以太网带宽管理机制下 TCP 性能的建模研究

张轶博　王海霞　雷振明

（北京邮电大学信息工程学院 ATM 中心　北京 100876）

摘　要　随着以太网技术的迅速发展，基于以太网的链路层流量控制技术成为流量工程的一个重要组成部分，基于漏桶式流控的以太网带宽控制算法被应用于驻地网环境当中。考虑到互联网中长时 TCP 业务流量增长迅速这一背景，评估以太网流量控制算法对长时 TCP 性能的影响有其积极意义。论文通过建模分析的方法，对于 NewReno TCP 算法在漏桶式带宽控制机制下的性能进行了分析，并通过仿真对结论进行了说明。

关键词　TCP，带宽分配机制，以太网，性能分析

Simulation Evaluation of TCP Throughput: A Simple Model of Bandwidth Sharing Mechanism in IEEE 802. 3 Ethernet

ZHANG Yi-Bo　WANG Hai-Xia　LEI Zhen-Ming

(ATM R&D Center, School of Information Engineering, Beijing Univ. of Posts and Telecommunications, Beijing 100876)

Abstract　With the development of Ethernet, the technique of link-layer based traffic control becomes one part of Traffic Engineering. Meanwhile, as over 70 percents of Internet traffic is TCP, it is necessary to evaluate TCP throughput under traffic control environment. In this paper, a simple model is given to evaluate the performance of long-term Newreno TCP in bandwidth sharing Ethernet. Simulation result shows that the model works well.

Keywords　TCP throughput, Bandwidth sharing mechanism, Ethernet, Performance analysis

1　引言

近年来，以太网技术发展迅速，已经成为一种重要的宽带网络接入方式。随着用户群的扩大和增值业务的发展，驻地网的结构日趋复杂，用户对于业务质量的要求也愈发强烈。为了便于计费和管理，驻地网在链路层对于用户流量进行了隔离控制。同时，这一举措也有助于防止恶意用户对于网络资源的独占，提高了资源共享的公平性。

目前，以太网链路层流量控制方面的研究方兴未艾，一些研究成果已经应用到实际当中。IEEE 在 802. 3x 标准中使用 XOFF/ON 帧对发送行为进行控制[1]。在此基础上，文[2,3]提供更为精细的控制策略：管理实体可以根据报文的 MAC 地址、业务优先级等信息对于不同的业务流实施相应的控制行为。为了便于工程实现，文[4]对文[1]进行了扩展，并结合了文[2,3]的思想，提出了基于漏桶方式的带宽分配控制策略，该策略使得在大规模以太网中实施带宽管理控制成为可能[4]。

由于文[2～4]的研究均以链路层控制为出发点，因而并未考虑链路层流控对传输层性能的影响。然而无论是 IEEE 采用的 XOFF/ON 方式还是文[4]所采用的漏桶式流控机制都会造成连续的报文丢弃，这对于使用 PAR (Positive Acknowledgement and Retrans- mission)机制的 TCP 协议的影响是巨大的。另一方面，目前互联网流量绝大部分是 TCP，而且长时 TCP 业务的比例正在不断上升[5]。因此，研究以太网环境中带宽管理策略对于长时 TCP 性能的影响具有积极意义。

本文对于漏桶式带宽管理策略下的 TCP 性能进行分析，第 2 节介绍漏桶式带宽管理机制；第 3 节根据以太网带宽管理机制的特点建立分析模型；第 4 节基于已建立的模型对于采用 NewReno TCP 算法的长时流的吞吐量进行了分析；最后通过仿真对分析结果进行了验证。

2　以太网带宽分配管理机制

本节介绍漏桶式流控的原理和基于漏桶式流控的带宽管理机制。

首先说明令牌的概念，令牌一种虚拟资源，一个令牌代表了流控实体的流控粒度 L_w。漏桶中的令牌总量代表了当前系统允许通过的数据量，报文通过漏桶时将消耗令牌，流控实体可以通过控制令牌的发放对数据流进行控制。例如，若漏桶中的令牌个数为 N_t，则此时允许通过的最大数据量为 $L_m = L_w \cdot N_t$，当到达的报文长度小于 L_m 时，报文可以通过，反之，到达报文长度大于时 L_m，令牌个数不足，报文不能通过。

运用漏桶算法进行流控时，首先根据被控流的速率和突发程度设定漏桶令牌的发送速率 R_t 和漏桶的容量 C_t；之后流控实体将以设定的速率增加令牌的个数，当令牌个数超过漏桶的容量后，漏桶内的令牌个数不再增加(漏桶满)。当被控流报文到达时，流控实体将根据漏桶中的令牌个数和报文长度进行决策：如果漏桶中令牌充足则消耗令牌并允许报文通过，反之则丢弃报文。

