基于 Gilbert 丢包机制的 TCP 吞吐量模型

曾 彬[1],张大方[2],黎文伟[2],谢高岗[3]
(1.湖南大学计算机与通信学院,湖南长沙 410082; 2.湖南大学软件学院,湖南长沙 410082;
3.中国科学院计算技术研究所,北京 100190)

摘 要: 丢包机制是推导 TCP 吞吐量模型的关键,直接影响模型的准确性.本文利用四状态 Gilbert 丢包机制来描述端到端路径上的丢包行为,对 TCP 的拥塞控制过程进行建模,在此基础上提出了一种更精确的 TCP 吞吐量模型.实验表明,改进的模型能较好的与实际值相拟合,可以更精确地预测实际 TCP 数据流的吞吐量性能.

关键词: 网络测量; TCP 吞吐量; 拥塞控制; Gilbert 丢包机制
中图分类号: TP393 **文献标识码:** A **文章编号:** 0372-2112 (2009) 08-1728-05

A TCP Throughput Model Based on Gilbert Packet Loss Pattern

ZENG Bin[1], ZHANG Da-fang[2], LI Wen-wei[2], XIE Gao-gang[3]
(1. *School of Computer and Communication, Hunan University, Changsha, Hunan* 410082, *China*;
2. *Software School, Hunan University, Changsha, Hunan* 410082, *China*;
3. *Next Generation Internet Research Center, Institute of Computing Technology, Chinese Academy of Sciences, Beijing* 100190, *China*)

Abstract: TCP throughput is one of the key concepts of computer network, and its estimation is very important for many research and application domains. Packet loss pattern is crucial to deduce a TCP throughput model, and can affect model's precision directly. Therefore, a new TCP throughput model GT(Gilbert Throughput) is presented by modeling a TCP congestion control mechanism based on Gilbert four-state model that is used to represent the packet lost behavior of end-to-end internet path. Experiment results show that GT matches the results better than Goyal model, and can predict the throughput of TCP flows more precisely in actual network.

Key words: network measurement; TCP throughput; congestion control; gilbert packet loss pattern

1 引言

针对 TCP 性能的研究一直是 Internet 网络协议研究的重要方面,而 TCP 吞吐量是衡量 TCP 性能的一个重要指标.目前,已经提出了多种 TCP 吞吐量的理论分析模型[1~8].其中,最基本的模型是 Mathis 模型[1]和 Padhye 模型[2,3],其他几个模型基本上都是对这两个模型的改进.这些模型都是对 TCP 的动态行为进行一定的假设,然后根据丢包情况对拥塞控制的各个阶段进行建模,推导出 TCP 吞吐量的表达式[4].

由此可见,丢包机制是推导 TCP 吞吐量模型的关键,直接影响模型的准确性.Mathis 模型采用周期丢包机制,这种丢包模式假设链路的丢包率为 p,则每连续传输 $1/p$ 个报文后,将丢失一个报文;Padhye 模型采用的是简单丢包机制,即假设在一轮报文发送中,如果在发送第 k 个报文时发生了本轮发送过程的第一个丢包,则位于发送窗口中的其他还没有被发送的报文都将丢失;Goyal 模型[5]在 Padhye 模型基础上,采用了 Bernoulli 模型来描述丢包过程,考虑了超时重传概率的计算.Goyal 模型是当前应用比较广泛的一种,在许多 TCP 友好传输算法、媒体流自适应传输等方面作为一个基础性理论公式得到大量的应用与改进[6~8].但这些改进主要是针对提高拥塞避免阶段报文段计算的准确性,丢包机制采用的仍是 Bernoulli 模型.

现有的推导分析模型中采用的丢包机制并不很准确,特别是在 MANET、WSN 等低带宽、高误码率的无线环境下,与实际值误差较大;而且有些模型的参数难以在实际的网络测量中进行准确测量[9,10].因此,现有 TCP 吞吐量模型采用的丢包机制已经成为制约提高吞吐量模型精确度的重要原因.而文献[11,12]的实验数据表明:40% 的丢包行为符合 Bernoulli 丢包机制,89% 的丢包情况符合两个状态的 Gilbert 模型,而四状态

收稿日期:2008-07-18;修回日期:2009-03-29
基金项目:国家 973 重点基础研究发展计划(No.2007CB310702);国家自然科学基金重大研究计划(No.90718008);国家自然科学基金(No.60673155,No.60703097)

Gilbert 模型的丢包情况符合度更高.因此,本文采用四状态 Gilbert 模型来描述测量过程中端到端路径上的丢包行为,对 TCP 的拥塞控制过程进行建模,重新推导出一个基于四状态 Gilbert 丢包机制的 TCP 吞吐量模型.最后给出了实验结果和本文的结论.

