

一种多 TCP 连接的建模分析¹

林 宇 程时端

(北京邮电大学程控交换与通信网国家重点实验室 北京 100876)

摘 要 单 TCP 连接的建模方法和改进, 已经有较多的研究成果. 对于多 TCP 连接的建模, 研究还较少. 该文针对多 TCP 连接, 给出了在接受者特性相同和接受者特性不同两种情况下的理论分析, 得到了 RTT(Round Trip Time) 与长期吞吐量随接受者数量、丢包率、端到端时延等参数变化的公式, 并通过仿真验证分析的有效性.

关键词 TCP, RTT, 吞吐量

中图分类号 TN919.3

1 引 言

传输控制协议 (Transfer Control Protocol, TCP) 单连接的建模方法以及相关的改进, 已经有较多的成果^[1-5]. 本文针对多 TCP 连接进行建模分析, 模型简述如下: 多个接收主机与同一发送主机建立多个 TCP 连接 (参见图 1, 其中 p 为端到端的丢包率). 本文的分析针对 Reno TCP, 并且假定接收主机和发送主机间的网络带宽足够, TCP 连接发送的数据为无限, 在这些假设下来考察多 TCP 连接的 RTT(Round Trip Time) 和长期平均吞吐量.

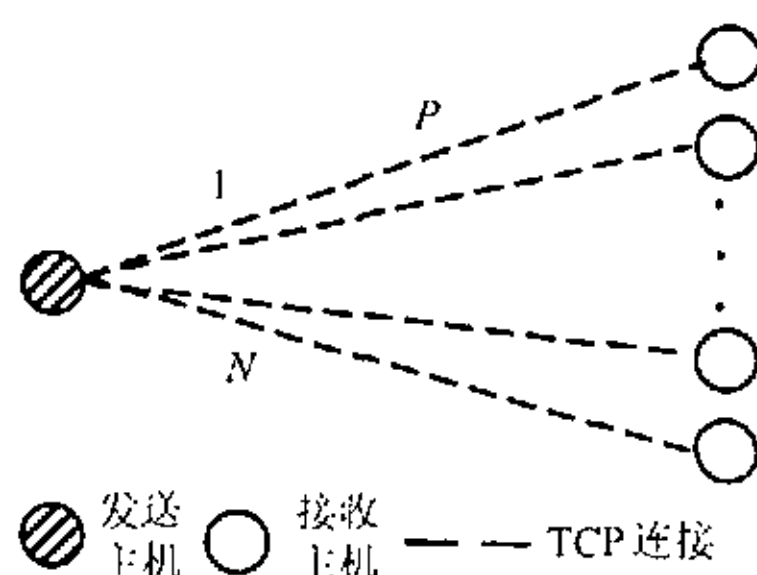


图 1 分析模型

本文的建模方法以文献 [1] 为基础, 使用的符号基本与文献 [1] 相同, 增加的符号如下:

Z_i^{TD} , Z_i^{TO} 为第 i 个 TD 和 TO 的时间长度 (定义同文献 [1]);

q 为随机选择的 TCP 连接处于 TD 状态的概率;

M^{TO} 为在 TO 状态下发送的数据包的个数;

N 为在发送主机和接收者间建立的 TCP 连接数;

N_{TD} , N_{TO} 为处于 TD 状态和 TO 状态的 TCP 连接数;

λ_{TD}^{TCP} , λ_{TO}^{TCP} 为处于 TD 状态和 TO 状态下的单 TCP 的数据包发送速率;

λ_S^R 为处于 TD 状态下 TCP 连接的接受端数据包的到达率;

¹ 2001-06-27 收到, 2001-11-26 定稿

$\lambda_{TD}^S, \lambda_{TO}^S$ 为发送端处于 TD 状态的所有 TCP 连接的发送速率之和或发送端处于 TO 状态的所有 TCP 连接的发送速率之和;

$\lambda_{RA}^S, \lambda_{RA}^R$ 为发送端或接收端 ACK(Acknowledgement, 确认) 包到达率;

X 为单一数据包的处理时间 (假定发送和接收时间相同);

Y 为单一 ACK 的处理时间 (假定其发送和接收时间相同);

p 为发送端和接收端间端到端的丢失率;

W^S, W^R 为在发送端和接收端数据包的排队等待时间;