简单归纳基于漏桶机制的带宽管理机制如下：网络管理实体将根据用户所需带宽情况设定各个受控流的控制参数(R_t 和 C_t)，并通过扩展的 MAC 控制帧对各个流控实体设置，此后流控实体将根据该配置进行带宽控制。在具体的带宽管理过程中，系统首先根据 MAC 地址、报文优先级等信息对到达的报文进行分类，之后将各个流送至相应的漏桶控制实体中完成具体的控制步骤。这样，网络管理实体就可以按照网络的实际情况和用户的需求在链路层对带宽进行分配和管理。目前，带宽分配机制在驻地网中的应用主要集中在用户隔离

张轶博　博士研究生，主要从事 TCP 性能评估与改善方面的研究。王海霞　硕士研究生，主要从事 TCP 性能评估方面的研究。雷振明　教授，主要从事 IP/ATM 宽带网络方面的研究。

和站点访问带宽控制两个方面，前者的目的是保障带宽占用的公平性，而后者则是为了管理驻地网用户访问特定服务器的总流量。

3 研究模型的建立

为了评估带宽管理机制对于长时 TCP 性能的影响，本节对于该机制进行抽象建模。

我们的目的是评估基于漏桶方式的带宽管理机制对于长时 TCP 传输的影响，因此，在讨论中我们将忽略除漏桶管理实体以外的因素对 TCP 传输的影响，首先给出以下前提。

1. 考虑到驻地网的实际情况，分配给用户或目标站点的带宽将远小于宽带接入网络的出口带宽。

2. 高层应用能够产生足够的数据使得 TCP 发送端总处于发送状态，且每个轮次的报文集中在一起发送[6,7]，需要注意的是此时发送端以最大报文长度 SMSS 发送报文[6]。

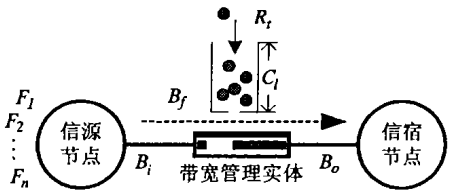

图 1 抽象的带宽管理模型

我们的研究模型如图 1 所示：信源通过带宽管理实体连接到信宿；带宽管理实体为信源节点分配带宽 B_f，并根据此带宽配置漏桶参数(R_t,C_t)；来自信源节点的多个长时 TCP 流(F_1 到 F_n)将共享此带宽资源。模型中 B_i 和 B_o 分别为信源、信宿与管理实体之间的物理链路带宽，有 $B_i>>B_f$，$B_o>>B_f$，因而 TCP 报文丢弃仅发生在带宽管理实体中。

接下来说明带宽管理机制中长时 TCP 流的环回时间(RTT)的特点。对目标站点进行带宽管理的情形而言，由于同一接入网内用户访问同一目标站点时，报文经相同的路由传播，因而 TCP 流的 RTT 时间近乎相等。对于用户隔离的情形而言，考虑到用户使用诸如 FlashGet、NetAnt 之类的多线程下载软件依次下载大数据文件的习惯，用户会同时开启多个长时 TCP 连接，而这些 TCP 连接具有相同目的地址，进而其 RTT 时间近乎相同。综上，为了简化问题，论文仅讨论长时 TCP 具有相近的环回时间的情形，并在此情形下假定多个长时 TCP 的环回时间相等。

基于上述讨论，我们做出以下分析：

当管理实体所分配的带宽被 N 个 RTT 相同的 TCP 流共享时，观察者在一个 RTT 时长内可以发现属于全部 N 个流的数据报文。又由前提 1 和 2，在物理传输速率远高于分配给用户的带宽的情况下，每个流的同轮次报文集中在一起发送，由于这些报文的到达间隔将非常之短，我们不妨将同一 TCP 的同轮次报文的到达事件视为在同一时刻发生，并简称其为轮次到达。

注意到 N 个 TCP 流的轮次到达相互独立且均匀分布在一个环回时间内，可近似认为两个相邻且不属于同一 TCP 流的轮次到达之间的间隔(T_i)服从负指数分布[10]。

考虑到未发生报文丢弃时，TCP 的发送速率是递增的，最终必然超过设定的发送速率。这样，对于漏桶式流控算法而言，递增的 TCP 流量将逐渐耗尽漏桶中累计的令牌。又考虑到 TCP 同一轮次的报文被集中发送，在如此短的时间内，流控实体不会发放大量新的令牌，因此对该 TCP 系统将在令牌耗尽后丢弃该轮次剩余的所有报文，事实上这是一种类似 Drop Tail 的丢弃行为。

