计算机研究与发展
Journal of Computer Research and Development
ISSN 1000-1239/CN 11-1777/TP
42(4):647~654, 2005

基于分布式 PEP 的卫星网络 TCP 性能增强协议

霍龙社　郑燕峰　高　文
(中国科学院计算技术研究所　北京　100080)
(中国科学院研究生院　北京　100039)
(lshuo@jdl.ac.cn)

A Distributed PEP-Based TCP Performance Enhancing Protocol for Satellite Networks

Huo Longshe, Zheng Yanfeng, and Gao Wen
(*Institute of Computing Technology, Chinese Academy of Sciences, Beijing* 100080)
(*Graduate School, Chinese Academy of Sciences, Beijing* 100039)

Abstract Based on distributed performance enhancing proxies, a new transport protocol for the satellite environment, the XP protocol is presented. It is designed and optimized to overcome the performance degradation that TCP experiences in satellite networks due to their long latency, high bit error rate, and bandwidth asymmetry characteristics. Main contributions include a reliable two-and-half-way handshake connection establishment mechanism, rate control and a measurement-based adaptive bandwidth allocation algorithm, and a delayed acknowledgements technique based on active sender-request. Interactions with the TCP connections in terrestrial segments are also considered. Emulation and real operating environment experiments show that the throughputs achieved by XP improve substantially in comparison with the regular TCP, and the fairness is also maintained perfectively when the bandwidth is shared among multi-connections.

Key words satellite networks; TCP protocol; performance enhancing proxy; throughput

摘　要　提出了一个基于分布式性能增强代理的卫星网络专有通信协议:XP 协议,用于解决卫星网络环境中因长时延、高误码率和非对称信道带宽等因素所导致的 TCP 传输性能低下问题. 协议的设计考虑到了与地面链路上 TCP 连接的接口关系和多连接共享同一卫星信道时的带宽分配问题. 主要贡献包括两路半握手连接建立机制,速率控制和基于测量的动态带宽分配算法,以及基于发送方主动请求的延迟确认技术等. 仿真和真实环境实验表明,分布式性能增强代理和 XP 协议的使用可显著提高网络中下行卫星链路的吞吐量,多数情况下带宽资源利用率可提高至 85% 以上,且在多连接共享带宽的情况下能够保持较好的公平性.

关键词　卫星网络;TCP 协议;性能增强代理;吞吐量

中图法分类号　TP393

收稿日期:2003-12-01;**修回日期**:2004-11-03

基金项目:国家自然科学基金项目(69983008);国家"八六三"高技术研究发展计划基金项目(2001AA112100);中国科学院知识创新工程基金项目(KGCXZ-103)

1 引 言

卫星通信具有覆盖面广、可扩展性强、用户接入方便和不受地域条件限制等优点,是地面光纤难以敷设或人口稀疏等偏远地区进行长距离宽带网络接入的一种重要补充手段. TCP 协议是目前 Internet/Intranet 网络中应用最为广泛的传输层协议,已成为端到端可靠数据传输的事实标准. 然而,当将 TCP 应用于具有卫星链路的网络环境时,由于卫星信道区别于地面链路的一些固有特征,其传输性能会受到严重影响,吞吐量急剧下降.

卫星网络的特点及其对 TCP 性能的影响主要体现在以下几个方面[1]:①长传输时延. 这会影响 TCP 的慢启动算法[2],使其需要花费较长时间才能达到最优的传输速率. 由于长时延而带来的大的带宽时延乘积要求发送方必须维持一个大的发送窗口才能充分利用带宽. 但标准 TCP 的最大发送窗口为 64KB,许多 TCP 实现中缺省的发送窗口是 8KB,在卫星通信中只能获得较小的吞吐量.②高误码率. 与其他地面通信介质相比,卫星信道具有较高的误码率. TCP 默认数据包丢失为网络拥塞的标志,并调用慢启动算法来减小发送窗口值以避免拥塞. 这对于由于信道传输误码所引起的数据包丢失来说并不合适,而只会使传输性能更加恶化.③非对称上下行信道. 用于向卫星发射信号的设备价格昂贵,故具有卫星链路的双向通信一般采用非对称信道. 例如在下行链路采用高速卫星信道,而在上行链路采用低速地面连接(如电话拨号等). 由于 TCP 协议需要利用反向链路中的确认消息(ACK)来触发发送方数据的持续发送,因此这种上下行链路带宽的不对称性会导致上行链路中 ACK 的拥塞,并反过来影响下行链路的传输性能.

