如,Gupta 等人[58]在实际网络中对各种 TCP 改进协议与 UDTv2 的公平性进行了实验评价,Wu 等人[59]采用 NS2 模拟实验方法对两个基于终端节点性能进行速率调节的协议 GTP 和 endpointXCP[59]与两个 TCP 改进协议 New Reno TCP 和 HSTCP 协议,在高速网络中的收敛性、效率、稳定性、公平性等性能指标上进行了比较评价.

文献[60]对基于 UDP 的改进协议的性能进行比较评价.在 FLDnet Testbed 中,采用仿真实验的方法比较了

目前主要的 3 种基于 UDP 的改进协议,即 RBUDP,UDT 和 Tsunami 协议,侧重于对 FLDnet 环境中协议的吞吐率以及不同时延及丢包率等条件下各种基于 UDP 的改进协议及 TCP 协议的性能进行比较评价.

4 总结与展望

本文综合国内外的相关研究,从传输协议的改进和性能评价两个方面系统地阐述了当前 FLDnet 传输协议的研究进展.通过对现有研究工作的分析,我们认为 FLDnet 传输协议在以下几个方面还有待进一步加以研究:

(1) 传输协议公平性等性能评价

由于目前 e-Science 科研应用对网络性能最紧迫的需求是传输速率,所以许多改进协议,尤其是基于 UDP 的改进协议主要是为了提高传输速率而设计的,很少在公平性等方面加以考虑.在现有的许多性能评价中,主要考虑的是吞吐率指标,对公平性、友好性等指标方面的评价还比较欠缺.然而,这并不意味着这些指标就是可以忽略的.

(2) 适用于 Lightpath 的传输协议

到目前为止,大部分改进的传输协议是针对分组交换的路由共享网络所作的改进,针对 Lightpath 上的传输协议的研究相对很少.最近,也有研究者提出了一种试图适用于 Lightpath 的 TCP 改进协议——C-TCP(circuit TCP)协议[61].C-TCP 完全取消了 TCP 的拥塞控制,以固定的网络容量(network capacity,简称 ncap)值代替 TCP 的拥塞窗口值 cwnd.实验结果表明,C-TCP 可以维持比较稳定的吞吐率和较高的带宽利用率.但是 C-TCP 存在一个严重问题,就是必须设置 ncap 值.由于取消了 TCP 的慢启动和拥塞控制机制,也就取消了 TCP 的可用带宽自适应功能, C-TCP 必须在使用前采用 Iperf 等测量工具测量可用带宽.如果连接改变,ncap 值又必须重新加以设置.C-TCP 的这个问题使其难以广泛应用.所以在设计适用于 Lightpath 的传输协议方面,还有很大的研究空间.

(3) 终端性能自适应传输协议

文献[33]中提出的 CDRA 拥塞控制算法能够针对终端的拥塞状况进行检测并做出拥塞反应,是一种针对终端性能的拥塞控制算法.采用 CDRA 算法的 RTsunami 协议就是一个终端性能自适应的传输协议.另外,RAPID[62]和 PA-UDP[63]两个协议也是根据终端性能调整发送速率的协议,但是尚不成熟,传输速率不高.由于光网络和高速以太网技术的飞速发展,网络带宽越来越高,终端系统的性能逐渐成为传输速率的瓶颈.因此,迫切需要在终端性能自适应的传输协议方面进行更多的研究.

(4) 应用层传输协议

目前,e-Science 科研应用中使用的一些应用层传输协议,比如在高能物理网格传输中使用 GridFTP 传输协议,在超级计算等应用中,SSH 是一个广泛使用的传输协议,也需要进行研究和改进.比如,Rapier 等人对 SSH 传输协议进行了改进,提出了 HPN-SSH(high performance networking SSH)协议[64].影响 FLDnet 传输性能的不只是传输层传输协议,应用层传输协议同样需要进行研究和改进.