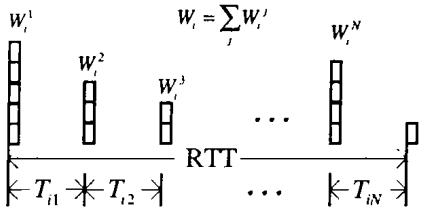

图 2 一个发送轮次中 N 个 TCP 流地报文到达示意

我们所建立的分析模型采用 Drop Tail 方式丢弃报文。图 2 是该模型一个发送轮次中报文到达的分布示意。其中，随机变量 T_i 代表轮次到达时间间隔，服从负指数分布；随机变量 W_i^j 为第 j 个 TCP 流在第 i 个轮次的发送窗口，W_i 为第 i 个轮次所有 TCP 流发送窗口的总和。

首先给出随机变量 T_i 的分布密度。由于轮次到达间隔小于环回时间 RTT 且服从负指数分布，可得轮次到达间隔的概密函数如式(3.1)，由于在环回时间 RTT 内将发生 N 次到达，易知到达强度 $\lambda=N/RTT$。

$$f(t)=\frac{\lambda e^{-\lambda\cdot t}}{1-e^{-\lambda\cdot RTT}} \quad t\in(0,RTT) \tag{3.1}$$

接下来讨论 TCP 报文丢失的概率。在多个 TCP 的总流量大于限定带宽后，漏桶中积累的令牌最终将被耗尽，此时，报文的通过量取决于该报文所在轮次到达与前一轮次到达之间的间隔时间内所能够积累的令牌量。易知，ΔT 时间内累积的令牌所允许通过的流量为 $\Delta T\cdot B_f$，若此后一个轮次到达的数据量多于 $\Delta T\cdot B_f$，则必然有报文被丢弃。由此，某 TCP 流某轮次的拥塞窗口为 W 时，发生报文丢弃的概率如式(3.2)，其中 T 为常量，代表按照给定带宽发送一个报文所需的时间，有 $T=SMSS/B_f$。

$$F_L(W)=P\{t<WT\}=\int_0^{WT}f(t)\mathrm{d}t=\frac{1-e^{-\lambda WT}}{1-e^{-\lambda RTT}} \tag{3.2}$$

4 TCP 性能的建模分析

本节基于上节的内容对 Newreno TCP 的吞吐量进行分析。

4.1 吞吐量计算的基本思路

为了得出多个 TCP 流经过总带宽限制为 B_f 的漏桶流控实体时的吞吐量，我们采用以下步骤分析。

首先，我们把相邻的两次丢弃之间的间隔定义为一个拥塞周期，这样一个拥塞周期内将包含多个报文发送轮次，而只有最后一个轮次发生报文丢弃。若记 Y_n 某拥塞第 n 个周期中通过漏桶的报文个数，R_n 为拥塞周期所包含的轮次个数，则整个通信过程中的平均吞吐量可以由(4.1)式给出[9]，其中 SMSS 为报文长度，RTT 为平均环回时间：

$$B=\frac{E[Y]\cdot SMSS}{E[R]\cdot RTT} \tag{4.1}$$

接下来详细讨论 $E[Y]$的求取，根据 TCP 流控算法的特点，发生报文丢弃后 TCP 流可能成为以下三种流之一：

1. 未发生报文丢弃的流，简记为 NL。

2. 有报文丢失但未引起超时重传的流，简记 LO。

3. 发生超时重传的流，简单记 TO。

我们根据上述原则将 N 个 TCP 流分为三类，并分别求

出一个拥塞周期内的三类流通过漏桶的数据量，最后将它们相加就得到 Y_n 如式(4.2)。

$$Y_n=Y_{NL}^n+Y_{TO}^n+Y_{LO}^n \tag{4.2}$$

以下说明一个拥塞周期内三类 TCP 通过数据量的计算。以 TO 流(超时流)为例介绍具体的计算方法：

1. 计算丢弃发生后，TO 流个数的期望。

2. 给出一个拥塞周期内单个 TO 流通过漏桶的数据量。

3. TO 流的总通过数据量等于 TO 个数与单个 TO 流通过数据量的乘积。

本节的后续部分将根据此思路进行吞吐量的推导。

4.2 三类流通过数据量的数学期望

本小节首先计算三类流个数的期望，然后计算单个流的通过数据量，最后得出总的拥塞周期通过数据量的期望。

首先计算流个数的期望。对于 Newreno TCP 而言，同一轮次发生一个或多个报文丢弃时，拥塞窗口将减半；当同一轮次中通过流控实体的报文小于 3 时，将触发超时重传[8,9]。由上述讨论，并结合式(3.1)、(3.2)，若拥塞发生时某 TCP 流的拥塞窗口为 W，则拥塞后该 TCP 流成为 TO 流、LO 流和 NL 流的概率如下式，其中 $K=\min(3,W)$：