针对如何提升卫星网络环境中 TCP 协议传输性能的研究主要可分为两大类. 第 1 类仅对终端节点上的 TCP 协议本身进行改进,例如采用大的初始化窗口[3]、字节计数[4]、慢启动之后的延迟确认[4]、T/TCP[5],选择性确认[6]和前向确认[7]等. 这类算法需要修改终端节点上的系统软件,会带来兼容性问题. 另一类则可对终端节点之间的整个网络系统进行调整,其中较典型的是在网络系统的中间节点处引入性能增强代理(performance enhancing proxy,PEP). RFC 3135[8]中根据 PEP 的分布性又将其分为两类:集中式 PEP 和分布式 PEP. 集中式 PEP 一般位于靠近 TCP 连接服务器一侧的卫星基站上. 例如 Snoop[9]通过在基站路由器的链路层插入代理程序来缓存分组和执行本地重传,将无线链路上的错误与有线端的服务器隔离;I-TCP[10]则在传输层将一条端到端的 TCP 连接分为两截,然后针对有线和无线两部分分别进行处理. 如果整个网络中存在多个 PEP,则称之为分布式 PEP. 分布式 PEP 一般驻留在网络中卫星链路的两侧,即卫星基站和卫星接收站中. 这种方式可以在两个 PEP 之间的卫星链路这一段采用不同于 TCP 的专有协议,专门针对卫星信道的固有特点进行优化,因此可取得更好的性能增强效果. STP[11]是一个适用于分布式 PEP 之间卫星数据通信的专有传输协议,它通过由发送方定时向接收方发送 POLL 分组来主动请求接收方发送确认分组,从而减缓反向链路的 ACK 拥塞. 但 STP 协议的设计仅考虑了两个 PEP 之间的通信问题,而没有考虑与 PEP 两侧地面链路上 TCP 连接之间的接口关系,也没有考虑多连接共享同一卫星链路带宽的情况. PERTA[12]是最近提出的一个卫星网络性能增强传输结构,它也采用分布式 PEP 模式,并在两个 PEP 之间的卫星链路上采用了一个基于 STP 进行改进的传输协议 STPP. 但它为 PEP 两侧的地面链路也设计了专有的传输层协议,而不是采用标准的 TCP 连接,因此仍然存在兼容性问题,仅适合于一个企业或集团内部所组建的私有网络系统.此外 STPP 协议也没有考虑多连接共享卫星链路带宽的情况.

本文采用基于分布式 PEP 的网络结构并提出一个用于 PEP 间通信的卫星网络 TCP 性能增强协议:XP 协议. 其性能优化的主要目标在于提高网络中卫星下行链路的吞吐量和带宽资源利用率. 具体做法为在卫星链路的两侧接入点各设置一个 PEP 网关,从而将一条端到端的 TCP 连接分为 3 段:PEP 网关两侧的地面链路上仍采用标准 TCP 协议,两个 PEP 之间则采用 XP 协议. XP 协议的设计除了考虑 PEP 间卫星网络通信的优化问题外,还考虑了与 PEP 两侧地面链路上 TCP 连接的接口关系. PEP 网关软件负责对 TCP 连接进行透明拦截和协议转换,而不需要对终端节点的协议栈和应用程序做任何改动,因此兼容性好,便于实际应用.

2 网络结构

XP 协议所适用的基于分布式 PEP 的卫星网络

结构如图 1 所示. 图中 Client 通过非对称混合卫星网络接入 Internet,访问其上的 Server 资源. 在靠近 Client 一侧的每个卫星接收站各配置一个接入 PEP 网关,称之为 GW1;在靠近 Server 一侧的卫星基站处配置一个中心 PEP 网关,称之为 GW2. XP 协议的实现在 GW1 和 GW2 之间完成. 与 TCP 类似,XP 是一个双向交互协议:从 GW2 发送至 GW1 的数据通过卫星信道来传递,称之为下行链路;从 GW1 发送至 GW2 的数据通过低速地面链路(如电话拨号)来传递,称之为上行链路. 作为中心接入网关,一个 GW2 可以与多个 GW1 进行交互,此时多个 GW1 共享同一下行链路带宽资源.

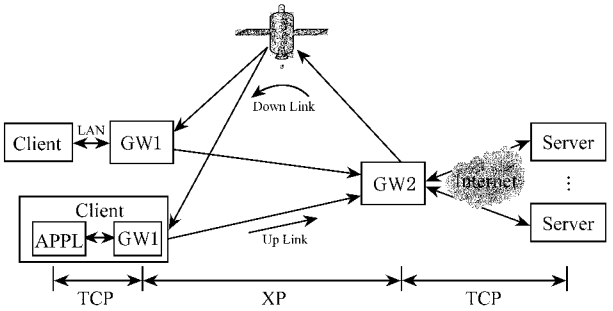

Fig. 1 The satellite network architecture suitable for XP.

图 1 XP 协议适用的卫星网络结构

该网络结构的协议栈模型如图 2 所示. XP 协议建立在 UDP 协议之上并在应用层实现. GW2 网关软件包括 XP 协议和协议转换模块两部分. GW1 中除了这两部分外,还有一个重定向驱动模块,位于数据链路层和 IP 层之间,用于透明拦截 Client 向 Server 发起的 TCP 连接请求并把它重定向至上层的协议转换模块. 所有软件模块的安装不需要对系统中原有的 TCP/IP 协议栈和系统软件做任何改动. GW1 既可作为一个单独的设备存在,也可以作为独立的软件模块安装于配置有卫星接收卡的单机 Client 中.