2 四状态 Gilbert 丢包机制

有两种类型的 Gilbert 模型:连续成功收到情形,连续丢失情形.本文用到的模型是连续成功收到的类型,图 1 是它的状态变迁图.

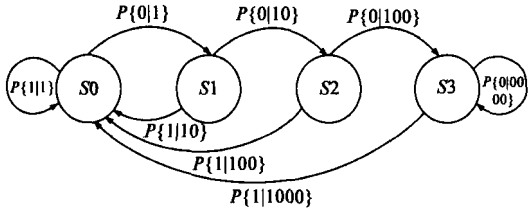


图1 四状态的Gilbert丢包机制

其中 $S0, S1, S2, S3$ 表示数据包的状态,1 表示包丢失,0 表示包没有丢失,$P\{1|1\}$表示第一个包丢失下一个包也丢失的情况下的概率,$P\{1|0\}$表示第一个包收到而下一个包丢失的情况下的概率.相应的 $P\{1|10\}$表示第二个包收到而第三个包丢失情况下的概率,其余与此类似.

假设网络中数据包到达路由是服从 Poisson 分布的,μ 代表该路由的平均数据处理能力,$b(i)$是在一个时间间隔内 i 个数据包被系统处理的概率,并且在这个时间间隔开始前至少有 i 个数据包在路由中,K 是最大队列值,$0 \leqslant i \leqslant K$.

$$b(i) = \int_0^{\infty} \frac{e^{-\mu}(\mu t)^i}{i!} e^{-t} dt \ , 0 \leqslant i \leqslant K \tag{1}$$

在这种情形下,状态转换概率计算如下:

$$p\{1|1\} = b(0) \tag{2}$$

因为 $P\{0|1\}$与 $P\{1|1\}$互为逆,所以

$$p\{0|1\} = 1 - p\{1|1\} = 1 - b(0) \tag{3}$$

类似的,下面的概率值也是成对出现的.

$$p\{1|10\} = \frac{p\{101\}}{p\{10\}} = \frac{b(1)\cdot b(0)}{1 - b(0)} \tag{4}$$

$$p\{0|10\} = 1 - p\{1|10\} = 1 - \frac{b(1)\cdot b(0)}{1 - b(0)} \tag{5}$$

而在实际应用中,则不需要关心路由的具体状态,只是用测量得到的数据包相关值计算出 Gilbert 模型特定概率值即可.如设从 1 状态到 0 状态变换的概率为 p,0 状态到 1 状态变换的概率为 q,则有:

$$p\{0|1\} = p = \frac{\text{发生丢包事件的次数}}{\text{1 状态出现的次数}} \tag{6}$$

$$p\{1|0\} = q = \frac{\text{发生丢包事件的次数}}{\text{0 状态出现的次数}} \tag{7}$$

对于连续丢包的情形都可以通过测量中得到的丢包统计数据计算出,在此不再赘述.

3 TCP 吞吐量模型推导

定义 1 TCP 吞吐量是一个描述端到端路径的性能参数.IPPM 在 RFC3148 中定义块传输能力 BTC 为 $8\cdot B/\Delta T$,其中 B 为在 ΔT 时间段内节点 s 成功地向节点 t 传输的字节数,不包括包头,并且重传的数据包只应该计算一次.

根据定义 1 的说明,达到流控稳定状态的 TCP 吞吐量的计算公式可以表示为:

$$B = \lim_{t\to\infty} B_t = \lim_{t\to\infty} \frac{N_t \cdot MSS}{t} = \frac{N \cdot MSS}{T} \tag{8}$$

其中 N 为接收方成功接收的报文数的期望,T 为接收这些报文所用的时间的期望,MSS 是报文段大小值[7].