$$F_{TO}(W)=P\{0<t<KT\}=\frac{1-e^{-\lambda KT}}{1-e^{-\lambda RTT}} \tag{4.3}$$

$$F_{LO}(W)=P\{KT<t<WT\}=\frac{e^{-\lambda KT}-e^{-\lambda WT}}{1-e^{-\lambda RTT}} \tag{4.4}$$

$$F_{NL}(W)=1-F_{TO}(W)-F_{LO}(W) \tag{4.5}$$

先计算 TO 流(超时流)的个数，考虑到 λ_{WT} 的分布区间宽度有限，负指数函数在较小的区间内可以近似成为线性函数(可在此区间中点展开为泰勒级数)，因而有：

$$P_{TO}=E[F_{TO}(W_{TO})]\approx F_{TO}(E[W_{TO}]) \tag{4.6}$$

注意到 TCP 流控算法自身具有的公平性，知发生报文丢弃后，所有 N 个 TCP 流中 m 个流发生超时的概率为二项分布，即：

$$P\{k=m\}=\binom{m}{N}P_{TO}^m(1-P_{TO})^{N-m} \tag{4.7}$$

由此求得超时流的个数期望 N_{TO} 如式(4.8)，同理可得 N_{LO} 和 N_{NL}。

$$N_{TO}=E[N_{TO}(k)]=NP_{TO} \tag{4.8}$$

$$N_{LO}=E[N_{LO}(k)]=NP_{LO} \tag{4.9}$$

$$N_{NL}=N(1-P_{TO}-P_{LO}) \tag{4.10}$$

接下来介绍单个流的通过数据量。

计算单个 NL 流的通过数据量。由于报文未被丢弃，NL 流的带宽将线性增长，则一个拥塞周期通过的报文数量 Y_{NL}^n 如式，其中 W_{NL}^{n-1} 和 W_{NL}^n 分别代表第 $n-1$ 个和第 n 个拥塞周期结束时 NL 流的发送窗口：

$$Y_{NL}^n=\frac{1}{2}(W_{NL}^{n-1}+1+W_{NL}^n)\cdot R_n \tag{4.11}$$

计算 TO 流的数据量。经历超时的 New-reno TCP 流在丢弃发生后的 α 个轮次内不发送报文，接着经过慢启动过程其窗口恢复到 $W_{TO}^{n-1}/2$，此后 TO 流线性增长，直至报文丢弃发生。由于超时时长可能大于拥塞周期时长，不妨令 $q=\lfloor \alpha/E[R]\rfloor+1$，则必有式(4.12)。因此，TO 流的通过数据量如式(4.13)。

$$(q-1)E[R]\leqslant\alpha\leqslant qE[R] \tag{4.12}$$

$$Y_{TO}^n=\frac{1}{q}\Big(\sum_{i=0}^{\log_2 W_{TO}^{n-1}/2-1}2^i+\frac{1}{2}\Big(\frac{W_{TO}^{n-1}}{2}+W_{TO}^{n+q-1}\Big)(R_{n+q-1}-(\alpha-(q-1)E[R])-\log_2\frac{W_{TO}^{n-1}}{2}\Big)\approx\frac{1}{2}q\Big(\frac{W_{TO}^{n-1}}{2}+W_{TO}^{n+q-1}\Big)(R_{n+q-1}-\alpha+(q-1)E[R]) \tag{4.13}$$

计算 LO 流的数据量。发生多个报文丢弃后，LO 流进入快速恢复此时发送窗口为 $W_{LO}^{n-1}/2-L^{n-1}$，其中 L^{n-1} 为 LO 流上一拥塞周期结束时丢弃的报文数。此后，LO 流将线性恢复至 $W_{LO}^{n-1}/2$ 并进入拥塞避免状态继续线性增长至 W_{LO}^n[7]。

$$Y_{LO}^n=\frac{1}{2}\Big(\frac{1}{2}W_{LO}^{n-1}-L^{n-1}+W_{LO}^n\Big)\cdot R_n \tag{4.14}$$