在上述网络结构中,从 Client 到 Server 之间的一次 TCP 会话过程如下:当 Client 向 Server 发起 TCP 连接请求时,GW1 对其进行透明拦截,代理 Server 与 Client 建立一条 TCP 连接. 然后 GW1 向 GW2 发起连接请求,双方建立一条 XP 连接. 接下来 GW2 再向 Server 发起 TCP 连接请求,代理 Client 与 Server 建立一条 TCP 连接. 当 Client 向 Server 发送数据时,实际上是首先通过 TCP 发送至 GW1,GW1 将其转换成为 XP 格式后发送至 GW2,再由 GW2 将其转换成为 TCP 发送至 Server. Server 至 Client 的数据传输也是同理.

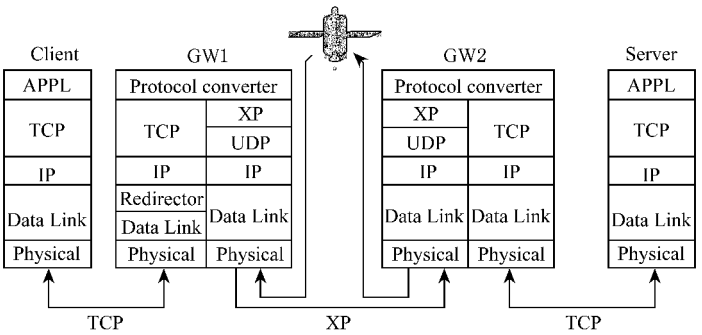

Fig. 2 Network protocol stack model.

图 2 网络协议栈模型

3 XP 协议主要实现技术

3.1 连接建立

为加快连接建立过程,缩短因长时延而带来的连接等待时间,XP 中提出了一个称之为两路半握手的连接建立机制,如图 3 所示.

当 Client 向 Server 发起 TCP 连接请求时,该请求被 GW1 截获,并在 Client 和 GW1 之间建立起一条 TCP 连接. 然后 GW1 向 GW2 发送 XP 连接请求 CONN_REQ,并捎带从 Client 接收到的第 1 个数据分组. GW2 收到该 CONN_REQ 后,首先与真正的 Server 建立 TCP 连接,然后向 GW1 发回连接确认 CONN_ACK. 这时不必等待从 GW1 返回连接再确认,GW2 立即进入数据传输状态,开始把从 Server 接收到的数据向 GW1 转发. 当 GW1 接收到从 GW2 返回来的 CONN_ACK 后,首先向 GW2 发送连接再确认 CONN_ACK,然后转入数据传输状态,开始接收从 GW2 转发来的数据分组.

从上述过程可以看到,自 GW1 发起连接建立请求至它接收到来自 Server 的第 1 个数据分组,中间只有一个 RTT(往返时间)的等待时间. 与 TCP

3 次握手所需要的至少两个 RTT 等待时间相比,XP的连接建立反应速度可以提高一倍,这对于HTTP等交互式应用来说极为有利.

由于 GW2 在未接收到从 GW1 返回来的连接再确认之前便开始传输数据,因此该过程被称之为两路半握手. 上述过程会导致一种半连接情况的出现 如果从 GW2 向 GW1 返回来的 CONN_ACK 在途中丢失,GW1 在超时未接收到的情况下会向 GW2 请求重新建立另一条连接;但 GW2 并不知道这种情况而继续在原有连接上向 GW1 发送数据. 为防止这种半连接情况出现,我们在 GW2 上设定一个超时:在 GW2 向 GW1 发送 CONN_ACK 的同时启动一个定时器,如果在一定的超时时间间隔内未收到从 GW1 返回来的连接再确认,则 GW2 停止发送数据,并强行中止当前的半连接.

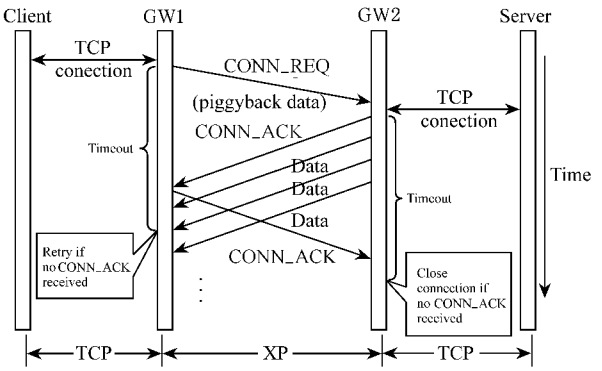

Fig. 3 Two-and-half-way handshake connection establishment.