2 接收者特性相同情况下多 TCP 连接的建模分析

首先考虑接收者特性相同的简单情况, 假定发送端接收端间的数据包丢失率为 p , 端到端时延为 τ . 在文献 [1] 中, TCP 吞吐量的近似模型为

$$B(p) \approx \min \left\{ \frac{W_{\max}}{RTT}, 1 / \left[RTT \sqrt{\frac{2bp}{3}} + T_0 \min \left(1, 3\sqrt{\frac{3bp}{8}} \right) p(1 + 32p^2) \right] \right\} \quad (1)$$

限于篇幅且最大窗口 W_{\max} 对 TCP 吞吐量影响的分析比较简单, 在本文中暂不讨论最大窗口 W_{\max} 对 TCP 吞吐量的影响, 取

$$B(p) \approx 1 / \left[RTT \sqrt{\frac{2bp}{3}} + T_0 \min \left(1, 3\sqrt{\frac{3bp}{8}} \right) p(1 + 32p^2) \right] \quad (2)$$

如图 2, 一个 TCP 连接交替处于 TD 和 TO 两个状态^[1], 为了方便叙述, 将处于 TD 状态的 TCP 连接为 TD TCP, 处于 TD 状态下的接收主机为 TD 接收主机. TO 状态的定义类似. 一个随机选定的 TCP 连接处于 TD 状态的概率为

$$q = E[Z_i^{TD}] / (E[Z_i^{TD}] + E[Z_i^{TO}]) = E[Z_i^{TD}] / (E[S]) \quad (3)$$

处于 TD 状态时长的数学期望 $E[Z_i^{TD}]$ 为^[1]

$$E[Z_i^{TD}] = E[A]/Q = E[RTT] \left\{ \frac{2+b}{6} + \sqrt{\frac{2b(1-p)}{3p} + \left(\frac{2+b}{6}\right)^2 + 1} \right\} / \left[\min \left(1, \frac{3}{E[W]} \right) \right] \quad (4)$$

假定各个 TCP 连接是独立的, 则 TD TCP 连接的数量为

$$E[N_{TD}] = \sum_{i=1}^N \binom{N}{i} q^i (1-q)^{N-i} \quad (5)$$

TO TCP 连接的数量为

$$E[N_{TO}] = N - E[N_{TD}] \quad (6)$$

图 3 中画出了系统中的数据流. 在发送端有 3 个数据流: 由 TD TCP 连接发送的数据流, 其到达率为 λ_{TD}^S ; 由 TO TCP 连接发送的数据流, 其到达率为 λ_{TO}^S ; 由 TD 接收主机返回的 ACK 流, 其到达率为 λ_{RA}^S . (对于接收主机, TO 接收主机收到的数据包远远小于 TD 接收