最后，考虑到发送窗口与拥塞周期长度无关[9]，给出通过数据量的期望如(4.15)：

$$\begin{aligned}E[Y_{NL}]&=N_{NL}\Big(E[W_{NL}]+\frac{1}{2}\Big)\cdot E[R]\\E[Y_{TO}]&=\frac{3N_{TO}}{4q}E[W_{TO}](qE[R]-\alpha)\\E[Y_{LO}]&=\frac{N_{LO}}{4}(3E[W_{LO}]-2E[L])\cdot E[R]\\E[Y]&=E[Y_{NL}]+E[Y_{TO}]+E[Y_{LO}]\end{aligned} \tag{4.15}$$

4.3 拥塞周期长度的数学期望

引入几个概念，我们称拥塞周期开始时，所有 TCP 流发送窗口的总和为拥塞恢复窗口，记为 W_r^n；并记拥塞周期结束时，所有 TCP 流发送窗口的总和为总拥塞窗口，并记其为 W_b^n。由 NewReno TCP 特点知 W_b^{n-1} 与 W_r^n 有以下关系(拥塞恢复窗口等于总拥塞窗口减去报文丢弃引起的发送窗口减小)：

$$W_r^n=W_b^{n-1}-\sum_{i=1}^{N_{TO}}W_{TO_i}^n-\sum_{i=1}^{N_{LO}}\Big(\frac{1}{2}W_{LO_i}^n+L_i^{n-1}\Big) \tag{4.16}$$

对于式(4.16)两端取期望，得式(4.17)：

$$E[W_r]=E[W_b]-N_{TO}E[W_{TO}]-N_{LO}(E[W_{LO}]/2+E[L]) \tag{4.17}$$

由于 W_r^n 经 R_n 个轮次增长后达到 W_b^n，考虑三类 TCP 流发送窗口的增长特性，可以得出以下讨论：LO 流和 NL 流的发送窗口在每一个轮次中都增长 1，而 TO 流在超时结束后，发送窗口指数阶增长至拥塞发生前的一半，并在以后的轮次里每个轮次将增长 1(需要说明的是这一过程需要 q 个拥塞周期)，再注意到 $N=N_{TO}+N_{NL}+N_{LO}$，有：

$$\begin{aligned}E[W_b]&=E[W_r]+(N_{NL}+N_{LO})\cdot E[R]+\frac{N_{TO}}{q}\cdot\Big(qE[R]-\alpha-\log_2\frac{E[W_{TO}]}{2}+\frac{E[W_{TO}]}{2}\Big)=E[W_r]+N\cdot E[R]+\frac{N_{TO}}{2q}(E[W_{TO}]-2\alpha-o(E[W_{TO}]))\end{aligned} \tag{4.18}$$

将式(4.17)代入(4.18)，可得到拥塞周期的数学期望，需要说明的是式(4.19)的进一步求解见 4.5 节。

$$E[R]\approx\frac{N_{TO}((2q-1)E[W_{TO}]+2\alpha)+qN_{LO}(E[W_{LO}]+2E[L])}{2qN} \tag{4.19}$$

4.4 吞吐量的求取

由式(4.1)、(4.15)和(4.19)知，总吞吐量 B 可以由 $E[W_{TO}]$、$E[W_{LO}]$、$E[W_{NL}]$ 以及 $E[L]$ 表示。以下求取这些期望。

首先给出总拥塞窗口的期望 $E[W_b]$。由于每个拥塞周期的总发送窗口 W_b^n 的大小主要取决于漏桶中积累令牌的数量，对于长时 TCP 的情形而言，考虑到漏桶的有限容量和基本稳定的积累时间，一个拥塞周期中所能够积累的令牌量基本稳定，因此有：

$$W_b^{n-1}\approx W_b^n\approx E[W_b] \tag{4.20}$$

又注意到 TCP 算法自身具有的公平性，则每个流在拥塞发生时的发送窗口的期望相等近似为 $E[Wb]/N$，因此拥塞周期中的三种流的发送窗口期望为：

$$E[W_{TO}]=E[W_{LO}]=\frac{E[W_b]}{N} \tag{4.21}$$

$$E[W_{NL}]=\frac{E[W_b]}{N}+\frac{E[R]}{2} \tag{4.22}$$

再计算 $N_{LO}E[L]$，由于 $N_{LO}E[L]$ 表示了拥塞中 LO 流丢弃的报文数的期望，如近似认为 TO 流丢弃的报文数为 $N_{TO}(E[W_{TO}]-3)$，并注意到拥塞中丢弃的报文为超过分配带宽 B_f 的部分，则可由式(4.23)近似确定 LO 流丢弃的报文。其中 $M=B_fRTT/SMSS$ 代表流控实体一个轮次中允许通过的报文数目。