图 3 两路半握手连接建立示意图

3.2 速率控制与带宽分配

当 XP 连接进入数据传输状态之后,由于 GW2 至 GW1 之间的卫星链路只有一个跳段,且带宽和时延均保持基本不变,因此在这一段链路中不存在网络拥塞问题,也就不需要采用类似 TCP 的慢启动算法来探测拥塞和增大减小发送窗口,而是可以在数据传输的一开始便采用最大可能的窗口值或速率来发送数据,从而充分利用带宽资源,获得最大可能的吞吐量.

虽然卫星链路内部不存在拥塞,但在 GW2 中的协议转换接口处却仍存在着拥塞问题. 即当 Server 至 GW2 之间的所有 TCP 连接的总吞吐量大于 GW2 至 GW1 之间卫星链路的固定带宽时,如果对于每个连接在 GW2 上都仍按照它从 TCP 连接接收数据的速率在卫星链路上进行转发,则必然会在该节点链路层导致拥塞和分组丢失,从而使传输性能恶化. 为解决上述问题,我们在 GW2 处引入数据缓冲和速率控制机制,过程如下(参见图 4 所示):

(1) 为每个连接分配两个发送缓冲区——待发送队列(SQA)和已发送队列(SQB). 从 TCP 侧接收到的数据分组首先进入 SQA 进行排队,等待发送.

(2) 调用带宽分配算法计算各连接在卫星链路上的传送速率.

(3) 对于每一连接,按照上述计算的传送速率将 SQA 中的数据分组发送至卫星链路. 发送后的分组缓存于 SQB 中,以备出错时重传.

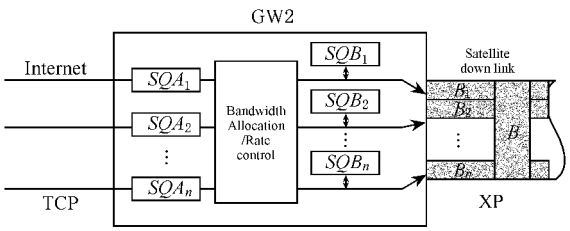

Fig. 4 Rate control and bandwidth allocation.

图 4 速率控制与带宽分配

上述第(2)步需要调用一个合适的带宽分配算法来为每一条连接分配带宽. 由于在系统运行过程中不断有连接加入和退出,且每一连接在 TCP 侧的速率也随时间而波动,因此该带宽分配算法也应该能够动态适应,以保证每一时刻下行卫星链路的带宽利用率都能够达到最优且每个连接都能够得到较为公平的服务. 该问题可以形式化如下:

设下行链路总带宽为 B,某一时刻 t 存在 n 条连接. 则问题的目标是为每条连接分配带宽 B_1, B_2,…,B_n,使得 $\sum_{i=1}^{n} B_i$ 最大,且满足如下两个条件:

(1) $\sum_{i=1}^{n} B_i \leqslant B$;

(2) 每条连接的带宽分配尽可能公平.

其中条件(2)没有一个现成的评价标准. 考虑到 XP 与 TCP 之间的转接关系,我们将该公平性定义为使 GW2 至 GW1 之间下行链路上各 XP 连接的带宽分配比例尽可能与 Server 至 GW2 之间各 TCP 连接的到达速率比例相一致,并提出一个基于测量的动态带宽分配算法 MBABA(measurement-based adaptive bandwidth allocation)如下:

(1) 按固定时间间隔 Δt 将时间轴分割为离散的时间片:$[t_0, t_1]$, $[t_1, t_2]$,…, $[t_{k-1}, t_k]$, $[t_k, t_{k+1}]$,….

(2) 在每一时间点 t_k,统计当前活动的连接数 n,以及在时间间隔 $[t_{k-1},t_k]$ 内各连接从 TCP 一侧接收到的数据字节数,记为:$D_{k,1},D_{k,2},\cdots,D_{k,n}$.

令 $D_k=\sum_{a=1}^{n}D_{k,a}$.

(3) 如果存在某条连接 $i(1\leqslant i\leqslant n)$,有 $D_{k,i}=0$,且 SQA_i 不为空,则令 $D_{k,i}=D_{k,i-1}$.

(4) 分别为每一连接 $i(1\leqslant i\leqslant n)$ 计算带宽:$B_i=B\times(D_{k,i}/D_k)$.

上述算法实际上是根据各连接在 TCP 侧上一个时间片的吞吐量比例值来预测它们在 XP 侧下一个时间片的带宽比例值. 当某连接在上一时间片没有接收到数据但其 SQA 中仍有数据等待发送时,则不能简单地将其带宽值置为 0,而是按其在上上个时间片接收到的数据量作为权值来参与带宽分配,从而保证其 SQA 中缓存的数据还能够继续发送. 由于 TCP 连接在经过一个较短时间的慢启动之后传送速率一般可趋于稳定(在一个较为稳定的范围内做锯齿状波动),故只要时间片参数选择适当,该算法能够做到使下行链路中各 XP 连接带宽所占比例与该连接从 TCP 侧接收数据速率所占比例基本一致,且总的带宽利用率可达到最优.