根据 TCP 拥塞控制机制,整个 TCP 传输过程可以模型化为几个阶段:慢启动阶段、拥塞避免阶段、超时重传阶段、快速重传和恢复阶段.整个传输过程就是在这几个状态之间不停的进行变换,从而完成对传输过程中出现的各种情况的响应,如图 2 所示.根据是否丢包,将模型推导分为两个部分:无丢包时,慢启动过程一直持续到传输结束,此时要考虑最大拥塞控制窗口的限制;有丢包时,慢启动结束,进入拥塞避免阶段.根据丢包情形的不同,传输过程有可能进入快速重传阶段,也有可能进入超时重传阶段.因此,N 和 T 又可以如下表示:

$$\begin{aligned} N &= N_{ss} + N_{cc} + N_{TO}\cdot P_{TO} + N_{FR}\cdot P_{FR} \\ T &= T_{ss} + T_{cc} + T_{TO}\cdot P_{TO} + T_{FR}\cdot P_{FR} \end{aligned} \tag{9}$$

其中,N_{ss}, T_{ss}分别为慢启动阶段接收方接收的报文数和所用时间的数学期望;N_{cc}, T_{cc}分别为拥塞避免阶段接收的报文数和所用时间的数学期望;N_{TO}, T_{TO}分别为超时重传阶段接收的报文数和所用时间的数学期望;N_{FR}, T_{FR}分别为快速重传和恢复阶段接收的报文数和所用时间的数学期望;P_{TO}, P_{FR}分别为包丢失之后进入

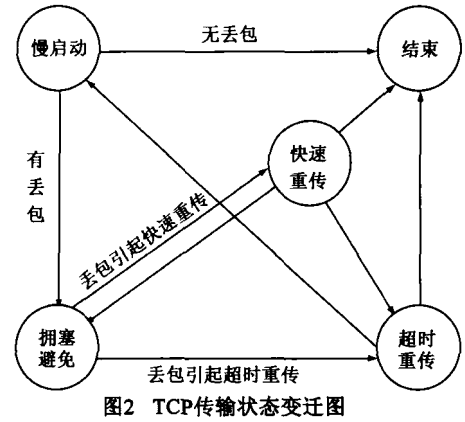


图2 TCP传输状态变迁图

超时重传阶段的概率和进入快速重传和恢复阶段的概率.显然,$P_{TO}+P_{FR}=1$.

因此,在模型的推导中,需要解决的关键问题是:(1)丢包概率的推导;(2)进入超时重传概率的推导(快速重传和恢复阶段的概率即可得).这两个概率与四状态 Gilbert 模型关系密切,是测量模型的基础.

定义 2 丢包概率 $P_{LO}(n,w)$是假设在第一个包丢失之后,在拥塞窗口大小为 w 的情况下,总共丢失 n 个包(包括第一个丢失的包)的概率.首先,我们计算 $P_{LO}(1,w)$.显然这个概率跟拥塞窗口的大小有关.当 w 为 1 时,根据 $P_{LO}(1,w)$的定义,这个包必然丢失,则 $P_{LO}(1,w)$的值为 1;当 $w>1$ 时,根据概率论的知识,这个概率可以表示为 $p(1|0)p(0|0)^{w-2}$.

$$P_{LO}(1,w)=\begin{cases}1,\text{当 } w=1 \text{ 时}\\ p(1|0)p(0|0)^{w-2},\text{otherwise}\end{cases}\tag{10}$$

而计算 $P_{LO}(2,w)$则要分四种情况来讨论.当 $w=1$ 时,不可能丢两个包,因而概率为 0;当 $w=2$ 时,这两个包都将丢失,因此此时的概率即等于在前一个包丢失的情况下后一个包继续丢失的概率,即为 $p(1|1)$;当 $w=3$ 时,要分两种情况考虑:丢失的两个包是连续丢,这种情况下,$P_{LO}(2,w)$的值应为 $p(0|11)$,如果丢失的两个包不是相邻的,则这个概率应为 $p(1|10)$,综合这两种情况,则 $W=3$ 时 $P_{LO}(2,w)$的值应为 $p(0|11)+p(1|10)$.同理,我们可以推导出当 $w=4$ 时 $P_{LO}(2,w)$的表达式为 $p(0|110)+p(0|101)+p(1|100)$.