$$N_{LO}E[L]\approx E[W_b]-M-N_{TO}(E[W_{TO}]-3) \tag{4.23}$$

利用式(4.21)对 P_{TO}，P_{LO}，P_{NL}进行化简，并注意到 $T=SMSS/B_f=RTT/M$，$\lambda=N/RTT$ 以及并记突发因子记 $b=E[W_b]/M$，得：

$$P_{TO}\approx F_{TO}(E[W_{TO}])=\frac{1-e^{-\min(\frac{3N}{M},b)}}{1-e^{-N}} \tag{4.24}$$

$$P_{LO}\approx F_{LO}(E[W_{LO}])=\frac{e^{-\min(\frac{3N}{M},b)}-e^{-b}}{1-e^{-N}} \tag{4.25}$$

$$P_{NL}=1-P_{TO}-P_{LO}\approx\frac{e^{-b}-e^{-N}}{1-e^{-N}} \tag{4.26}$$

将式(4.15)代入式(4.1)并根据(4.22)(4.21)(4.23)，化简得：

$$B\approx\frac{SMSS}{RTT}\left(\frac{E[W_b](1+2P_{TO}+P_{NL})+2M}{4}+\frac{N_{NL}(E[R]+1)-3N_{TO}}{2}-\frac{3P_{TO}\alpha E[W_b]}{4qE[R]}\right) \tag{4.27}$$

4.5 进一步的讨论

为了便于评价，用 B_f 对总吞吐量进行归一化，得到吞吐率 ρ 如(4.28)式，其中突发因子 $b=E[Wb]/M$。

$$\rho=B/B_f\approx\frac{b(1+2P_{LO}+P_{NL})+2}{4}+\frac{N_{NL}(E[R]+1)-3N_{TO}}{2M}-\frac{3P_{TO}ab}{4qE[R]} \tag{4.28}$$

由于根据式(4.19)以及 $q=\lfloor\alpha/E[R]\rfloor+1$ 求取 $E[R]$ 的解析解较为复杂，这里根据应用环境的实际情况予以简化求解。

对于同时存在的长时 TCP 较少($M>>N$)，超时时长小于拥塞周期时长的情形而言，$q=1$。利用(4.21)(4.23)进行化简，得 $E[R]$、ρ 如下式：

$$E[R]=(\alpha+3)P_{TO}+\frac{(2+P_{LO}-P_{TO})Mb-2M}{2N} \tag{4.29}$$

$$\rho\approx\frac{(1+2P_{TO}+P_{NL})b+2}{4}-\frac{3P_{TO}ab}{4E[R]}+\frac{NP_{NL}(E[R]+1)-3NP_{TO}}{2M} \tag{4.30}$$

对于 TCP 连接数量较多，超时时长大于拥塞周期时长的情形，利用 $q\approx\alpha/E[R]$ 以及(4.21)(4.23)对(4.19)中的 $E[R]$ 进行求解，并利用 $qE[R]\approx\alpha+E[R]$ 化简(4.28)，得到最终的分析结果。

$$E[R]=\frac{\alpha M[(2+P_{LO})b-2]+6\alpha NP_{TO}}{2N(1-P_{TO})\alpha+P_{TO}bM} \tag{4.31}$$

$$\rho\approx\frac{b(1+2P_{TO}+P_{NL})+2}{4}+\frac{N_{NL}(E[R]+1)-3N_{TO}}{2M}-\frac{3P_{TO}ab}{4(\alpha+E[R])} \tag{4.32}$$

5 仿真结果

为了验证论文的结论，我们使用 NS-2(版本 2.1b8)进行了仿真。仿真拓扑如图 3 所示，我们在信源和信宿节点上依次部署 2—10 个长时 TCP 流，总的带宽限制为 1M。

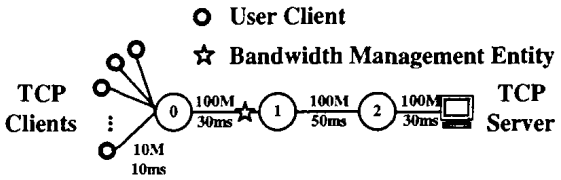

图 3 仿真拓扑示意

表 1 仿真测量参数一览

TCP 流数 N	2	3	4	5	6
突发因子 b	1.1214	1.2028	1.2690	1.3767	1.4972
超时轮次 α	4.1377	3.9973	3.8771	3.6056	3.5007
吞吐量 ρ_1	0.8332	0.8783	0.8955	0.9008	0.9093
TCP 流数 N	7	8	9	10	
突发因子 b	1.6056	1.6826	1.7997	1.8505	
超时轮次 α	3.3276	3.2207	3.1000	2.7996	
吞吐量 ρ_1	0.9541	0.9558	0.9797	0.9877	