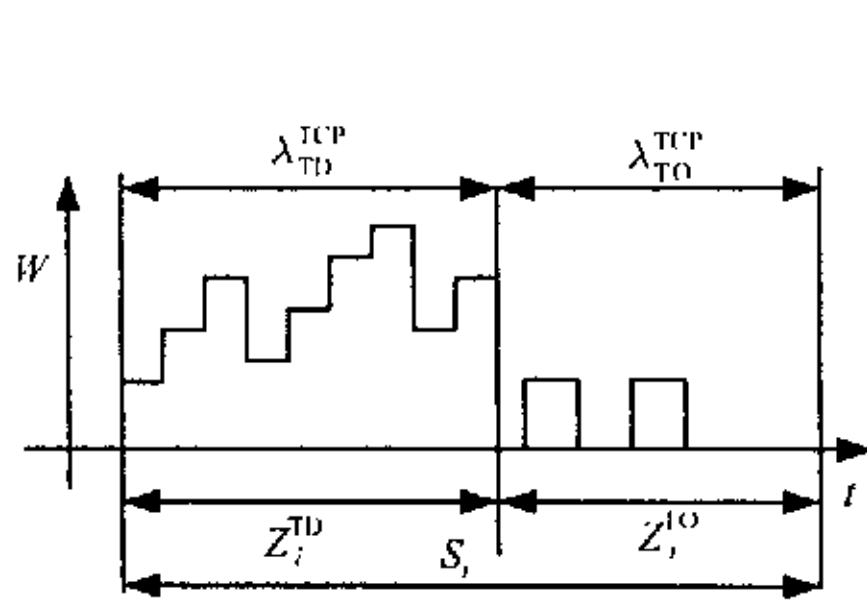


图 2 假定丢失指示为 3 个重复 ACK 或超时, TCP 连接的状态变化

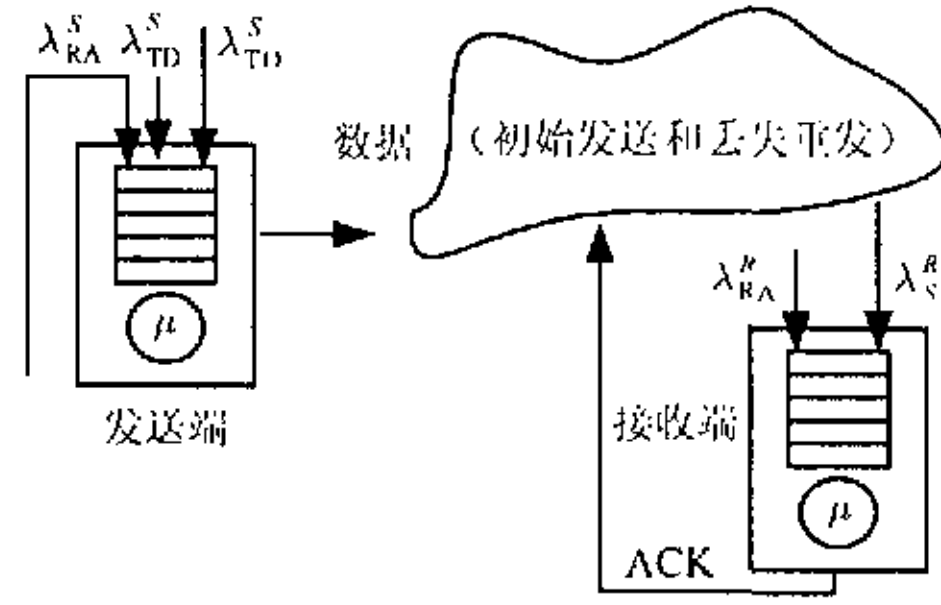


图 3 接收端和发送端的数据流

主机, 这里忽略了 TO 接收主机返回的 ACK 包流.) 在接收端有 2 个数据流: 接收端收到的来自发送方的数据包, 其到达率为 λ_S^R ; 接收端返回给发送端的数据包, 其到达率为 λ_{RA}^R .

对于随机选定的 T_i , 其处于 TD 状态和 TO 状态下发送数据包的速率分别为 (参见图 2)

$$\lambda_{TD}^{TCP} = \frac{B \times E[S] - E[M^{TO}]}{E[Z_i^{TD}]}, \quad \lambda_{TO}^{TCP} = \frac{E[M^{TO}]}{E[Z_i^{TO}]} \tag{7}$$

当 T_i 处于 TD 状态, 其对应的接收端收到数据包的到达率为

$$\lambda_S^R = \lambda_{TD}^{TCP} (1 - p) \tag{8}$$

如果假定接收端每正确收到 b 个数据包, 发送一个 ACK, 则接收端返回 ACK 的到达率为

$$\lambda_{RA}^R = \lambda_S^R / b \tag{9}$$

则在发送端, 由所有接收端返回 ACK 流到达率为

$$\lambda_{RA}^S = E[N_{TD}] \lambda_{RA}^R (1 - p) \tag{10}$$

对于 Reno TCP 而言, 在收到一个新的 ACK 后, 将发送后续的 b 个数据包, 并将发送窗口增加 b/W . 则发送端所有 TD TCP 发送数据包的到达率如 (11) 式. 通常, 处于 TD 状态的 TCP 连接, W 将远大于 b , (11) 式中进行了近似处理.

$$\lambda_{TD}^S \approx \lambda_{RA}^S (b + b/W) \approx \lambda_{RA}^S b \tag{11}$$

