doi:10.3969/j.issn.1000-436x.2014.04.010

TCP 改进协议在高速长距离网络中的性能研究

王国栋[1,2]，任勇毛[1]，李俊[1]
(1.中国科学院 计算机网络信息中心，北京 100190；2.中国科学院 研究生院，北京 100049)

摘 要：随着高速网络的发展，一系列适合于高速长距离网络的 TCP 传输协议被相继提出。分类总结了近年来提出的各种改进的 TCP 传输协议。在此基础之上，分别基于仿真工具和真实网络对目前存在的传输协议性能进行了评价，并做出了详细的评价分析。在归纳和总结目前 TCP 传输协议研究中存在的问题的同时，提出了下一步研究的方向。

关键词：高速长距离网络；传输协议；拥塞控制；性能评价
中图分类号：TP393 **文献标识码**：A **文章编号**：1000-436X(2014)04-0081-10

Performance evaluation of TCP congestion control algorithms in fast long distance network

WANG Guo-dong[1,2], REN Yong-mao[1], LI Jun[1]
(1.Computer Network Information Center, Chinese Academy of Sciences, Beijing 100190, China;
2.Graduate School, Chinese Academy of Sciences, Beijing 100049, China)

Abstract: With the development of high-speed networks, a series of high-speed TCP transport protocols have been proposed. Firstly, the improved TCP variants was classified according to their characteristics, and each protocol was given a brief introduction. Then, detailed evaluation and analysis were performed through the simulation and in real network separately. Finally, the shortcomings of these protocols and their future research directions were provided.

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

1 引言

随着信息技术的快速发展和科研活动信息化水平的日益提高，互联网逐渐成为科研工作者进行科学研究的工具。例如在高能物理方面，欧洲粒子物理研究中心(CERN)的大型粒子对撞机每年产生几十 PB 的数据[1]，这些数据需要通过高速网络传送到全球各个研究中心进行处理和分析。在天文学领域，天文学家采用甚长基线干涉测量法（VLBI, very long baseline interferometry）由分布于全球的仪器采集数据并通过高速网络进行汇聚，从而获得详细的天文图像[2]。在生物信息领域，全球各大测序中心每年都产生大量的基因序列数据，这些数据通过高速网络供全球用户获取[3]。这些科研应用的发展，也对高速网络的传输性能提出了越来越高的要求。

在高速网络传输方面，高速网络基础设施已经出现，但是研究人员却无法有效利用网络带宽，因为传统的传输协议严重影响着高速网络的传输性能。针对传输协议在高速网络中的各种缺陷，研究人员提出了许多基于标准 TCP 的改进协议。这些改进协议针对现有协议的某些缺点做出了改进，使得部分性能有所提高，但是这些改进协议在高速长距离网络环境中的性能如何，能否满足 e-Science 等科

收稿日期：2012-08-31；**修回日期**：2013-04-20
基金项目：国家发改委基金资助项目(GC-HG100452)；科技部国际合作基金资助项目(2012FDA11090)；中国科学院计算机网络信息中心主任基金资助项目(CNIC_ZR_201204)
Foundation Items: The National Development and Reform Project (GC-HG100452); International Cooperation Project of the Ministry of Science and Technology (2012FDA11090); Director Fund Project of CNIC, CAS (CNIC_ZR_201204)

研应用的需要，针对这些问题，迫切需要对现有的改进协议做出评价。

对 TCP 传输协议的评价，研究人员已经做了一部分工作，其中，文献[4]对 CUBIC 协议进行了评价并与标准 TCP 的性能进行了比较。文献[5]通过 NS2 仿真软件对 TCP 传输协议进行了简单的分析，但是其仅仅评价了 TCP 的效率和稳定性 2 个方面。文献[6]对 Illinois 和 CTCP 的性能进行了评价。文献[7]在无线局域网络环境下对 Vegas、Fast TCP、CTCP 和 ARENO 的性能进行了评价。文献[9]和文献[10]在实验室环境下对快速启动 TCP 协议进行了评价。文献[11]评价了 3 种不同 TCP 协议在进行 H.264 SVC（H.264 可分级编码）流传输时的性能。综上所述，目前对 TCP 传输协议的性能评价主要存在以下不足。首先是评价的范围不够广，没有对目前存在的协议进行全面系统的分析，而往往是针对某一个或几个协议进行比较。其次是评价的手段不够全面，由于真实的高速长距离网络资源的稀缺，大部分评价工作采用仿真软件进行仿真或者在实验室环境下进行模拟，而对 TCP 传输协议在真实网络环境中的性能的评价还不多见。针对以上问题，本文首先利用仿真软件对目前典型的 TCP 传输协议进行了全面的、系统的评价分析。在此基础之上，充分利用中国科技网的网络资源优势在实际网络中对目前典型的 TCP 传输协议的性能进行了研究。

2 TCP 改进协议分类描述

标准 TCP 协议在高速网络中存在的主要问题是其加性增、乘性减(AIMD，additive increase multiplicative decrease)的窗口调整策略。Floyd[12]指出，在高速网络中，如果要充分利用带宽，这个约束条件会导致不现实的分组丢失率(2×10^{-10})。拥塞减半后加性增加拥塞窗口需要大量时间(1.16 h)才能恢复到 10 Gbit/s 的吞吐率。因此对标准 TCP 协议的改进，主要集中在拥塞控制窗口(congestion control window)的管理机制上。根据拥塞检测机制的不同，改进协议可以分为以下几类：基于分组丢失反馈的改进协议(LCA，loss-based congestion algorithm)，基于时延反馈的改进协议(DCA，delay-based congestion algorithm)，基于 LCA 和 DCA 的混合反馈改进协议(CCA，compound congestion algorithm)，基于可用带宽测量的改进协议及基于路由器显式反馈(explicit notification)的改进协议。表 1 对各类协议进行了分类总结。

2.1 基于分组丢失反馈的改进协议

2.1.1 HSTCP

HSTCP[12]采用高/低速模式切换的工作方式，通过 $\alpha(\omega)$和 $\beta(\omega)$调整窗口变化速率。当拥塞窗口较小时，HSTCP 采用类似于传统 TCP 协议的窗口增长和分组丢失递减方式，当拥塞窗口较大时，采用更加积极的窗口增长和更加缓和的窗口减少算法。