(5) 下一代互联网传输层协议

由于到目前为止,海量数据传输应用还相对较少,而且还采用了像 Lightpath 这种专用的链路来传输.因此,针对海量数据传输设计的各种传输协议大都将这种应用视为特殊应用,强调追求传输速率,对协议的公平性、友好性等问题考虑不足.由于对 TCP 协议的各种改进方案传输速率仍然难以达到海量数据传输的需求,而基于 UDP 的各种传输协议在 TCP 和 UDP 等传输层协议之上,在应用中需要对应用程序进行修改,这极大地限制了这些传输协议的应用.下一代互联网正朝着高速的方向发展,各种普通的大数据量传输应用(比如 HDTV)将越来越普及,因此,设计满足普及的大数据量传输应用、适合于各种底层网络环境中(不只是 FLDnet,还有高速无线网络、传感器网络等)的下一代互联网传输层协议是一个非常重要的研究方向.

(6) 拥塞控制机制

拥塞控制是传输协议需要考虑的一个重要方面.无论技术如何发展,互联网的各种资源(包括带宽、节点处理能力、缓存空间等等)总是有限的,而应用流量的增长却是无限的,因此,拥塞控制的问题是一个可持续的研究

课题.

致谢 感谢审稿专家和编辑老师给本文提出的宝贵意见.

References:

[1] Ren YM, Qin G, Tang HN, Li J, Qian HL. Performance analysis of transport protocol over fast long distance optical network. Chinese Journal of Computers, 2008,31(10):1679−1686 (in Chinese with English abstract).

[2] Stewart R, Xie Q, Morneault K, Sharp C, Schwarzbauer H, Taylor T, Kytina I, Kalla M, Zhang L, Paxson V. Stream control transmission protocol. RFC 2960, Internet Engineering Task Force, 2000.

[3] Schulzrinne H, Casner S, Frederick R, Jacobson V. RTP: A transport protocol for real-time applications. RFC 3550, Internet Engineering Task Force, 2003.

[4] Schulzrinne H, Rao A, Lanphier R. Real time streaming protocol (RTSP). RFC 2326, Internet Engineering Task Force, 1998.

[5] Kohler E, Handley M, Floyd S. Designing DCCP: Congestion control without reliability. ACM SIGCOMM Computer Communication Review, 2006,36(4):27−38.

[6] Floyd S. High speed TCP for large congestion windows. IETF RFC 3649, 2003.

[7] Kelly T. Scalable TCP: Improving performance in high-speed wide area networks. Computer Communication Review, 2003,33(2): 83−91. [doi 10.1145/956981.956989]

[8] Xu L, Harfoush K, Rhee I. Binary increase congestion control for fast long-distance networks. In: Proc. of the INFOCOM. 2004. 2514−2524. http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?arnumber=135467

[9] Ha S, Rhee I, Xu LS. CUBIC: A new TCP-friendly high-speed TCP variant. ACM SIGOPS Operating System Review, 2008,42(5): 64−74. [doi: 10.1145/1400097.1400105]

[10] Floyd S, Allman M, Jain A, Sarolahti P. Quick-Start for TCP and IP. RFC 4782, 2007.

[11] Hacker TJ, Smith PM. Stochastic TCP: A statistical approach to congestion avoidance. In: Proc. of the PFLDnet2008. Manchester, 2008. http://www.cs.unc.edu/~aikat/diss/docs/papers-notes/papers/hacker_pfldnet_2008_stochasticTCP.pdf

[12] Shorten RN, Leith DJ. H-TCP: TCP for high-speed and long-distance networks. In: Proc. of the PFLDnet. Argonne, 2004. http://www.hamilton.ie/net/htcp3.pdf

[13] Jin C, Wei DX, Low SH. FAST TCP: Motivation, architecture, algorithms, performance. IEEE/ACM Trans. on Networking, 2006, 14(6):1246−1259. [doi: 10.1109/TNET.2006.886335]

[14] Bhandarkar S, Reddy ALN, Zhang Y, Loguinov D. Emulating AQM from end hosts. ACM SIGCOMM Computer Communication Review, 2007,37(4):349−360.

[15] Katabi D, Handley M, Rohrs C. Congestion control for high bandwidth-delay product networks. ACM SIGCOMM Computer Communication Review, 2002,32(4):89−102.