对于处于 TO 状态的 TCP 连接, 在 TO 期间, 只发送 M^{TO} 个数据包, 则发送端由所有 TO TCP 发送的数据包流的到达率为

$$\lambda_{TO}^S = E[N_{TO}] E[M^{TO}] / E[Z^{TO}] \tag{12}$$

如果假定数据包的产生是泊松流, 并且数据包的服务时间是独立于到达过程的随机变量, 则系统可以用 M/G/1 排队系统近似. 发送端发送队列的负荷 ρ^S 为 [6]

$$\rho^S = \lambda_{TO}^S E[X] + \lambda_{RA}^S (E[Y] + bE[X]) \tag{13}$$

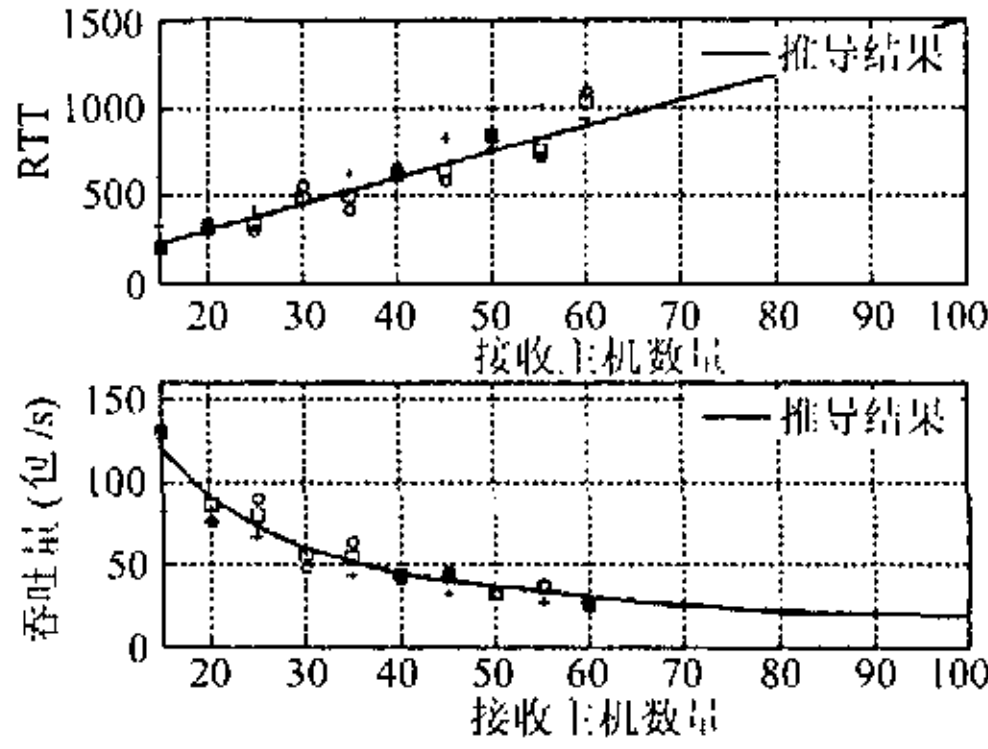


图 4 RTT 与吞吐量随接收主机数量变化曲线

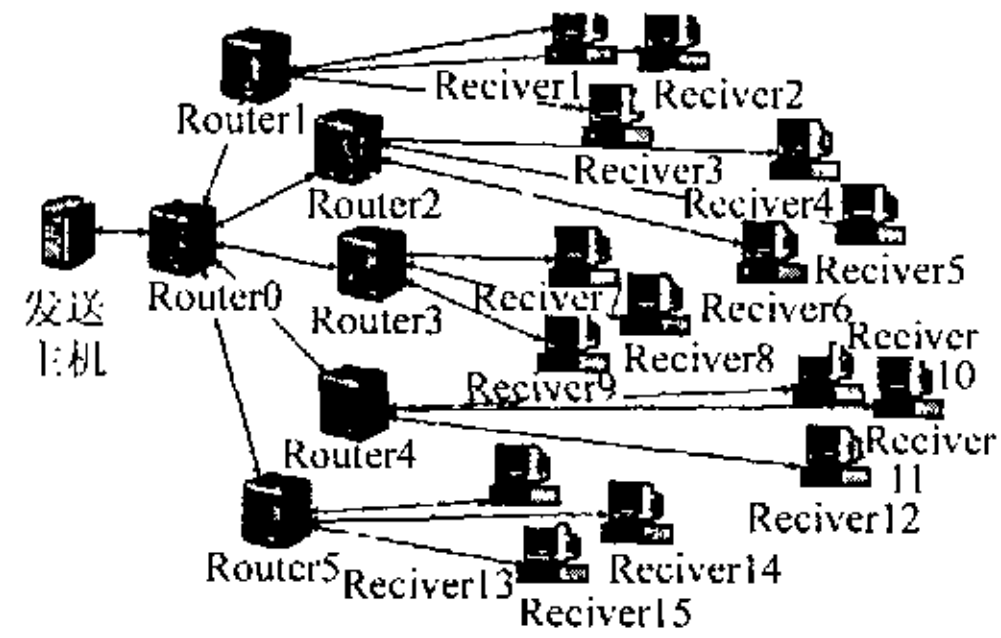


图 5 仿真模型

一个数据包发送请求在发送队列中的等待时间 $E[W^S]$ 为

$$E[W^S] = \frac{\lambda_{RA}^S E[(bX + Y)^2] + \lambda_{TO}^S E[X^2]}{2(1 - \rho^S)} \tag{14}$$

由于 TD 接收主机的负荷将远远大于 TO 接收主机，所以这里仅仅考虑 TD 接收主机的负荷，其值为

$$\rho^R = \lambda_S^R E[X] + \lambda_{RA}^R E[Y] \tag{15}$$

数据包在接收端的平均等待时间为

$$E[W^R] = \frac{\lambda_{RA}^R E[(X + Y/b)^2]}{2(1 - \rho^R)} \tag{16}$$

将 RTT 看成与在发送端缓冲区等待时间、发送时延、端到端时延、在接收端缓冲区等待时间等相关的随机变量，RTT 的数学期望为