2.1.2 STCP（scalable TCP）

STCP[13]采用 MIMD(multiplicative increase multiplicative decrease)代替 AIMD 来调整拥塞窗口。对于每个 RTT，在拥塞避免阶段，$cwnd = cwnd + \alpha\times cwnd(\alpha=0.01)$，在快速恢复阶段，$cwnd = cwnd - \beta\times cwnd(\beta=0.125)$。

2.1.3 BIC

BIC[14]的主要特点是它独特的窗口增长函数，它由二进制搜索增长和线性增长两部分组成。二进制搜索类似于经典的折半查找算法，首先假设拥塞窗口的最大值 max 和最小值 min，并取其中值 TW(target window = (max +min)/2)作为目标窗口增长 cwnd。如果没有发生分组丢失，则取当前窗口作为 min 继续进行二进制搜索增长 cwnd；如果发生分组丢失，则更改当前窗口为 max，并重新进行二进制搜索增长 cwnd。如果当前窗口距离 TW 过大，则采用线性增长辅助增加 cwnd。

2.1.4 CUBIC

CUBIC[15]是 BIC 的一个改进版本，试图在保留 BIC 优点的基础上，简化窗口控制并增强它的 TCP 友好性和 RTT 公平性。

表 1 TCP 改进协议分类描述

TCP 分类		协议
隐式反馈	基于分组丢失反馈的改进协议	HSTCP、STCP、BIC、CUBIC、HTCP、Hybla、Libra
	基于时延反馈的改进协议	Hybrid Slow Start TCP、Vegas、Fast TCP
	基于分组丢失和时延混合反馈的改进协议	CTCP、Illinois、Africa、YeAH
	基于可用带宽策略的改进协议	Westwood、ARENO、Fusion
显式反馈	基于显式反馈的改进协议	XCP、VCP、JetMax、EVLF-TCP、CLTCP

2.1.5 HTCP

HTCP[16]采用上次拥塞事件以来逝去的时间 Δ 来检测网络拥塞程度。AIMD 的调整因子为 $\alpha(\Delta)$，$\alpha(\Delta)$随着 Δ 的变化进行动态调整，从而调整 AIMD 的窗口变化速率。

2.1.6 Libra

Libra[17]是在 New Reno[18]的基础上改进而来的，其主要目的是为了提高 New Reno 的 RTT 公平性和可扩展性。

2.1.7 Hybla

与 Libra 类似，Hybla[19]的主要目的也是为了解决 New Reno 的 RTT 公平性问题。Hybla 通过引入 RTT 的参考值 $RT\mathrm{T}_{\mathrm{ref}}$(默认为 25 ms)，并取 $\rho=RTT/RTT_{\mathrm{ref}}$来调节不同 RTT 的拥塞窗口增长速率。

2.2 基于时延反馈的改进协议

2.2.1 Fast TCP

Fast TCP[8, 20]以队列延时作为反馈因子，利用发送端检测到的 ACK 时延变化来调整拥塞窗口。其拥塞窗口调整算法如下：

$$w(k+1)=\frac{1}{2}\left(\frac{w(k-1)\times baseRTT}{RTT}+\alpha+w(k)\right)$$

其中，$baseRTT$ 表示所观测到的最小 RTT，α 表示一个非负的修正因子。

2.2.2 Vegas

Vegas[21]是基于时延反馈协议的代表。它通过观测 TCP 连接中的 RTT 时延变化来调节拥塞窗口。如果发现 RTT 变大，则认为网络发生拥塞，相应的减小拥塞窗口。如果发现 RTT 变小，则认为拥塞已经解除，并增加拥塞窗口。如果 RTT 保持不变，则不改变拥塞窗口的大小。

2.2.3 Hybrid Slow Start TCP

Hybrid Slow Start TCP[22]利用了网络测量技术中的 packet-pair[23]测量技术和 packet-train[24]测量技术，通过“ACK train length”和“Delay increase”来决定 TCP 何时从慢启动阶段转换到拥塞避免阶段。

2.3 基于分组丢失和时延混合反馈的改进协议

2.3.1 CTCP

CTCP[25](compound TCP)将基于分组丢失的 CAA(congestion avoidance algorithm)和基于时延的 CAA 相结合，在保证了较高的带宽利用率的情况下，又获得了较好的公平性。基于时延的 CAA 是通过引入 dwnd(delay window)来实现的，协议中的发送窗口为 $win=min(cwnd+dwnd, awnd)$，其中，$awnd$ 是来自接收端的广播窗口。相应的在拥塞避免阶段，每收到一个 ACK，窗口大小调整为 $cwnd=cwnd+1/win$，而在慢启动阶段，CTCP 保留了 Reno TCP 的慢启动策略。

2.3.2 Africa

TCP Africa[26](adaptive and fair rapid increase congestion avoidance)将网络状态分为拥塞和无拥塞 2 个状态，并且利用 Vegas 的 RTT 延时来判断这 2 种状态。当网络无拥塞时，协议进入类似于 HSTCP 的快速增长模式；在网络逼近拥塞时，TCP Africa 进入类似于传统 TCP 的慢速增长模式。

2.3.3 Illinois

Illinois[27]同样考虑到 LCA 和 DCA 的局限性，提出了采用两者相结合的算法。与 CTCP 和 Africa 类似，都是将网络状态分为拥塞和无拥塞 2 种状态，但是在这 2 种状态下对拥塞窗口的调整策略不同。对于 Illinois 而言，在拥塞避免阶段，每个 RTT: $cwnd=cwnd+\alpha$，而当检测到分组丢失时，$cwnd=cwnd-\beta\cdot cwnd$。

2.3.4 YeAH

YeAH[28](yet another high-speed TCP)将网络划分为“Fast”和“slow”2 个状态。在“Fast”状态下，拥塞窗口使用更加迅速的方式增加(STCP)，在“slow”状态下，拥塞窗口使用较温和的方式增加(Reno TCP)，以避免网络拥塞的产生。在此基础上，当网络拥塞超过阈值时，将路由器缓存中的分组取出，来进一步缓解拥塞。