综合上面的分析,我们可以得到如下的表达式:

$$P_{LO}(2,w)=\begin{cases}0,w=1\\ p(1|1),w=2\\ p(0|11)+p(1|10),w=3\\ p(0|110)+p(0|101)+p(1|100),w=4\\ p(0|0)^{w-4}(p(0|110)+p(0|101))+\\ \quad p(1|100)p(0|1)p(0|0)^{w-5},\text{otherwise}\end{cases}\tag{11}$$

$P_{LO}(3,w)$与 $P_{LO}(4,w)$的分析与此类似,不再赘述.

定义 3 超时重传概率 $P_{TO}(w)$表示在拥塞窗口大小为 w 时,发生丢包之后,出现超时的概率.

当发生第一次丢包之后,在丢失包之前发送的数据包都会被目的端接收到,返回相应的 ACK,于是发送端根据收到的 ACK 继续发送 $w-1$ 个包,直到检测到丢包事件.如果 $w<4$,一个包丢失,也必将导致超时发生,这是因为根据假设,包丢失之后,在收到三个重复的 ACK 之前,将没有足够的包发送,所以:

$$P_{TO}(w)=1,w<4\tag{12}$$

当丢失两个包时,则收到第一个丢失包的重复 ACK 的个数为 $w-2$.假设这 $w-2$ 个 ACK 足以触发对第一个丢失的包的快速重传,那么慢启动阈值的大小将减小到 $w/2$.由于已经收到对第一个丢失包确认的 $w-2$ 个重复的 ACK,因此拥塞窗口的大小变成了 $3w/2-2$,而在重传第一个丢失的包时,还有 w 个包没有得到确认,因此,在快速重传第一个丢失的包的时间里,还可以发送 $w/2-2$ 个包.而要使通过这 $w/2-2$ 个包能检测到第二个包丢失,必须有 $w/2-2\geqslant 3$,即 $W\geqslant 10$.因此当 $w<10$ 时,肯定只能丢失一个包,否则将不能触发第二个丢失的包的快速重传.因此,当 $w<10$ 时,出现超时重传的概率为:

$$P_{TO}(w)=1-P_{LO}(1,w),4\leqslant W<10\tag{13}$$

当丢失三个数据包时,同理可得,在拥塞窗口 $w<20$ 时,必将进入超时重传阶段.即:

$$P_{TO}(w)=1-P_{LO}(1,w)-P_{LO}(2,w),10\leqslant W<20\tag{14}$$

综合上面的分析,可以得到 $P_{TO}(w)$的表达式为:

$$P_{TO}(w)=\begin{cases}1,\text{当 } w<4 \text{ 时}\\ 1-P_{LO}(1,w),\text{当 } 4\leqslant w<10 \text{ 时}\\ 1-P_{LO}(1,w)-P_{LO}(2,w),\text{当 } 10\leqslant w<20 \text{ 时}\\ 1-P_{LO}(1,w)-P_{LO}(2,w)-P_{LO}(3,w),\text{otherwise}\end{cases}\tag{15}$$

应用丢包概率 $P_{LO}(n,w)$和超时重传概率 $P_{TO}(w)$推导结果,即可在定义 1 的 TCP 吞吐量推导中完成 TCP 吞吐量的计算.下面将通过实验来验证建立模型的有效性.

4 实验分析

4.1 模拟实验

根据模型的特点,设计了如图 3 所示的仿真环境.为提高仿真实验的真实性,让被测量传输过程与其他 TCP 传输和 UDP 数据流在瓶颈链路上竞争带宽.在图 3 中,*Sr* 代表发送结点,*Ro* 代表路由器,*De* 代表接受结点,link 代表链路,在发送结点旁边标明的是结点上的应用类型和传输类型,接受结点旁边标明的是接受类型.