首先我们对仿真数据进行分析，从中测量出突发因子 b、超时轮次 α 和仿真吞吐率 ρ_1，结果如表 1 所示。

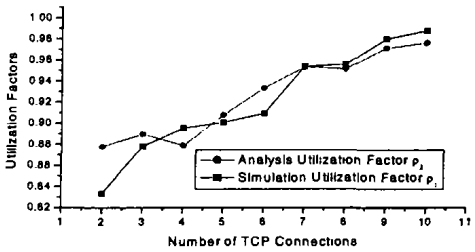

图 4 仿真吞吐率 ρ_1 与模型吞吐率 ρ_2 的比较

基于测量到的数据，我们将 b 和 α 代入模型，获得模型吞吐率 ρ_2。图 4 是仿真吞吐率 ρ_1 与模型吞吐率 ρ_2 的比较。比较结果说明论文所建立的模型是有效的。

结束语 论文通过建模的方法对于以太网带宽管理机制下 Newreno TCP 的性能进行了研究。以太网带宽管理机制采用漏桶式流控算法，该算法对于长时 TCP 的控制方式类似 Drop Tail。模型分析和仿真结果都说明，对于多个长时 TCP 共享带宽的情形而言，在 TCP 流数量较少的情况下，该管理机制对于 TCP 的性能具有一定的影响(吞吐率较低)；在数据流数量较多的情况下，系统可以达到较高的吞吐率。换言之，对于少数长时 TCP 流共享带宽的情形，以太网带宽管理机制的性能不尽如人意，有待改进。

参 考 文 献

1 IEEE 802.3x. Specification for 802.3 Full Duplex Operation. IEEE Standard 802.3,1998

2 Noureddine W, Tobagi F. Selective Back-Pressure in Switched Ethernet LANs. GLOBECOM,1999

3 Wechta J, Fricker M, Halsall F. Hop-by-Hop flow Control as a Method to Improve QoS in 802.3 LANs. IWQoS,1999

4 Yoshigoe K, Christensen K. RATE Control for Bandwidth Allocated Services in IEEE 802.3 Ethernet. LCN,2001

5 Marina F, Ken K, David M, Claffy K. Longitudinal study of Internet traffic from 1998-2003. the Cooperative Association for Internet Data Analysis (CAIDA). http://www.caida.org/outreach/papers/2003/nlanr/nlanr_overview.pdf

6 Allman M, Paxson V, Stevens W. TCP Congestion Control. RFC 2581, April 1999

7 Fall K, Floyd S. Simulation-based Comparisons of Tahoe, Reno and SACK TCP. Computer Communication Review, July 1996.

ftp://ftp.ee.lbl.gov/ papers/sacks.ps.Z
8 Floyd S,Henderson T. The NewReno Modification to TCP's Fast Recovery Algorithm. RFC 2582,April 1999
9 Jitendra P,Victor F,Don T,Jim K. Modeling TCP Throughput: A Simple Model and its Empirical Validation. In: Proc. of the ACM SIGCOMM '98 conf. on Applications,Technologies,Architectures,and Protocols for Computer Communication
10 陆传赉.工程系统中的随机过程.电子工业出版社,2000.72～73

(上接第 36 页)

据。若使用 9600bps 的通信速率,以每个字节占用 10bit 计算,则该缓冲区可以存储约 8.53s 的连续串口数据。若串口速率为 115200bps 则能存储约 711ms 的连续数据;从另一个方面说,如果我们定义向网络发送的最大业务数据报文长度为 1k 字节,那么该缓冲环可以保留 8 帧完整的数据。当缓冲区中的数据满足条件时,处理器才进行相应的操作流程。这使得处理器在处理串口数据时能享有更大的弹性,并且将运算资源用于其它的工作。

4.3 SDTP 协议

图 6 描述了在本 DAS 服务器中实现的 SDTP(Serial Data Transmit Protocol)协议格式。所有针对数据传输的报文均按照该协议结构进行封装。SDTP 的报文以 1 字节长度的类型字段开始,包括了 2 字节的序号字段,2 字节的长度字段和不定长度的数据字段。