2.4 基于可用带宽测量的改进协议

2.4.1 Westwood

Westwood[29]和 westwood+[30]是典型的基于网络带宽测量的 TCP 协议。Westwood 通过计算最近过去(recent past)的带宽来调整拥塞窗口的变化。网络带宽是通过记录一段时间之内(以 ACK 为标准)所发送的数据来得到的。Westwood+改进了网络带宽的测量机制，分别用“被确认了”的数据来代替发送的数据，用 RTT 来代替 ACK 来获得传输时间，这一改进大大提高了网络带宽测量的准确性。

2.4.2 ARENO

ARENO[31](adaptive RENO)是一个基于带宽测量的 TCP 改进协议，旨在改进 TCP 的效率和 TCP 的友好性。它具有 2 个窗口增加机制，W_{base} 和 W_{probe}。在 W_{base} 阶段，每个 RTT 增加一个 CWND，在 W_{probe} 阶段，每个 RTT 增加的 CWND 根据所测量的网络带宽动态调整。ARENO 的带宽测量机制类似于 Westwood。

2.4.3 Fusion

Fusion[32]与 ARENO 类似，结合 westwood 的带宽测量和 vegas 的网络缓存预测机制。Fusion 定义了 3 个线性增长函数，根据不同的队列延时，动态切换这 3 个增长函数。另外，当网络分组丢失时，根据 RTT 的不同，将拥塞窗口减小到相应的程度。

2.5 基于显式拥塞通告的改进协议

基于显式拥塞通告的改进协议的代表主要有 XCP[33]和 VCP[34]。XCP 为数据分组增加了拥塞报头，由发送端写入当前的窗口大小和 RTT 估计值，为路由器计算可用带宽提供信息。VCP 采用 IP 报头的冗余位作为负载因子向发送端传递网络拥塞状况。基于显示反馈的 TCP 协议还有 JetMax[35]、EVLF-TCP[36]、CLTCP[37]，这类协议的最大问题就是可扩展性差。由于这类协议需要路由器支持，实际部署会比较困难，因此本文不再详述和评价此类协议。

3 TCP 传输协议评价标准及评价方法

对于传输协议的性能评价是个很大的研究课题，一直是比较热门的研究领域。进行性能评价，需要考虑 2 个因素：一是评价标准，二是评价方法。

3.1 评价标准

评价标准是影响评价结果的一个重要因素。对于传输协议的评价标准，目前并无定论[38, 39]。但是对于 e-Science 科研应用而言衡量传输协议优劣的一个重要指标是传输协议的效率;对于网络自身的性能而言，RTT 公平性和 TCP 友好性也是必要的考虑因素。综合以上考虑，本文着重从协议的吞吐率、RTT 公平性和协议之间的友好性进行评价。

3.2 评价方法

评价方法也是影响评价结果的重要因素。一般而言，对于传输协议性能评价的方法主要有以下几种：一种是利用理论模型分析法[40]。一种是采用模拟的实验方法，NS2 是目前学术界广泛使用的一种网络模拟平台，其实验结果受到专业领域的普遍认可。另一种是通过搭建模拟网络环境进行实验[6, 41]。还有一种是在实际网络中进行测试[42]。

理论模型分析方法需要建立传输模型，这种方法比较复杂，而且难以保证模型的有效性；采用软件仿真虽然比较容易控制网络的各种参数，但是其性能依赖所使用的仿真环境；采用实验床的方法也仅仅模拟了网络的环境，与实际网络还有差距。在真实网络中进行测试最大的问题是网络资源有限，另外背景流量也会对实验结果产生较大影响。综合以上分析，为了全面、客观地反应 TCP 改进协议的性能，本文分别采用仿真和在真实网络中对 TCP 传输协议进行评价。

4 基于 NS2 的 TCP 传输协议评价

本节采用 NS2 仿真工具对目前流行的 TCP 改进协议进行评价。网络拓扑为单瓶颈主干链路、多接入终端的哑铃型拓扑。根据队列大小理论[43]，队列大小为 100%BDP（bandwidth delay product），TCP 连接的应用使用 FTP，实验网络拓扑如图 1 所示。

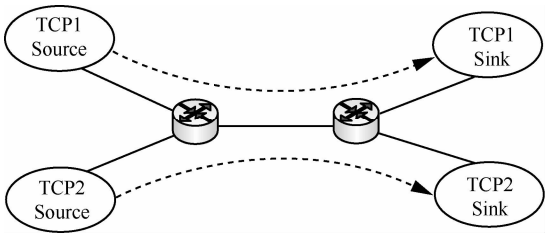


图 1 NS2 仿真实验拓扑

4.1 单流带宽利用率

单流带宽利用率是 TCP 传输协议传输效率的重要指标，在没有背景流量干扰的情况下，通过改变影响 TCP 性能的参数，考察单个 TCP 流所能获得的带宽利用率，能反应 TCP 在数据传输中的效率。由于链路带宽和 RTT 时延是影响 TCP 传输效率的 2 个主要因素，因此本小节主要从这两方面进行考察。

4.1.1 链路瓶颈带宽对带宽利用率的影响

本实验主要考察链路瓶颈带宽对 TCP 协议带宽利用率的影响。为了减小过大的链路时延对传输性能的影响，本实验选取 RTT 时延为 64 ms，瓶颈带宽分别选取为 10 Mbit/s (Ethernet)，100 Mbit/s (FE)，155 Mbit/s (OC-3 Wan)，622 Mbit/s (OC-12)，1 Gbit/s (GE)。实验时间为 1 200 s，通过带宽利用率来刻画单流 TCP 的传输效率。