与需要有集中控制的 round robin 算法相比,MBABA 可以方便地在多个进程或线程间进行分布式的带宽分配. 具体做法是将某一时刻的活动连接数和当前时间片内所有连接从 TCP 侧接收到的数据总量 D_k 定义为所有连接都可以访问的共享变量,然后各连接便可以独自运行,且仅通过对这两个共享变量的访问来完成动态带宽分配任务.

3.3 流量控制

受 GW2 内存容量限制,需要为每个连接的发送缓冲区 SQA 和 SQB 大小设置上限. 其中 SQB 的上限设置为该连接的带宽时延乘积,SQA 的上限作为系统参数在安装时根据设备的实际内存大小进行配置. 当 GW2 从 TCP 连接接收数据的速率大于 GW2 向卫星链路发送数据的速率时,除了速率控制之外还需要进行适当的流量控制,以保证系统不会因为 SQA 或 SQB 缓冲区的溢出而丢失数据.

XP 协议中的流量控制较为简单,通过设置两个标识 ES 和 ER 来控制数据的收发. 当 SQB 队列满时,复位 ES,不再向卫星链路发送数据;当 SQB 队列大小减小至最大队列长度的一半时,置位 ES,重新开始向卫星链路发送数据. 当 SQA 队列满时,复位 ER,禁止从 TCP 连接接收数据;当 SQA 队列大小减小至最大队列长度的一半时,置位 ER,重新开始从 TCP 连接接收数据.

3.4 基于发送方主动请求的延迟确认

TCP 协议要求数据接收方每接收到一个分组便对其进行确认. 在上下行带宽不对称的网络中,频繁的确认分组会导致上行链路中 ACK 拥塞,进而影响发送方数据的正常发送.

XP 协议不需要根据接收方的确认来触发发送方数据的持续发送,但仍然要求接收方对其接收到的数据分组进行确认,因为发送方需要根据这些确认信息来清除其 SQB 队列中缓存的接收方已正确接收的数据分组,以避免缓冲区空间增长过大. 但与 TCP 不同,XP 并不要求接收方每接收到一个分组便返回一个确认,而是只有当发送方主动向接收方提出确认请求时,接收方才向其发回一个确认分组,具体过程描述如下:

(1) 在 GW2 中计算一个周期值;

(2) GW2 每发送 K 个数据分组后,在其下一个数据分组中捎带 ACK 请求信息(将分组类型从 DATA 改为 DATA_ASKACK);

(3) GW1 只有在接收到一个带 ACK 请求的数据分组时,才向 GW2 返回一个 ACK 分组,对在此之前所收到的所有数据分组进行确认;

(4) GW2 接收到一个 ACK 后,从其 SQB 队列中清除由该 ACK 所确认的所有数据分组.

其中 K 的取值应该在保证上行链路不发生拥塞的前提下尽量缩短发送方等待 ACK 的时间. 设下行链路带宽为 B_d,下行数据分组大小为 P_d,上行链路带宽为 B_u,上行 ACK 分组大小为 P_u,则 K 的计算公式为 $K=\left\lceil (B_d/B_u)/(P_d/P_u)\right\rceil+\alpha$. 其中 α 为调整因子,用于兼顾上行链路中除了 ACK 之外还存在的其他类型数据,具体实现中通常取一个较小的整数值.

上述延迟确认的采用可以在不影响下行链路吞吐量的前提下大大缓减对带宽不对称网络中上行链路的带宽需求,从而提高系统传输性能.

3.5 差错控制

由于使用了延迟确认,故类似 TCP 中基于超时重传的差错控制在 XP 协议中不再可行. 为此 XP 采用了基于数据驱动 ARQ(自动请求重传)的差错控制. 由于下行链路的数据收发方之间只有一跳卫星链路,其间不存在拥塞,故分组在该链路上的传输必定按顺序进行而不会失序. 如果接收方接收到的

数据分组序列号之间出现间隔,则可认定中间序列号的分组是因误码而丢弃,而不是因为失序而引起的. 这时接收方就可以立即发送 NACK,主动请求发送方重传丢失的数据报文. 为提防 NACK 或重传的数据分组丢失,XP 在接收方为每个 NACK 设置了一个超时. 若接收方在发出 NACK 之后一个预期的超时时间间隔内仍未收到由它指定重传的数据报文,则重新发送该 NACK. 该机制使得在网络中出现因误码而丢包的情况时,发送方不必降低发送速率,而出错重传的反应时间则可缩至最短. 为防止因下行卫星链路误码或上行链路丢包严重恶化而导致的连接死锁,XP 协议在 GW1 和 GW2 中各设置了一个空闲超时. 如果某个连接在该超时时间间隔内未接收到任何来自对方的 XP 分组,则强行中止该连接.

4 实 验

4.1 仿真实验

在 Linux 操作系统中运行 NIST Net 软件[13]来进行仿真实验. NIST Net 可以在同一台计算机的两个网络接口间插入模拟的带宽、时延和丢包率等网络参数,而网络接口之上运行的仍然是真实的协议栈和应用程序.