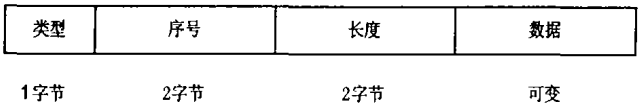

图 6 SDTP 协议结构

SDTP 序号字段与长度字段的使用能保证数据的有序传输,以及通过对称重传机制来防止报文的丢失。

5 性能分析

我们主要分析 DAS 系统对串口数据的处理能力。为了简化计算做如下假设:设串口的波特率为 B。传输每个字节需要 10bit,则通过串口每秒能传输 $B/10$ 字节。根据本文设置的串口传输规则,数据报文最大长度设定为 L 字节,等待超时值设定为 D 秒。那么,系统启动传输的两个条件的最短延迟时间分别为:

条件 1:收到 L 字节时发送,延迟为:$10*L/B$;

条件 2:超时 D 秒发送,延迟为:$10/B+D$。

系统的最短延迟时间为上述两个参数的最小值,即 $\min(10*L/B,10/B+D)$。

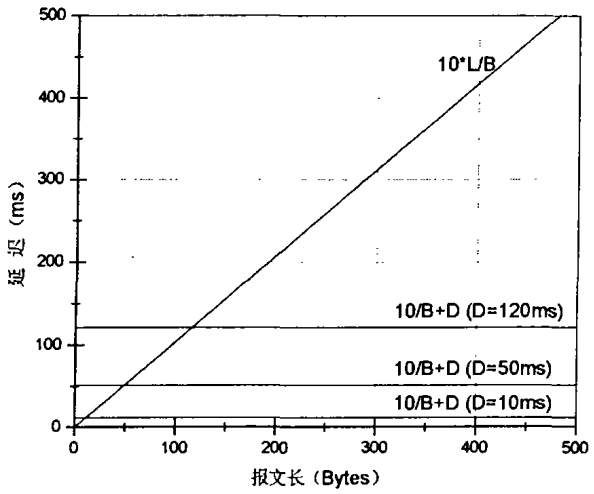

图 7 不同参数下的系统最小延迟时间

考虑最坏情况下的延迟时间:系统在收到第一个字符后,每次在超时前的一瞬间收到了串口来的下一个字符,那么整个数据帧的延迟为$(D-10/B)*(L-1)$。

图 7 描述了系统在 9600bps 波特率下,不同参数设置对系统最小延迟的影响。从图 7 中的曲线可以看出:对于条件 1,系统的最小延迟是随报文长度增加而线性增加的;对于条件 2,系统的最小延迟为固定值。由于系统总最小延迟为二者的最小值,因此可以在波特率一定的情况下通过确定 D 的值来确定系统的最小延迟不大于某个值。

同时,我们针对 Moxa 公司的 NPort Server 产品和 Digi International 公司的 PortServer TS 产品,以及本文的 DAS 服务器进行了性能比较与测试,结果如表 2 所示。

表 2 三种产品的比较

	DAS	Moxa	Digi
处理器	8 位 AVR	32 位 ARM	16 位 x86
网络延迟	820us	270us	530us
串口速率	2400-115200bps	300-1.5Mbps	300-921Kbps
网络接口	10M	10/100M	10M
价格	¥100	¥2400	¥1000

通过比较,可以看出,DAS 与别的产品相比虽然在功能和性能的绝对值上都不及同类产品,但是,其所提供的性能已经能够满足很多现场设备的运行要求,并且在价格因素的比较中具有很大的优势,能够占有一定的市场份额。

结论 DAS 通用 Internet 设备接入服务器无需对现有设备做任何修改,不依赖于串行设备自身数据帧格式的结构,"透明"地实现了面向流传输方式的串行接口和面向包传输方式的网络接口之间的数据转换。该研究成果已应用在东大新业的嵌入式 Internet 系列产品中,部分产品获国家实用新型专利,并通过了辽宁省科委组织的科学技术成果鉴定。

参考文献

1 Lee B H. Embedded Internet System: Poised for Takeoff[J]. IEEE Internet Computing,May 1998.29
2 Agranat I. Embedded web servers in network devices[J]. Communication Systems Design, March 1998. 30～36
3 Wolf W. Hardware-software co-design of embedded systems. Proc. of the IEEE,1994,82(7): 967～989
4 Robert F. Embedded Internet systems come home[J]. IEEE Internet Computing,2001,40(14):52～53
5 Jacek W. Embedded Internet technology in process control devices [J]. IEEE,2000,34(3):301～308
6 赵海,陈飞鸣. Embedded Internet 的体系结构及其 ONDC 模型的实现[J]. 东北大学学报,1999,20(3):257
7 张德干,郝先臣,赵海. 一种基于 EI 技术的 EIDI 模型的研究及实现[J]. 电子学报,2002,30(5):749
8 金欢,阮冠春,徐凌宇,赵海. 基于嵌入式 Internet 技术的 Webit 体系结构研究与实现[J]. 控制与决策,2002,17(5):541