[16] Xia Y, Subramanian L, Stoica I, Kalyanaraman S. One more bit is enough. ACM SIGCOMM Computer Communication Review, 2005,35(4):37−48. [doi: 10.1145/1090191.1080098]

[17] Zhang Y, Leonard D, Loguinov D. JetMax: Scalable max-min congestion control for high-speed heterogeneous networks. In: Proc. of the IEEE INFOCOM. 2006. 1−13.

[18] Huang XM, Lin C, Ren FY. A novel high speed transport protocol based on explicit virtual load feedback. Computer Networks, 2007,51(7):1800−1814. [doi: 10.1016/j.comnet.2006.11.003]

[19] Huang XM, Lin C, Ren FY, Peter D, Wang YZ. Improving the convergence and stability of congestion control algorithm. In: Proc. of the 15th IEEE Int'l Conf. on Network Protocols (ICNP 2007). Beijing, 2007. 206−215.

[20] Krishnan R, Sterbenz J, Eddy W, Partridge C, Allman M. Explicit transport error notification (ETEN) for error-prone wireless and satellite networks. Computer Networks, 2004,46(3):343−362.

[21] Fox R. TCP big window and nak options. RFC 1106, 1989.

[22] Mathis M, Mahdavi J, Floyd S, Romanow A. TCP selective acknowledgement options. RFC 2018, 1996.

[23] Durst RC, Miller GJ, Travis EJ. TCP extension for space communications. Wireless Networks, 1997,3(5):389−403.

[24] SCPS transport protocol (SCPS-TP). 1997. http://www.scps.org/index.html

[25] Cheng RS, Lin HT. TCP selective negative acknowledgment over IEEE 802.11 wireless networks. In: Proc. of the Int'l Conf. on Networking and Services (ICNS 2006). Silicon Valley, 2006. 98.

[26] Cheng RS, Lin HT. Improving TCP performance with bandwidth estimation and selective negative acknowledgment in wireless networks. Journal of Communications and Networks, 2007,9(3):236−246.

[27] Sun FL, Li VOK, Liew SC. Design of SNACK mechanism for wireless TCP with new snoop. In: Proc. of the IEEE WCNC 2004. Atlanta, 2004. 1051−1056.

[28] Ren YM, Tang HN, Li J, Qian HL. Improving TCP performance with selective negative acknowledgement in hybrid optical packet network. In: Proc. of the Int'l Conf. on Computer and Network Technology (ICCNT 2009). Chennai: World Scientific Press, 2009. 122−128.

[29] He E, Leigh J, Yu O, DeFanti T. Reliable blast UDP: Predictable high performance bulk data transfer. In: Proc. of the IEEE Int'l Conf. on Cluster Computing. 2002. 317−324. [doi: 10.1109/CLUSTR.2002.1137760]

[30] Meiss MR. Tsunami: A high-speed rate-controlled protocol for file transfer. 2002. http://www.evl.uic.edu/eric/atp/TSUNAMI.pdf/

[31] Gu YH, Grossman RL. UDT: UDP-Based data transfer for high-speed wide area networks. Computer Networks, 2007,51(7): 1777−1799. [doi: 10.1016/j.comnet.2006.11.009]

[32] Larzon LA, Degermark M, Pink S, Jonsson LE, Fairhurst G. The lightweight user datagram protocol (UDP-Lite). RFC 3828, 2004.

[33] Ren YM, Tang HN, Li J, Qian HL. A novel congestion control algorithm for high performance bulk data transfer. In: Proc. of the IEEE Int'l Symp. on Network Computing and Applications (NCA 2009). Cambridge: IEEE Computer Society, 2009. 288−291.

[34] Gu YH, Grossman R. SABUL: A transport protocol for grid computing. Journal of Grid Computing, 2003,1(4):377−386. [doi: 10.1023/B:GRID.0000037553.18581.3b]

[35] Wu RX, Chien AA. GTP: Group transport protocol for lambda-grids. In: Proc. of the 4th IEEE/ACM Int'l Symp. on Cluster Computing and the Grid. Washington: IEEE Computer Society, 2004. 228−238.