根据对不同链路带宽的竞争和与不同类型传输流的竞争,模拟了在不同瓶颈带宽、路由器排队机制,背景流量等条件下的情形.为了表述的方便,对不同条件下

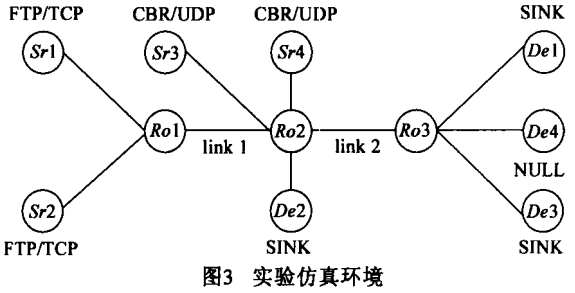


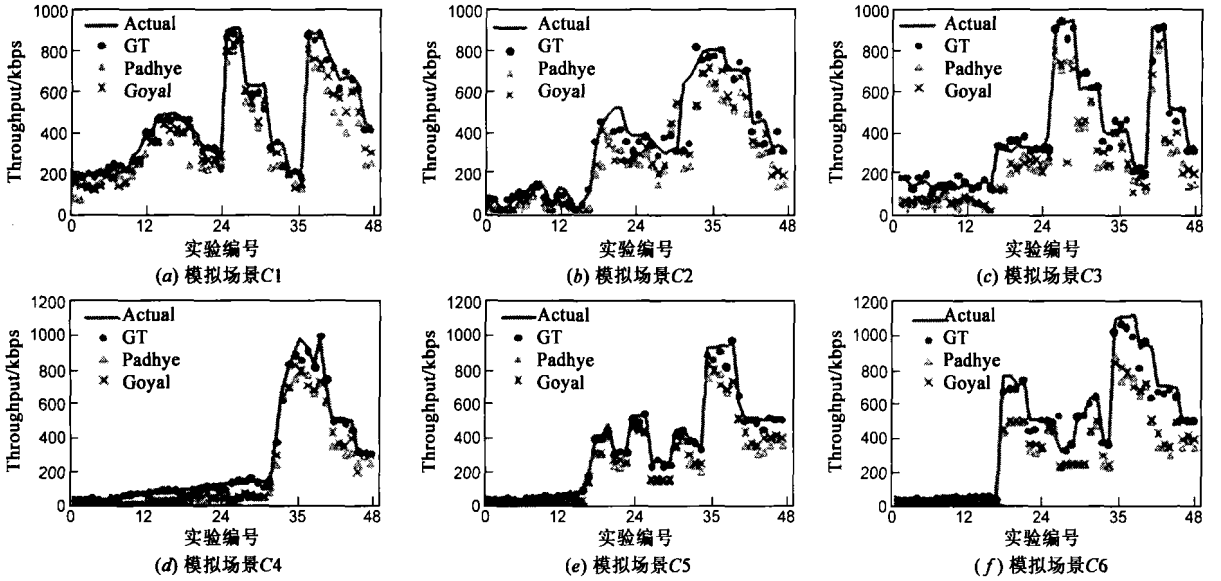
图3 实验仿真环境

的模拟进行了编号,如表 1.对这 6 种仿真场景,我们完成了每种仿真场景下 48 组不同参数条件的仿真实验.

表 1　模拟场景 ID 表

场景	参与结点	传输类型	瓶颈链路	队列方式
C1	*Sr2-De2*	TCP	link1	DropTail
C2	*Sr4-De4*	UDP	link2	DropTail
C3	*Sr3-De3*, *Sr4-De4*	TCP UDP	link2	DropTail
C4	*Sr4-De4*	UDP	link2	RED
C5	*Sr3-De3*, *Sr4-De4*	TCP, UDP	link2	RED
C6	*Sr1-De1*, *Sr2-De2*	TCP, TCP	link1	RED

为了验证本文提出的模型的有效性,我们将它的结果和与采用简单丢包模式的 Padhye 模型以及采用 Bernoulli 丢包模式的 Goyal 模型在相同条件下进行比较.为了表述的方便,下面将本论文提出的模型简称为 GT 模型.仿真结果如图 4(*a*)~(*f*)所示.图中实线代表实验中被监测传输流的实际 TCP 吞吐量,小圆形代表文中提出的新模型 GT,叉号和小三角形分别代表 Padhye 模型和 Goyal 模型,纵坐标代表吞吐量,横坐标代表不同实验条件.从仿真结果图也可以看出,由于采用了更为准确的丢包模型来描述端到端路径的丢包行为,本文提出的 GT 模型比 Padhye 模型和 Goyal 模型更加准确,与实际值最为接近.特别在高带宽高延时或者低带宽低延时情形下的精确度与 Padhye 模型和 Goyal 模型相比均有较大提升.