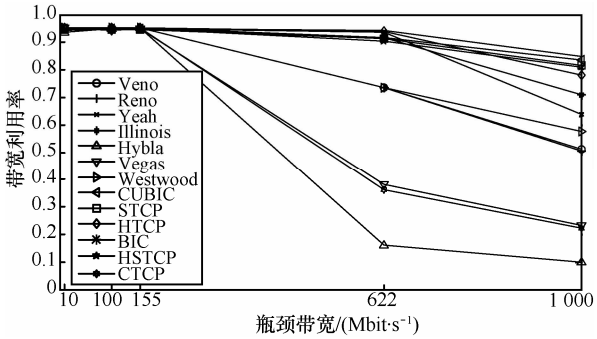


图 2 瓶颈带宽对带宽利用率的影响

从图 2 所示的实验结果可以看出，在低带宽(带宽小于 155 Mbit/s)的环境下，各类协议都具有非常好的带宽利用率(带宽利用率超过 0.9)。但是随着链路瓶颈带宽的增加，尤其在带宽大于 155 Mbit/s 之后，各种协议的带宽利用率都呈下降趋势。其中 Hybla 下降的最为明显，在 622 Mbit/s 的瓶颈带宽中，Hybla 仅仅得到了 0.15 的带宽利用率，而在 1 Gbit/s 的瓶颈带宽中，Hybla 的带宽利用率只有 0.1。这在一定程度上反映了 Hybla 通过降低 RTT 较小的流的拥塞窗口增长速率的方法，影响了其在较小网络时延环境中的传输效率[44]。其次是 CTCP 和 Vegas，在 622 Mbit/s 的瓶颈带宽中，它们的带宽利用率均低于 0.4，而在 1 Gbit/s 的带宽中，它们的带宽利用率仅仅为 0.2。专门为高速长距离网络设计的 BIC，HTCP 和 HSTCP 均获得了不错的带宽利用率，在 1 Gbit/s 的链路中，它们的带宽利用率均超过了 0.8。

4.1.2　RTT 对带宽利用率的影响

本实验选取瓶颈带宽为 622 Mbit/s，RTT 时延从 2 ms 到 512 ms 递增，实验时间为 1 200 s，通过吞吐率来刻画 RTT 对单流 TCP 传输效率的影响。

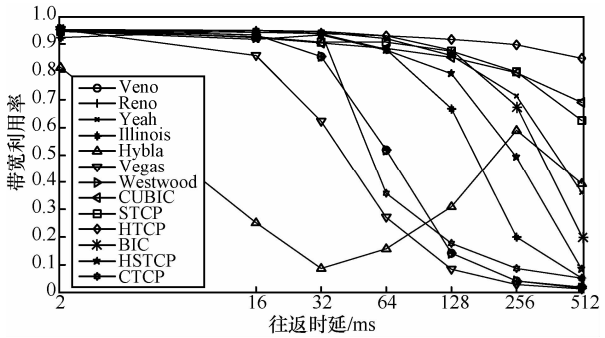


图 3　RTT 对带宽利用率的影响

从图 3 所示的结果可以看出，随着 RTT 的增加，除了 Hybla 之外，其余各类 TCP 协议的带宽利用率均呈下降趋势。Hybla 之所以出现带宽利用率先下降再上升的情况，是和其采用的算法分不开的。Hybla 引入了参数 RTT_{ref}(默认为 25 ms)，并取 $\rho=RTT/RTT_{ref}$ 来调节不同 RTT 环境下拥塞窗口的增长速率，其自身建立的模型虽然保证了 RTT 的公平性(见 4.2)，但是无法保证在所有 RTT 环境下都能达到很好的带宽利用率[44]。图 2 中 Hybla 的性能欠佳也反映了相同的问题。需要强调的是，随着 RTT 的增加，Vegas 的吞吐率下降明显，这表明在高延时环境中，依靠 RTT 的变化来判断网络拥塞的准确率下降，因此其性能受到了影响。HTCP 在 RTT 为 512 ms，瓶颈带宽为 622 Mbit/s 的网络环境下依然获得了不错的带宽利用率(0.85)。其次 CUBIC 和 STCP 也获得了不错的带宽利用率。

4.2　公平性

TCP 协议的公平性主要有 2 种；一种是协议内部的公平性(intra-protocol fairness)，即 2 个完全相同的 TCP 流在竞争通过瓶颈路径时，是否能公平的分享瓶颈带宽；另一种是 RTT 的公平性(RTT fairness)，即相同的 TCP 协议在不同的 RTT 时延下竞争通过瓶颈路径时的带宽分配情况。

4.2.1　Intra-protocol fairness

本实验选取路由器 buffer 为 100%BDP，2 个相同的 TCP 流竞争通过带宽为 622 Mbit/s 的瓶颈链路，分别考察在不同 RTT 环境下协议之间的公平性，实验时间为 1 200 s。本文采用 Jain[45,46]所提出的公平性指标来衡量 TCP 协议的公平性：

$$F=\frac{(\sum_{i=1}^{n}x_i)^2}{n(\sum_{i=1}^{n}x_i)^2}$$

其中，F 的值越接近于 1，表明协议公平性越好。

从图 4 所示的实验结果可以看出，随着 *RTT* 的增加，各协议的公平性总体上呈上升趋势，这是因为随着 *RTT* 的增加，处于竞争的 TCP 协议有充足的时间调整拥塞窗口，以占据竞争流所让出的带宽。

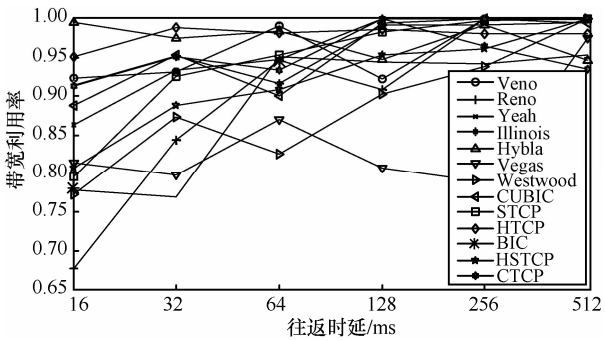


图 4　协议内部公平性

其中，HTCP 的公平性最佳，即使在 RTT 较小的情况下依然具有较好的公平性。而 Vegas 的协议内部公平性较差，其原因在于仅依靠链路中 RTT 的变化来判断网络的拥塞程度存在一定的误差。尤其是随着 RTT 的增加，Vegas 无法通过链路时延的变化准确估计网络的拥塞程度，这也影响了其公平性。而当 RTT 为 512 ms 时，Vegas 又表现出了一种“良好”的协议内部的公平性，其原因在于此时 2 个 Vegas