[36] Dickens PM. FOBS: A lightweight communication protocol for grid computing. Lecture Notes in Computer Science 2790, 2003. 938−946.

[37] Vishwanath V, Leigh J, He E, Brown MD, Long L, Renambot L, Verlo A, Wang X, DeFanti TA. Wide area network experiments with Lambdastream over dedicate high bandwidth networks. In: Proc. of the IEEE INFOCOM 2006. Barcelona, 2006. http://www.startap.net/translight/papers/Vishwanath_IEEEInfocom2006.pdf

[38] Zheng X, Mudambi AP, Veeraraghavan M. FRTP: Fixed rate transport protocol—A modified version of SABUL for end-to-end circuits. In: Proc. of the Pathnets2004 on Broadnet2004. San Jose, 2004. http://www.ece.virginia.edu/mv/pubs/workshops/pathnets04/pathnets2004.pdf

[39] Lopez-Pacheco DM, Pham C. Performance comparison of TCP, HSTCP and XCP in high-speed, highly variable-bandwidth environments. In: Proc. of the IEEE 3rd Int'l Conf. on Network Protocols (ICNP 2004). Berlin, 2004. http://web.univ-pau.fr/~cpham/Paper/ICNP04.pdf

[40] Floyd S, Henderson T. The NewReno modification to TCP's fast recovery algorithm. RFC 2582, 1999.

[41] Chuvpilo G, Lee JW. A simulation based comparison between XCP and HighSpeed TCP. In: Computer Networks Final Project. Cambridge: Massachusetts Institute of Technology, 2002. http://www.how2setup.org/users/chuvpilo/papers/chuvpilo-2002-6.829-project.pdf

[42] Zhang FJ, Pan L, Li JH. Performance comparison of TCP, high-speed TCP and XCP in high BDP network. Computer Engineering, 2006,32(2):113−116 (in Chinese with English abstract).

[43] Jacobson V. Congestion avoidance and control. In: Proc. of the ACM SIGCOMM'88. 1988. 314−329.

[44] Bullot H, Cottrell RL, Hughes-Jones R. Evaluation of advanced TCP stacks on fast long-distance production networks. Journal of Grid Computing, 2003,1(4):345−359.

[45] Kuzmanovic A, Knightly EW. TCP-LP: A distributed algorithm for low priority data transfer. In: Proc. of the IEEE INFOCOM, Vol.3. San Francisco, 2003. 1691−1701.

[46] Li YT, Leith D, Shorten RN. Experimental evaluation of TCP protocols for high-speed networks. IEEE/ACM Trans. on Networking, 2007,15(5):1109−1122. [doi: 10.1109/TNET.2007.896240]

[47] Mathis M, Heffner J, Reddy R. Web100: Extended TCP instrumentation for research, education and diagnosis. ACM Computer Communications Review, 2003,33(3):69−79.

[48] Ha S, Le L, Rhee I, Xu LS. Impact of background traffic on performance of high-speed TCP variant protocols. Computer Networks, 2007,51(7):1748−1762. [doi: 10.1016/j.comnet.2006.11.005]

[49] Yang Z, Wu LD. Simulation-Based performance evaluation of TCP protocols for high-speed long distance networks. Computer Science, 2007,34(1):67−70 (in Chinese with English abstract).

[50] Ren YM, Tang HN, Li J, Qian HL. Performance comparison of TCP variants for high-speed network by NS2 simulation. Computer Engineering, 2009,35(2):6−9 (in Chinese with English abstract).

[51] Kumazoe K, Hori Y, Tsuru M, Qie YJ. Transport protocols for fast long distance networks: comparison of their performance in JGN. In: Proc. of the 2004 Int'l Symp. on Applications and the Internet Workshops (SAINTW 2004). Tokyo, 2004. 645. http://doi.ieeecomputersociety.org/10.1109/SAINTW.2004.1268701

[52] Huang R, Chien A. Benchmarking high bandwidth delay product protocols. Technical Report, San Diego: Concurrent Systems Architecture Group, University of California, 2003. http://www-csag.ucsd.edu/papers/MicroGrid-p.html

[53] Sivakumar H, Bailey S, Grossman RL. PSockets: The case for application-level network striping for data intensive applications using high speed wide area networks. In: Proc. of the SC 2000. Dallas, 2000. 38.