$$E[RTT] = bE[X] + 2\tau + (bE[X] + E[W^R] + E[Y]) + (E[Y] + E[W^S]) \tag{17}$$

利用 (17) 式和 (2) 式联立，可以获得 $E[RTT]$ 和 B ，但通常只能获得数值解。

图 4 画出了解方程后 RTT 与吞吐量随接收主机数量的变化情况，其中，实线是推导的结果，离散点是仿真的结果（下同）。参数选取如下： $p = 0.001$ ， $E[X] = 500\mu s$ ， $E[Y] = 100\mu s$ ， $\tau = 300ms$ ， $b = 2$ 。为了便于对比，参数的选取参考文献 [7]。随着连接数量的增加，RTT 迅速增长，TCP 连接的长期平均吞吐量迅速下降。这是由于此时发送主机的处理能力已经成为 TCP 吞吐量的瓶颈，发送端需要处理大量的 ACK，这使得各个 TCP 连接的 RTT 迅速增长，最终导致 TCP 吞吐量的下降。

仿真的模型如图 5 所示，我们使用 OPNET 实现了该仿真模型。发送主机与路由器 Router0(离发送主机最近的路由器)间带宽为 150Mbps，Router0 与 5 个路由器 (Router1-Router5) 间带宽为 30Mbps，Router1-Router5 每个路由器连接 3 个接收主机，它们之间的带宽为 10Mbps。IP 包的包长为 300 字节，发送主机和接收主机间使用 FTP(File Transfer Protocol) 业务。各条链路的时延为 10ms，在 Router0 处以 0.001 的概率随机地进行 IP 包丢弃，其他路由器不进行丢弃，在发送主机和接收主机间建立数量为 15, 20, 25, ..., 60 个的 TCP 连接。图 5 中曲线为推导值，离散点为随机选择的 3 个 TCP 连接在不同连接数量情况下长期吞吐量的变化情况。