流的吞吐率均很小（如 4.1.2 节所示），其竞争瓶颈带宽的能力减弱所致。值得一提的是基于 Vegas 改进之后的 YeAH TCP 的公平性得到了很大提高。

4.2.2　RTT fairness

本实验中 2 个使用相同协议的 TCP 流在不同 RTT 的环境下竞争通过带宽为 622 Mbit/s 的瓶颈链路。其中，一个 TCP 流的 RTT 固定为 256 ms，另一个 TCP 流的 RTT 从 16 ms 到 512 ms 之间变化，分别考察相同协议在不同延时下的 RTT 公平性。本文采用 Chiu 所提出的公平性指标来刻画 RTT 的公平性

$$A=\frac{x_1-x_2}{x_1+x_2} \tag{1}$$

其中，x_1 为 RTT 变化的 TCP 流，x_2 为 RTT 固定为 256 ms 的 TCP 流，A 值越接近 0 表明 RTT 公平性越好。

由图 5 所示的实验结果可知，除了 Hybla 的 A 值始终接近 0 之外，其余各种协议的 RTT 公平性均不理想。Hybla 的设计目的就是为了提高 TCP 的 RTT 公平性，此实验进一步验证了 Hybla 在提高 TCP 的 RTT 公平性方面的有效性，但是从 4.1 节实验结果可知 Hybla 存在严重的效率问题。

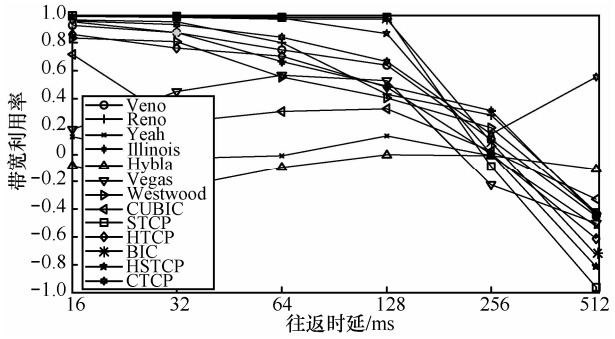


图 5　RTT 平性比较

随着 TCP 流 $x1$ 的 RTT 逐渐增加，各协议的 RTT 公平性也趋近于 0，当 2 个流的 RTT 相同时(256 ms)，大多数的 A 值都汇聚与 0，当 x_1 的 RTT 再进一步增加，A 值进一步减小并偏离 0 变为负值，这说明大多数 TCP 都存在 RTT 公平性问题，RTT 小的流在瓶颈带宽竞争中存在优势。由图 5 还可以看出，YeAH 也具有不错的 RTT 公平性，虽然和 Hybla 相比还有差距，但是 YeAH 的效率却比 Hybla 具有优势。

4.3　TCP 友好性

本实验中 2 个使用不同协议的 TCP 流在相同的 RTT 时延下竞争通过带宽为 622 Mbit/s 的瓶颈链路。其中，一个 TCP 固定为传统的 Reno，另一个 TCP 流为待测试的 TCP 协议。选取 RTT 从 16 ms 到 512 ms 之间变化，分别考察各协议在不同 RTT 时延下的友好性问题。本文采用式(1)刻画 TCP 的友好性。其中，x_1 为待测试协议的 TCP 流，x_2 为 Reno。A 值越趋于 0 表明其友好性越好，A 值为 0 表明所测试的协议与 Reno 具有相同的友好性。

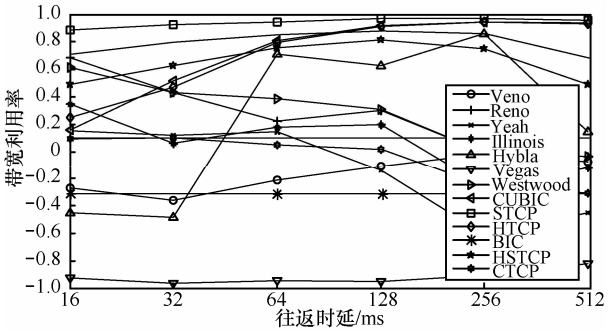


图 6　TCP 友好性比较

从图 6 所示的实验结果可知，随着 RTT 的增加，所测试协议的友好性逐渐提高。具体来讲，所测试的 TCP 协议可以分为 2 类：一类是 A 为负值的协议，这类协议比 Reno 具有更好的 TCP 友好性；另一类是 A 为正值的协议，这类协议的 TCP 友好性较差，比较典型的是 STCP，其使用 MIMD 的拥塞窗口调整策略，并且拥塞后减小 0.125 倍当前窗口的方式，虽然获得了不错的效率，但是其 TCP 友好性极差。Vegas 采用基于时延的反馈算法，具有较好的 TCP 友好性，但是其传输效率较差。TCP 的友好性和效率是一对矛盾体，在提高效率的同时很难保证协议的友好性，反之，在提高友好性的时候，传输效率也往往会受到影响。

5　基于真实网络的 TCP 传输协议评价

本节主要考察 TCP 协议在真实网络中的传输效率。为了更好的反应 TCP 传输协议在不同网络环境中的传输性能，本实验采用两段真实链路对 TCP 的性能进行评价。第一段链路是 GLORIAD（global ring network for advanced application development）从香港到芝加哥的国际链路，链路带宽为 1 Gbit/s，往返时延为 139 ms(±1 ms)；第二段链路是从北京到上海的国内链路，链路带宽为 1 Gbit/s，往返时延为 20 ms(±1 ms)。具体的链路拓扑示意如图 7 所示。

所采用的测试方法也分为 2 种：一种是通过网络测量工具 Iperf 来测试，以最大限度的降低不同应用程序对 TCP 传输性能的影响；考虑到在海量数据传输时，FTP 是一种常用的手段，因此第 2 种方法采用 FTP 来传输实际文件（大小为 2 Gbyte）来进行测试。FTP 服务器采用 CentOS 系统默认的 VSFTP，由于不同的 FTP 客户端对传输性能影响较大，为了真实反映 TCP 协议的传输性能，选用 LFTP 作为客户端。因为 LFTP 未进行任何的加密处理，其传输效率受到加密等外界因素的影响最小，因此能更好地反映 TCP 传输协议的性能。

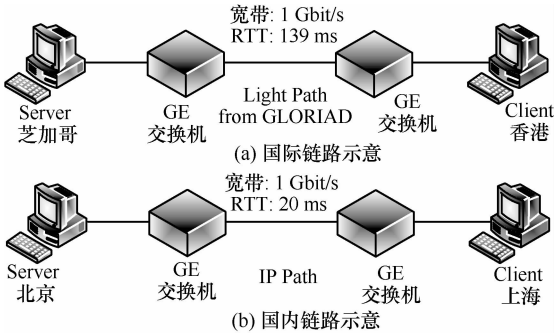


图 7　国际和国内链路示意

在实际网络中进行海量数据传输时，传输效率是用户最为关注的性能指标，因此本次的性能评价也主要以 TCP 的吞吐率作为考察指标。在真实网络中对 TCP 进行测试，TCP 的吞吐率不可避免地会受到背景流量的影响，同一协议的测试结果会有所变化，为了反映出 TCP 较为真实传输性能，以下测试均进行 5 次，去掉最大值和最小值，取其余 3 个有效值的平均值作为最终结果。