(*a*) 模拟场景*C1*　(*b*) 模拟场景*C2*　(*c*) 模拟场景*C3*

(*d*) 模拟场景*C4*　(*e*) 模拟场景*C5*　(*f*) 模拟场景*C6*

图4　不同场景下模型精确度比较图

4.2　测量实验

进一步分析实际测量实验中,在不同丢包率条件下 TCP 吞吐量测量结果的准确性,比较的对象包括本文提出的 GT 模型、Goyal 模型以及直接测量方法.实验的环境如图 5 所示.路由器由 NistNet[13] 搭建,利用它可以提供任意大小的包丢失率且包的丢失是随机的.

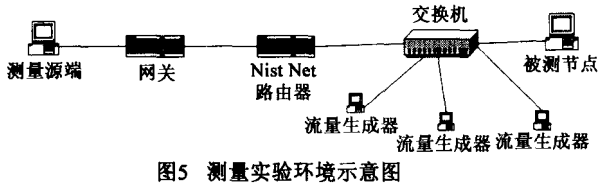
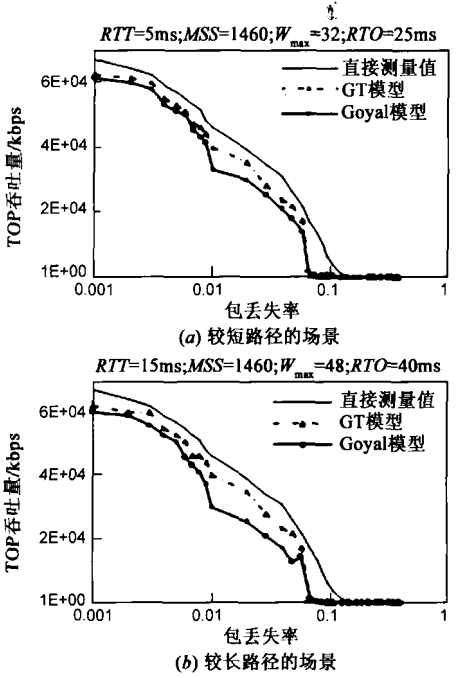


图5　测量实验环境示意图

图 6 给出了三种方法所测得的 TCP 吞吐量值,其中横轴代表包丢失率(对数刻度),纵轴代表以 kbps 为单位的 TCP 吞吐量.图中的每一个点代表给定包丢失率情况下 10 次吞吐量测量结果的平均值.图 6(*a*)中配置的模型参数是仿真局域网的较短路径场景,图 6(*b*)中配置的模型参数是仿真广域网的较长路径场景.从图中可以看出,随着包丢失率的增大,TCP 吞吐量呈迅速



(*a*) 较短路径的场景

(*b*) 较长路径的场景

图6　不同包丢失率下TCP吞吐量测量结果

下降的趋势.其中,由于直接测量方法是直接统计单位时间内发送的 TCP 报文段数目,因此它的测量结果最接近于实际情况,而且该方法不受模型参数的影响,结果接近于真实值.图 6(a)中,当包丢失率处于 0.7%到 6%之间时,GT 方法相对与 Goyal 模型更接近直接测量结果,而当前实际网络中的丢包率大部分都处于该区间;图 6(b)中,网络延迟的变化增大,Goyal 模型与实际测量值的偏差更大.因此在同等条件下,改进模型 GT 的计算结果曲线与实际测量值更加相符,能更好地反应实际 TCP 数据流的吞吐量.

5 结论

本文利用四状态 Gilbert 丢包机制对 TCP 吞吐量模型进行了重新推导,该模型考虑了 TCP 拥塞控制的所有阶段,采用了更精确的丢包机制.实验结果表明,本文提出的 GT 模型在多种情形下的测量结果与 Padhye 模型和 Goyal 模型相比,精确度都有一定提升,能更好地应用于实际 Internet.