(1) 传输性能测试

在图 5 所示的仿真网络环境中,用以下 3 组实验来测试和比较在不同参数条件下 XP 和标准 TCP 的传输性能.

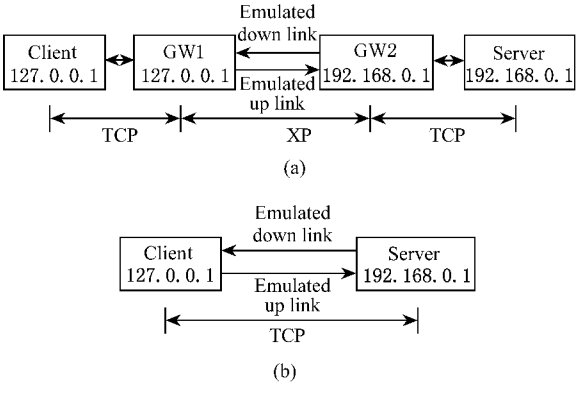

Fig. 5 NIST net emulation platform.
(a) For XP and (b) For TCP.
图 5 NIST net 仿真实验环境.
(a) XP 协议;(b) TCP 协议

① 固定上行链路带宽为 34Kb/s,*RTT* 为 320ms,通过改变下行链路带宽来测试其对下行链路吞吐量的影响. 如图 6 所示,采用 XP 时的信道吞吐量随着带宽的增大基本呈线性增长,带宽利用率始终保持在 90% 以上;而仅采用 TCP 时的吞吐量在带宽增大到一定值之后就基本不再增加.

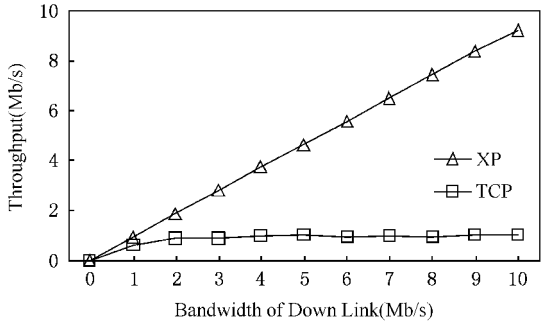

Fig. 6 Throughput vs. bandwidth.
图 6 带宽变化对性能的影响

② 固定下行链路带宽为 10Mb/s,上行链路带宽为 34Kb/s,通过改变下行链路的时延来调整 *RTT* 值,测试其对下行链路吞吐量的影响. 如图 7 所示,随着时延的增大,采用 XP 时的吞吐量缓慢下降,在 *RTT* 到达 1000ms 之前其信道利用率都仍保持在 80% 以上;而仅采用 TCP 时的吞吐量则随着时延的增大而迅速降低.

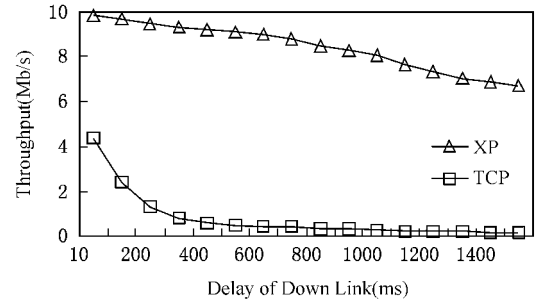

Fig. 7 Throughput vs. *RTT*.
图 7 *RTT* 变化对性能的影响

③ 固定下行链路带宽为 10Mb/s,上行链路带宽为 34Kb/s,*RTT* 为 320ms,通过改变下行链路的丢包率来测试其对下行链路吞吐量的影响. 如图 8

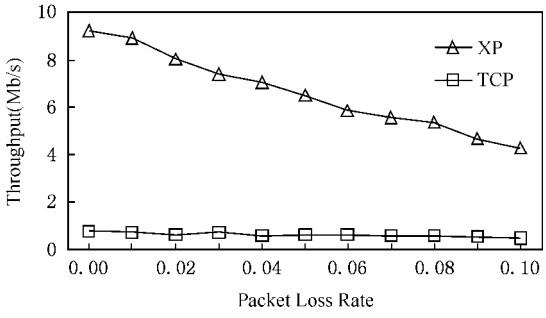

Fig. 8 Throughput vs. packet loss rate.
图 8 丢包率变化对性能的影响

所示,随着丢包率的增大,采用 XP 时的吞吐量逐渐下降,但仍保持在一个较高的水平;而仅采用 TCP 时的吞吐量则始终保持在一个极低的水平.

上述仿真实验结果表明,在具有长时延、高误码率和上下行信道带宽不对称的网络环境中,加入 PEP 网关并采用 XP 协议与单纯采用 TCP 协议相比,下行卫星链路的吞吐量有大幅提高,性能改善较为明显.