5.1　TCP 协议在国际链路上的性能

图 8 和图 9 分别是采用 Iperf 和 FTP 对 TCP 在国际链路上进行数据传输的测试结果。从总体上看，采用 Iperf 所测量的 TCP 性能稍好于使用 FTP 进行实际文件传输的性能。这主要是由于 FTP 自身的开销导致了 TCP 的吞吐的降低。但是考虑到目前在高速长距离网络中进行海量数据传输时广泛采用 FTP 来进行，因此，图 9 所示的使用 FTP 对 TCP 进行的性能测试更有实际意义。

由图 8 和图 9 可知，在网络时延为 139 ms(±1 ms)，链路可用带宽约为 900 Mbit/s（均通过 Iperf 的 UDP 传输协议测量所得，在传输 900 Mbit/s 的 UDP 数据时链路无分组丢失）的国际网络链路上，TCP 改进协议的性能还有很大的提升空间。虽然在使用 Iperf 进行测量时 CUBIC、HSTCP、HTCP、Hybla 和 STCP 的性能相对突出，但是在实际传输 2 Gbyte 的文件时，STCP 的性能并没有优势，其原因在于 STCP 在拥塞避免阶段过于剧烈的窗口调整策略，极易导致网络处于拥塞的状态[44]。相反，HTCP 和 Hybla 的性能相对较好。其中，Hybla 为了提高 TCP 的 RTT 公平性，在链路中 RTT 较大时，Hybla 会加速拥塞窗口的调整，以提高其在大延时的链路中的性能[44]。值得注意的是，在传输 2 Gbyte 的文件时，Reno 也获得了不错的性能，这一方面反应了经过改进快速恢复阶段之后的 Reno，相较于传统的标准 TCP 在性能上有所改善；另一方面反应了在网络状态较好（无分组丢失且可用带宽较大）的情况下，Reno 也可以获得一定的吞吐率。

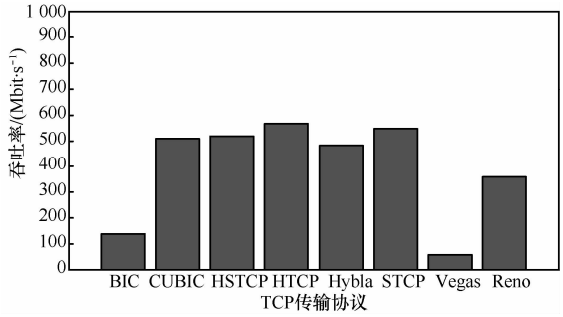


图 8　TCP 在国际链路上的吞吐率（Iperf 所测）

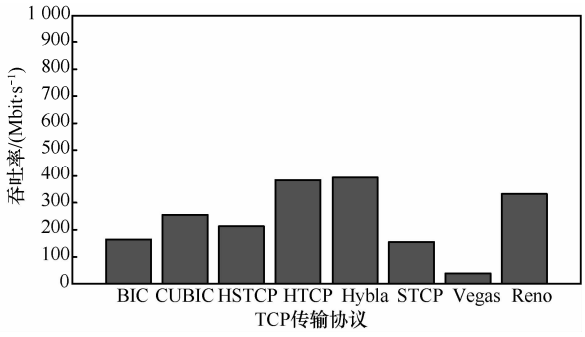


图 9　TCP 在国际链路上的吞吐率（FTP 所测）

5.2　TCP 协议在国内链路上的性能

图 10 和图 11 是在国内链路上分别采用 Iperf 和 FTP 所获得的 TCP 的传输性能。与 5.1 节所示的性能类似，采用 Iperf 所获得的传输性能也稍好于使用 FTP 所获得的性能。

由图 10 和图 11 可知，在网络时延为 20 ms，可用带宽约为 900 Mbit/s（均通过与 5.1 所示方法测量）的国内链路上，相较于国际链路，TCP 的性能有了明显的提高。这反应了 TCP 受到网络链路时延的影响这一事实：随着网络延时的降

低，TCP 的传输效率会相应的提高。图 10 使用 Iperf 所测试的 TCP 吞吐率中，BIC、CUBIC、HSTCP 和 STCP 均获得了不错的吞吐率，尤其是 HSTCP，其吞吐率已经接近 900 Mbit/s。STCP 的吞吐率也超过了 800 Mbit/s。与在国际链路上测试的结果类似，图 11 所示的是采用 FTP 传输 2 Gbyte 文件时，TCP 的性能有所下降，其中，STCP 的吞吐率仍然不够理想，其原因与在国际链路上 STCP 的吞吐率不够理想类似。相反，BIC 和 HSTCP 均获得了不错的性能，其吞吐率均超过了 700 Mbit/s，这反映了在中长距离网络中 BIC 和 HSTCP 均具有明显的传输优势，这也是 BIC 之所以能作为 CentOS 的默认 TCP 传输协议的原因所在。还需要指出的是 Hybla 传输协议的特点也得到了体现，对照 5.1 节中 Hybla 在国际链路上的性能，发现其吞吐率基本上保持不变，或者说还有所减小，其原因在于 Hybla 降低了在低延时链路上的拥塞窗口增长速率，以达到较好的 RTT 公平性[44]。这一结果与第 4 节采用 NS2 进行仿真的情况相吻合。