致谢:向对本文提供帮助与支持的中科院计算所的潘亚军及黄靖表示衷心谢意.

参考文献:

[1] M Mathis, J Semke, J Mahdavi, et al. The macroscopic behavior of the TCP congestion avoidance algorithm[J]. ACM Computer Communication Review, 1997, 27(3): 67 - 82.

[2] J Padhye, V Firoiu, D Towsley, et al. Modeling TCP performance: a simple model and its empirical validation[J]. IEEE/ACM Transation on Network, 2000, 8(2): 133 - 145.

[3] 韩涛,朱耀庭,朱光喜,等.考虑慢启动影响的 TCP 吞吐量模型[J].电子学报,2002,30(10):1481 - 1484.
Han Tao, Zhu Yao-ting, Zhu Guang-xi, et al. A TCP throughput model considering slow start[J]. Acta Electronica Sinica, 2002, 30(10): 1481 - 1484. (in Chinese)

[4] B Sikdar, S Kalyanaraman, K S Vastola. Analytic models and comparative study of the latency and steady-state throughput of TCP tahoe, reno and sack[A]. Proc of GLOBECOM[C]. San Antonio, Texas, 2001. 1781 - 1787.

[5] M Goyal, R Guerin, R Rajan. Predicting TCP throughput from Non-invasive network sampling[A]. Proc of IEEE INFOCOM2002, New York, USA, 2002. 180 - 189.

[6] Dong Lu, Yi Qiao, Peter Dinda, et al. Characterizing and predicting TCP throughput on the wide area network[A]. Proc of the 25th IEEE International Conference on Distributed Computing Systems[C]. Columbus, Ohio, 2005. 1 - 11.

[7] 潘亚军. TCP 吞吐量测量方法研究[D].北京:首都师范大学,2005.
Pan Ya-jun. Measurement Method Study of TCP Throughput [D]. Beijing: Capital Normal University, 2005. (in Chinese)

[8] Qi He, Constantinos Dovrolis, Mostafa Ammar. On the predictability of large transfer TCP throughput[J]. Computer Networks: The International Journal of Computer and Telecommunications Networking, 2007, 51(14): 3959 - 3977.

[9] Hannan Xiao, Kee Chaing Chua, Malcolm, J. A, et al. Theoretical analysis of TCP throughput in ad hoc wireless networks [A]. Proc of IEEE GLOBECOM 2005[C]. St Louis, 2005. 2714 - 2719.

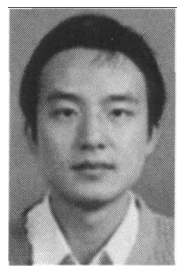
[10] Ding Lianghui, Zhang Wenjun, Xie Wei. Modeling TCP throughput in IEEE 802.11 based wireless ad hoc networks [A]. Proc of IEEE Communication Networks and Services Research Conference[C]. NW Washington, DC USA, 2008. 552 - 558.

[11] Yih-ching Su, Chu-sing Yang, Chen-wei Lee. The analysis of packet loss prediction for Gilbert-model with loss rate uplink [J]. Information Processing Letters, 2004, 90(3): 155 - 159.

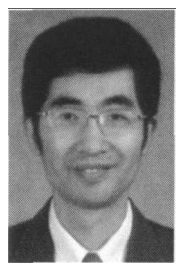
[12] Xunqi Yu, James W. Modestino, Xusheng Tian. The Accuracy of Gilbert models in predicting packet-Loss statistics for a single-multiplexer network model[A]. Proc of IEEE INFOCOM 2005[C]. Miami, 2005, 2602 - 2612.

[13] Mark Carson, Darrin Santay. NIST Net: A linux-based network emulation tool[J]. Aam Sigcomm Computer Communication Review, 2003, 33(3): 111 - 126.

作者简介:



曾 彬 男,1979 年生于湖南邵阳,湖南大学计算机与通信学院博士研究生,主要研究方向为网络性能监测与分析.
E-mail: zengbin604@hnu.cn



张大方 男,1959 年生于上海,教授,博士生导师,主要研究方向为软件容错、网络测试、软件测试、可信系统与网络.