图 6 画出了 RTT、吞吐量随丢包率变化的情况。随着丢包率从 0.001 增长到 0.1, TCP 吞吐量几乎不变, 当丢包率大于 0.1 后, TCP 吞吐量迅速下降。这是由于如果丢包率变小, 接收端数据包的到达率增大, 则接收端的 ACK 速率增大, ACK 在发送端平均等待时间增长, RTT 将增长。丢包率下降是 TCP 吞吐量增长的因素, 但是丢包率下降却带来了 RTT 增长这一使 TCP 吞吐量下降的因素, 这使得丢包率在 0.001 到 0.1 变化范围内, TCP 吞吐量变化不大, 这个特性同单 TCP 吞吐量随丢失率变化的特征是不同的(类似地, 如果丢包率增大, 将使得 RTT 变短, RTT 的变短补偿了由于丢包率增大对 TCP 吞吐量的影响)。当丢包率大于 0.1 后, 由于丢包过多, 大部分的 TCP 连接都长时间处于 TO 状态, TCP 的吞吐量迅速下降。在进行仿真时, 在 Router0 处分别以 0.001, 0.002, 0.003, ..., 0.01, 0.02, ..., 0.1, ..., 0.3 概率进行丢弃, 在发送主机和接收主机间建立 50 个连接, 各链路时延为 10ms。

图 7 画出了 RTT、吞吐量随端到端时延变化的情况。随着端到端时延的增大, 开始时 RTT 并没有明显的增大, TCP 的吞吐量也几乎不变, 直到端到端时延增长到约为 320ms 处, RTT 开始迅速增长, TCP 的吞吐量迅速下降。这是由于端到端时延的增大, 将导致吞吐量的下降, 吞吐量的下降使得接收端 ACK 速率下降, ACK 速率下降又将导致发送端排队时延的减小, 这在一定程度上补偿端到端时延增大的影响, 因此, 在端到端时延增大幅度不大时, RTT 并没有明显的增大。随着端到端时延的继续增大(超过 320ms), 这种反馈的作用逐渐变小, 导致 RTT 开始迅速增长。在仿真中, 取 $p = 0.001$, TCP 连接数为 50, IP 包在 Router0 处进行相应的延迟(从 20ms 到 1420ms)。