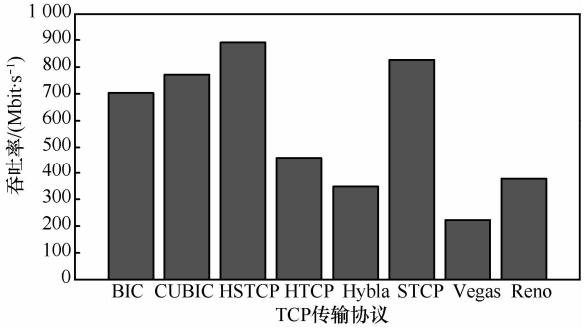


图 10 TCP 在国内链路上的吞吐率（Iperf 所测）

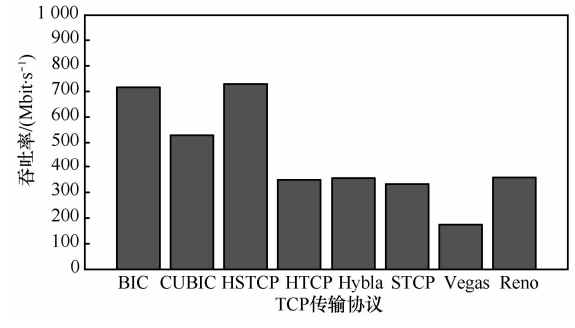


图 11 TCP 在国内链路上的吞吐率（FTP 所测）

值得注意的是，在本次实验中，发现不同的 FTP 客户端对传输效率的影响非常大，以 SFTP 客户端为例，从香港到芝加哥，HTCP 协议最大仅获得了 103 Mbit/s 的吞吐率，这与使用 LFTP 获得的接近 400 Mbit/s 的吞吐率相差甚远。究其原因是由于其自身采用了加密机制所产生的额外开销导致的。因此，除了传输协议之外，应用程序的选取在高速长距离网络中进行海量数据传输时也应该得到充分重视。

6 结束语

科研活动的信息化水平日益提高对高速长距离网络提出了新的需求，大量的高速长距离网络相继建立，与之相适应的高速 TCP 传输协议也被相继提出。本文分类讨论了目前流行的 TCP 改进协议，在此基础上分别通过仿真软件和真实网络对目前流行的 TCP 传输协议进行了系统的评价，本文发现，目前 TCP 改进协议还存在以下几个问题。

传输效率与友好性问题。高速网络首先体现在传输效率方面，如何将大量的科研数据进行高速有效的传输是用户首先考虑的问题。目前大量的改进协议主要集中在此，TCP 在高速长距离网络中的性能也得到了明显的提升。然而在提高传输效率的同时如何减小对与之竞争的数据流的影响，即提高 TCP 友好性，是一个亟待解决的问题。

RTT 的公平性问题。随着数据中心的兴起和以网络为工具的科研工作的深入，大量的数据需要被用户共享。分散在世界各地的用户由于距离数据源的距离不等，因此 RTT 也各不相同，这就存在 RTT 的不公平问题。距离数据源较近的用户往往比距离数据源较远的用户更容易抢占带宽资源，使得距离数据源较远的用户无法获得理想的传输效率[47]。目前的传输协议对此关注较少，即使一些研究人员提出了改进措施[21]，但是在提高 RTT 公平性的同时，其自身的传输效率受到了影响。

网络带宽测量技术与拥塞控制机制相结合的问题。目前流行的 TCP 改进协议的拥塞窗口调整策略所依据的反馈方式主要有分组丢失反馈、延时反馈和两者相结合的混合反馈。网络可用带宽测量技术日益成熟，使得网络带宽测量技术与拥塞窗口调整策略相结合成为可能。文献[21]使用了网络带宽测量技术，但是不管是在仿真环境中还是在实际网络中其优势并不明显，因此在此领域还有大量工作可做。

本文对 TCP 改进协议进行的评价结果以及所发现的问题，将会是下一步研究工作的重点。

参考文献：

[1] CERN[EB/OL]. http://home.web.cern.ch.

[2] eVLBI[EB/OL]. http://www.evlbi.org.

[3] Bio-mirror[EB/OL]. www.bio-mirror.net.

[4] LEITH D J, SHORTEN N, MCCULLAGH G . Experimental evaluation of cubic-TCP[A]. Protocols for Fast Long Distance Networks[C]. Los Angeles, USA, 2007.

[5] SANSA J, SZOMORU A, DER HULST J M V. An evaluation study of data transport protocols for e-vlbi[J]. International Journal of Computing and ICT Research, 2007,(5):68-75.

[6] LEITH D, ANDREW L, QUETCHENBACH T, *et al*. Experimental evaluation of delay/loss-based TCP congestion control algorithms[A]. PFLDnet 2008[C]. Tokyo, Japan, 2008.

[7] HASHIMOTO M, HASEGAWA G, MURATA M, *et al*. Performance evaluation and improvement of hybrid TCP congestion control mechanisms in wireless LAN environment[A]. IEEE Telecommunication Networks and Applications Conference[C]. Australasian, 2008. 367-372 .

[8] CHENG J, WEI D, LOW S H, *et al*. FAST TCP: from theory to experiments[J]. IEEE Network, 2005,19(1):4-11.

[9] SCHARF M, STROTBEK H. Performance evaluation of Quick-Start TCP with a Linux kernel implementation[A]. Proc IFIP Networking 2008[C]. Springer LNCS 4982. Singapore, 2008. 703-714.

[10] SCHARF M. Performance evaluation of fast startup congestion control schemes[J]. NETWORKING, 2009,(5):716-727.

[11] KUSCHNIG R, KOFLER I, HELLWAGNER H. An evaluation of TCP-based rate-control algorithms for adaptive Internet streaming of H.264/SVC[A]. Proceedings of the First Annual ACM SIGMM Conference on Multimedia Systems[C]. New York, 2010.

[12] FLOYD S, GURTOV A. RFC3649-HighSpeed TCP for Large Congestion Windows[S]. 2003.

[13] KELLY T. Scalable TCP: improving performance in highspeed wide area networks[J]. Computer Communications Review, 2003, 33(2): 83-91.

[14] XU L, RHEE I. Binary increase congestion control for fast long-distance networks[A]. IEEE INFOCOM 2004[C]. Hongkong, 2004. 2514-2524.

[15] HA S, RHEE I, XU L. CUBIC: a new TCP-friendly high-speed TCP variant[J]. SIGOPS Operating Systems Review, 2008, 42(5):64-74.

[16] LEITH D, SHORTEN R. H-TCP: TCP for high-speed and long distance networks[A]. PFLDnet 2004[C]. Illinois, USA, 2004.

[17] MARFIA G, PAU G, GERLA M, *et al*. TCP Libra: exploring rtt-fairness for UCLA Computer Science Department, Tech Rep[R]. 2005.