[54] Anglano C, Canonico M. A comparative evaluation of high-performance file transfer systems for data-intensive grid applications. In: Proc. of the 13th IEEE Int'l Workshops on Enabling Technologies: Infrastructure for Collaborative Enterprises (WET ICE 2004). 2004. 283−288. http://doi.ieeecomputersociety.org/10.1109/ENABL.2004.2

[55] Chun B, Culler D, Roscoe T, Bavier A, Peterson L, Wawrzoniak M, Bowman M. Planet-Lab: An overlay testbed for broad-coverage services. ACM SIGCOMM Computer Communications Review, 2003,33(3):3−12. http://www.planet-lab.org

[56] The bbFTP-large files transfer protocols. Web Site. 2005. http://doc.in2p3.fr/bbftp/

[57] Wu R, Chien A. Evaluation of rate-based transport protocols for Lambda-grids. In: Proc. of the 13th IEEE Int'l Symp. on High-Performance Distributed Computing (HPDC 2004). Honolulu, 2004. 87−96.

[58] Chen M, McIntosh R, Leers F. Characterization and evaluation of TCP and UDP-based transport on real networks. In: Proc. of the PFLDnet 2005. Lyon, 2005. http://www.slac.stanford.edu/cgi-wrap/getdoc/slac-pub-10996.pdf

[59] Wu R, Chien AA. Evaluation of end-node based rate allocation schemes for lambda networks. In: Proc. of the PFLDNet 2006. Nara, 2006. http://www-csag.ucsd.edu/papers/pfldnet2006.pdf

[60] Ren YM, Tang HN, Li J, Qian HL. Performance comparison of UDP-based protocols over fast long distance network. Information Technology Journal, 2009,8(4):600−604. [doi: 10.3923/itj.2009.600.604]

[61] Mudambi AP, Zheng X, Veeraraghavan M. A transport protocol for dedicated end-to-end circuits. In: Proc. of the IEEE Int'l Conf. on Communications (ICC 2006). 2006. 18−23.

[62] Banerjee A, Feng WC, Mukherjee B, Ghosal D. RAPID: An end-system aware protocol for intelligent data transfer over Lambda grids. In: Proc. of the 20th Int'l Parallel and Distributed Processing Symp. (IPDPS 2006). Rhodes Island, 2006.

[63] Eckart B, He XB, Wu QS. Performance adaptive UDP for high-speed bulk data transfer over dedicated links. In: Proc. of the IEEE Int'l Symp. on Parallel and Distributed Processing (IPDPS). Miami, 2008. 1−10.

[64] Rapier C, Bennett B. High speed bulk data transfer using the SSH protocol. In: Proc. of the 15th ACM Mardi Gras Conf. (MG 2008). Baton Rouge, 2008.

附中文参考文献:

[1] 任勇毛,秦刚,唐海娜,李俊,钱华林.高速长距离光网络传输协议性能分析.计算机学报,2008,31(10):1679−1686.

[42] 张福杰,潘理,李建华.大带宽时延积网络中 TCP,High-Speed TCP 及 XCP 性能比较.计算机工程,2006,32(2):113−116.

[49] 杨征,吴玲达.基于仿真的高速长距离网络中 TCP 协议性能评价.计算机科学,2007,34(1):67−70.

[50] 任勇毛,唐海娜,李俊,钱华林.高速网络 TCP 改进协议 NS2 仿真性能比较.计算机工程,2009,35(2):6−9.

任勇毛(1981－),男,湖南邵阳人,博士,助理研究员,主要研究领域为高速网络,传输协议.

李俊(1968－),男,博士,研究员,博士生导师,主要研究领域为下一代互联网,高速网络.

唐海娜(1977－),女,工程师,主要研究领域为网络管理,网络监测.

钱华林(1940－),男,研究员,博士生导师,主要研究领域为下一代网络体系结构.