(2) 多连接共享带宽测试

本实验用于测试 MBABA 算法的性能. 为了进行比较,我们提出另一个简单的平均带宽分配算法 ABA(average bandwidth allocation). ABA 在网络中有新的连接加入或老的连接退出时,按照当前共享带宽的所有连接数将下行带宽平均分配给每一连接.在图 5(a)所示 XP 仿真实验环境中,我们建立 4 个 Server,各个 Server 至 GW2 之间的链路带宽分别设为 1Mb/s,2Mb/s,3Mb/s 和 4Mb/s,并设定 GW2 与 GW1 之间的下行链路带宽为 10Mb/s. 然后在 Client 和各个 Server 之间建立连接,并下载大小分别为 10MB,20MB,30MB 和 40MB 的文件,观察各连接的下载时间以及吞吐量情况.

如表 1 所示,在采用 MBABA 算法的情况下,下行链路中各连接的带宽分配基本与它们在 Server 侧的 TCP 连接带宽比例相匹配,总的带宽利用率高达 93.7%. 而当采用 ABA 算法时,由于下行链路带宽平均分给了各连接,因此在第 1 和第 2 两个连接结束之前的 88.3s 时间段,前两个连接的带宽过剩而后两个连接带宽不足,从而导致总的带宽利用率明显下降.

Table 1 Results for Multi-Connections Shared Bandwidth
表 1 多连接共享带宽测试结果

Connection	TCP Bandwidth at the Server Side(Mb/s)	File Size (MB)	Download Time(s)	
			MBABA	ABA
1	1	10	89.3	88.3
2	2	20	89.4	88.3
3	3	30	89.4	103.6
4	4	40	89.5	122.1
Average Throughput of the Down Link(Mb/s)			9.37	6.87

4.2 真实卫星网络环境实验

真实卫星网络测试环境如图 9 所示. 下行链路采用中科院研究生院远程教育网所租用的鑫诺卫星信道,测试中分配的带宽为 2.5Mb/s;上行链路由 GW1 先通过电话拨号连接至 Internet 服务提供商(如 263),然后再通过 Internet 路由至 GW2,拨号连接带宽为 56Kb/s.

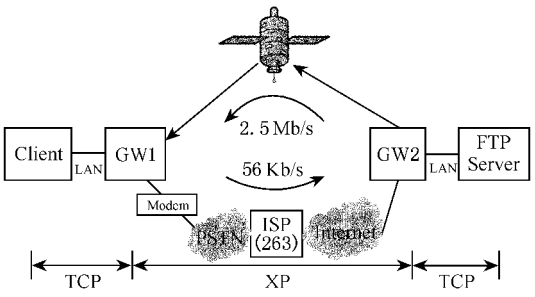

Fig. 9 Real satellite network experimental platform.
图 9 真实卫星网络实验环境

在上述卫星网络环境中,我们通过下载一组不同大小的文件来测试和比较 XP 和标准 TCP 的性能. 如图 10 所示,采用标准 TCP 时的下行链路吞吐量仅能达到 0.85Mb/s 左右,带宽利用率不足 35%;而采用 PEP 网关和 XP 协议之后的吞吐量平均可达 2.2Mb/s,带宽利用率高达 88%以上. 该测试结果表明本文所提出的分布式 PEP 网络结构和 XP 协议可用于在真实卫星网络环境下改善 TCP 协议的传输性能,提高卫星链路带宽资源的利用率.

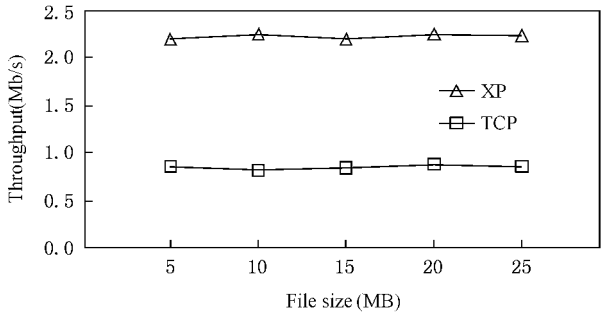

Fig. 10 Experimental results in real satellite network.
图 10 真实卫星网络实验结果

5 结 论

本文提出了一个基于分布式 PEP 的卫星网络环境 TCP 性能增强协议,并针对其所适应的网络结构和协议实现中的关键技术进行了介绍和分析. 基于两路半握手的连接建立机制可使连接建立等待时间至少缩短一半,并能有效处理与两侧 TCP 连接的衔接关系,保证连接的可靠性;速率控制和基于测量的动态带宽分配算法使得发送方能够始终以最大可能的速率发送数据,同时又保证了多连接共享带宽的公平性;流量控制使得中心 PEP 网关不至于因缓冲区溢出而丢失数据;基于发送方主动请求的延迟确认机制可用于缓减上下行带宽不对称网络中上行链路的带宽需求;基于数据驱动的差错控制使得在

传输信道出现误码时发送方不必降低发送速率,而出错重传的反应速度亦得以加快. 仿真和真实环境实验表明,本文所提出的 XP 协议,在具有长时延、高误码率和上下行带宽不对称的网络环境中其性能远远优于标准 TCP 协议,下行卫星链路的带宽资源能够得到充分利用.