[18] RFC2582-the NewReno Modification to TCP's Fast Recovery Algorithm[S]. 1999.

[19] CAINI C, FIRRINCIELI R. TCP Hybla: a TCP enhancement for heterogeneous networks[J]. International Journal of Satellite Communications and Networking, 2004, 22: 547-566.

[20] WEI D X, LOW S H, HEGDE S. FAST TCP: motivation, architecture, algorithms, performance[J]. IEEE/ACM Trans Netw, 2006, 14(6): 1246-1259.

[21] PETERSON L B. TCP Vegas: end to end congestion avoidance on a global Internet[J]. IEEE J Sel Areas Commun, 1995, 13(8):1465-1480.

[22] RHEE S. Hybrid slow start for high-bandwidth and long-distance networks[A]. PFLDnet 2008[C]. Tokyo, Japan, 2008.

[23] KESHAV S. Packet-pair flow control[EB/OL]. http://www.cs.cornell.edu/skeshar/doc/9412-17.ps, 1995.

[24] DOVROLIS C, RAMANATHAN P, MOORE D. Packet-dispersion techniques and a capacity-estimation methodology[J]. IEEE/ACM Transactions on Networking, 2004,12(6): 963-977.

[25] TAN K, ZHANG Q, SRIDHARAN M. Compound TCP: a scalable and TCP-friendly congestion control for high-speed networks[A]. PFLDnet 2006[C]. Nara, Japan, 2006.

[26] KING R, RIEDI R. TCP-Africa: an adaptive and fair rapid increase rule for scalable TCP[A]. IEEE INFOCOM 2005[C]. Miami, 2005. 1838-1848.

[27] LIU S, SRIKANT R. TCP-Illinois: a loss and delay-based congestion control algorithm for high-speed networks[A]. First International Conference on Performance Evaluation Methodologies and Tools[C]. Pisa, Italy, 2006. 417-440.

[28] BAJOCCHI A, VACIRCA F. YeAH-TCP: yet another highspeed TCP[A]. PFLDnet 2007[C]. Marina DelRey, California, 2007.37-42.

[29] CASETTI C, GERLA M, MASCOLO S, *et al*. TCP westwood: bandwidth estimation for enhanced transport over wireless links[A]. ACM MOBICOM 2001[C]. Rome, Italy, 2001. 287-297.

[30] GRIECO L A, MASCOLO A, *et al*. Performance evaluation and comparison of Westwood+, New Reno, and Vegas TCP congestion control[A]. The Seventh International Symposium on Computers and Communications[C]. Alexandria, 2004. 25-38.

[31] HIDEYUKI S, TUTOMU M. Improving efficiency-friendliness tradeoffs of TCP congestion control algorithm[A]. IEEE Globecom 2005[C]. St Louis, Missouri, 2005.

[32] KANEKO K, SU Z, KATTO J. TCP-Fusion: a hybrid congestion control algorithm for high-speed networks[A]. PFLDnet 2007[C]. Marina Del Rey, California, 2007.

[33] DINA K, MARK H, CHARLTE R. Congestion control for high bandwidth delay product networks[A]. ACM SIGCOMM 2002[C]. Pittsburgh, USA, 2002.

[34] XIA Y, KALYANARAMAN L. One more bit is enough[J]. ACM Sigcomm Computer Communication REVIEW, 2005, 34: 37-48.

[35] ZHANG Y, LOGUINOV L D D. JetMax: scalable max-min congestion

control for high-speed heterogeneous networks[A]. IEEE INFORCOM 2006[C]. Barcelona, Spain, 2006. 1193-1219.

[36] HUANG X, LIN C, REN F. A novel high speed transport protocol based on explicit virtual load feedback[J]. Computer Networks, 2007, 51:1800-1814.

[37] HUANG X, LIN C, REN F Y, *et al*. Improving the convergence and stability of congestion control algorithm[A]. The 15th IEEE Internet Conference on Network Protocols (ICNP 2007)[C]. Beijing, China, 2007. 206-215.

[38] FLOYD E. RFC 5166 Metrics for the Evaluation of Congestion Control Mechanisms[S]. RFC, 2008.

[39] LARRY L P, BRUCE S D. Computer Networks: A System Approach fourth edition[M]. San Francisco: Elsevier, 2007.

[40] STALLINGS W. High-Speed Networks and Internets: Performance and Quality of Service, Second Edition[M]. USA: Prentice Hall, 2002.

[41] LEITH D J, SHORTEN R N, MCCULLAGH G. Experimental evaluation of Cubic-TCP[A]. PFLDnet 2007[C]. Marina DelRey, California, 2007.

[42] COTTRELL R L, ANSARI S, KHANDPUR P, *et al*. Characterization and evaluation of TCP and udp-based transport on real networks[A]. PFLDnet2005[C]. France, 2005.

[43] VILLAMIZAR C, SONG C. High performance TCP in ANSNET[J]. ACM Sigcomm Computer Communication REVIEW, 1994, 24: 45-60.

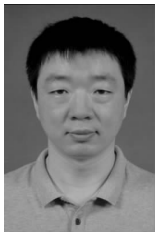
[44] AFANASYEV A, TILLEY N, REIHER P, *et al*. Host-to-host congestion control for TCP[J]. Communications Surveys & Tutorials, IEEE, 2010, 12: 304-342.

[45] CHIU D M, JAIN R. Analysis of the increase and decrease algorithms for congestion avoidance in computer networks[J]. Computer Networks and ISDN systems, 1989, 17:1-14.

[46] BULLOT H, COTTRELL R L, HUGHES J R. Evaluation of advanced TCP stacks on fast long-distance production networks[J]. Journal of Grid Computing, 2003, (1): 345-359.

[47] PAPADIMITRIOU D, WELZL M, SCHARF M, *et al*. Open Research Issues in Internet Congestion Control[S]. RFC6077, 2011.

作者简介：



王国栋（1981-），男，河南焦作人，博士，中国科学院助理研究员，主要研究方向为高速网络、传输协议。

任勇毛（1981-），男，湖南邵阳人，博士，中国科学院副研究员，主要研究方向为高速网络、传输协议。

李俊（1968-），男，安徽桐城人，博士，中国科学院研究员、博士生导师，主要研究方向为下一代互联网、高速网络。