致谢 感谢贺思敏博士以及课题组其他成员在论文工作中给予的大力支持和帮助. 感谢《计算机研究与发展》刊物的审稿人对论文修改所提出的宝贵意见.

参 考 文 献

1 M. Allman, D. Glover, L. Sanchez. Enhancing TCP over satellite channels using standard mechanisms. RFC 2488. http://www.faqs.org/rfcs/rfc2488.html, 1999

2 M. Allman, V Paxson, W Stevens. TCP congestion control. RFC 2581. http://www.faqs.org/rfcs/rfc2581.html, 1999

3 M. Allman, S. Floyd, C. Partridge. Increasing TCP's initial window. RFC 2414. http://www.faqs.org/rfcs/rfc2414.html, 1998

4 M. Allman, *et al*. Ongoing TCP research related to satellites. RFC 2760. http://www.faqs.org/rfcs/rfc2760.html, 2000

5 R. Braden. T/TCP—TCP extensions for transactions functional specification. RFC 1644. http://www.faqs.org/rfcs/rfc1644.html, 1994

6 M. Mathis, J. Mahdavi, S. Floyd, *et al*. TCP selective acknowledgment options. RFC 2018. http://www.faqs.org/rfcs/rfc2018.html, 1996

7 M. Mathis, J. Mahdavi. Forward acknowledgment: Refining TCP congestion control. ACM SIGCOMM 1996, Stanford, CA, USA, 1996

8 J. Border, M. Kojo, J. Griner, *et al*. Performance enhancing proxies intended to mitigate link-related degradations. RFC 3135. http://www.faqs.org/rfcs/rfc3135.html, 2001

9 H. Balakrishnan, S. Seshan, R. H. Katz. Improving reliable transport protocol and handoff performance in cellular wireless networks. ACM Wireless Networks, 1995, 1(4): 469~481

10 A. Bakre, B. R. Badrinath. Implementation and performance evaluation of indirect TCP. IEEE Trans. Computers, 1997, 46(3): 260~278

11 T. R. Henderson, R. H. Katz. Transport protocols for Internet-compatible satellite networks, IEEE J. Select. Areas Commn., 1999, 17(2): 326~344

12 M. Marchese, M. Rossi, G. Morabito. PERTA: Performance enhancing transport architecture for satellite communications, IEEE J. Select. Areas Commn., 2004, 22(2): 320~332

13 Internetworking Technology Group (ITG), National Institute of Standards and Technology (NIST). http://is2.antd.nist.gov/itg/nistnet/, 2003-12

Huo Longshe, born in 1968. Ph. D. candidate and senior engineer. His research interests are in the areas of computer networks and multimedia communications.

霍龙社,1968 年生,博士研究生,高级工程师,主要研究方向为计算机网络和多媒体通信.

Zheng Yanfeng, born in 1975. Ph. D. candidate. His research interests are in the areas of computer networks and multimedia communications.

郑燕峰,1975 年生,博士研究生,主要研究方向为计算机网络和多媒体通信.

Gao Wen, born in 1956. Professor, Ph. D. supervisor. His research interests are in the areas of multimedia data compression, image processing, computer vision, multimodal interface, and artificial intelligence.

高文,1956 年生,博士,教授,博士生导师,主要研究方向为多媒体数据压缩、图像处理、计算机视觉、多模式接口和人工智能等.

Research Background

This work is partially supported by the Knowledge Innovation Program of the Chinese Academic Science under grant No. KGCXZ-103, the National Natural Science Foundation of China under grant No. 69983008, and the National High Technology Development 863 Program of China under grant No. 2001AA12100. Satellite systems have been an important element of telecommunications networks for many years serving, in particular, long distance telephony, data, and television broadcasting. The involvement of satellite in Internet protocol (IP) networks is a direct result of new trends in global telecommunications, where Internet traffic will hold a dominant share in the total network traffic. The TCP protocol is the connection-oriented, end-to-end reliable transport protocol widely utilized in the Internet. Unfortunately, it has been designed to be effective over wired networks, and often experiences severely performance degradation when the underlying channel is characterized by features such as large bandwidth delay product, high error rates, and bandwidth asymmetry, which are the case of heterogeneous satellite networks. In this paper, we present a performance enhancing transport architecture for satellite environment, and a novel transport protocol specifically used for the satellite links. The main contributions include a reliable two-and-half-way handshake connection establishment mechanism, a measurement-based adaptive bandwidth allocation algorithm, and a delayed acknowledgements technique based on active sender-request. Emulation and real operating environment experiments show that the throughputs achieved by the proposed architecture improve substantially in comparison with regular TCP, and the fairness is also maintained perfectly when the bandwidth is shared among multiple connections.