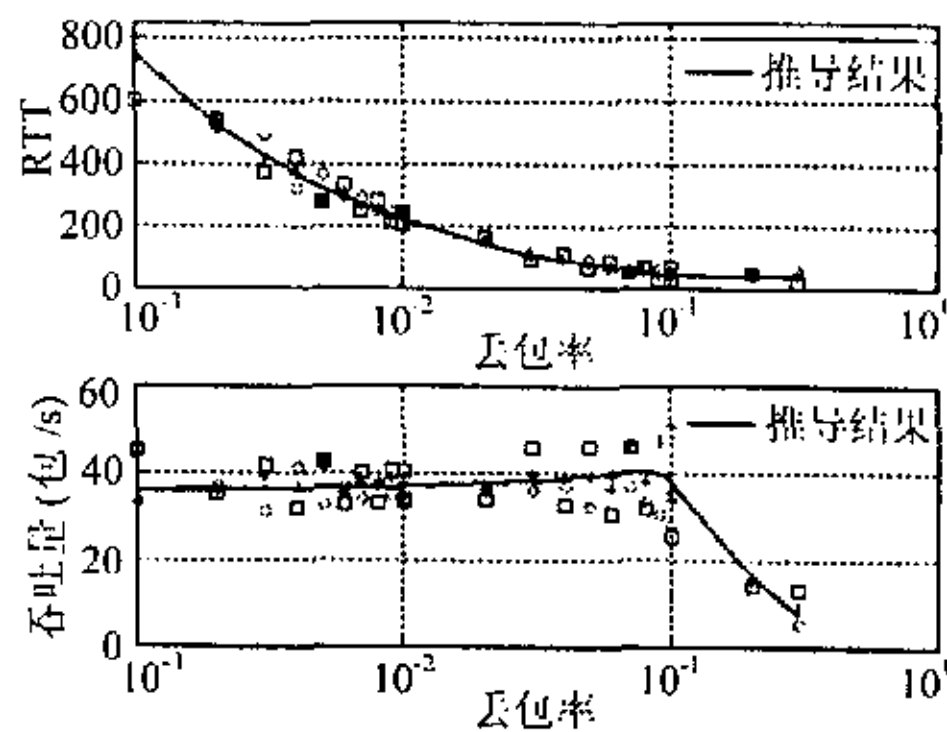


图 6 RTT, 吞吐量随丢包率变化曲线

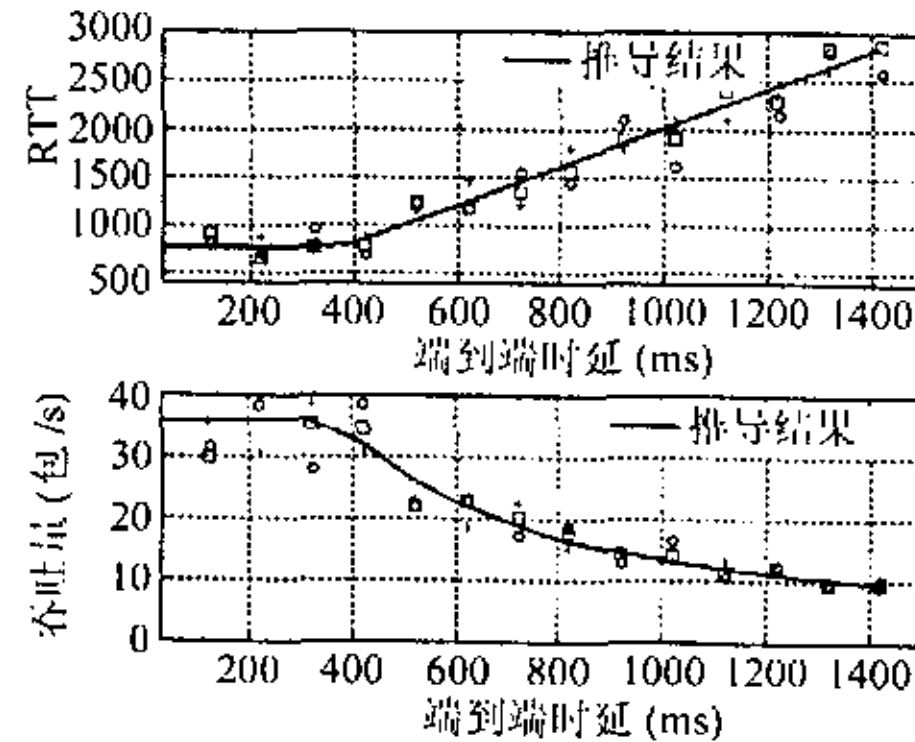


图 7 RTT, 吞吐量随端到端时延的变化曲线

3 接收者特性不同的情况下多 TCP 连接的建模分析

现在考虑接收主机特性不同的情况, 如果各个 TCP 连接具有不同的端到端时延和丢包概率, 则某个随机选定的连接 T_j 处于 TD 状态的概率为

$$q_j = \frac{E[Z_j^{TD}]}{E[Z_j^{TD} + Z_j^{TO}]} = \frac{E[Z_j^{TD}]}{E[S_j]} \tag{18}$$

则处于 TD 状态的 T_j 发送的数据包流的到达率为

$$\lambda_{TD-j}^{TCP} = \frac{B_j E[S_j] - E[M_j^{TO}]}{E[Z_j^{TD}]} \tag{19}$$

在接收端收到的数据包流的到达率为

$$\lambda_S^{R-j} = \lambda_{TD}^{TCP-j} p \quad (20)$$

其接收端返回的 ACK 流的速率为

$$\lambda_{RA}^{R-j} = \lambda_S^{R-j} / b \quad (21)$$

在发送端收到的 ACK 流的总速率为

$$\lambda_{RA}^S \approx \sum_{j=1}^N q_j \lambda_{RA}^{R-j} p_j \quad (22)$$

在发送端 TD TCP 发送的数据包流的总速率为

$$\lambda_{TD}^S \approx \lambda_{RA}^S (b + b/W) \approx \lambda_{RA}^S \times b \quad (23)$$

在发送端 TO TCP 发送的数据包流的到达率为

$$\lambda_{TO}^S \approx \sum_{j=1}^N (1 - q_j) E[Z_j^{TO}] / E[M_j^{TO}] \quad (24)$$

T_j 接收端的负荷 (由于接收端的负荷远远小于发送端, 这里只考虑 TD 接收端) 为

$$\rho^{R-j} = \lambda_S^{R-j} E[X] + \lambda_{RA}^{R-j} E[Y] \quad (25)$$

$$E[W^{R-j}] = \frac{\lambda_{RA}^{R-j} E[(X + Y/b)^2]}{2(1 - \rho^{R-j})} \quad (26)$$

$$E[RTT_j] = bE[X] + 2\tau_j + (bE[X] + E[W^{R-j}] + E[Y]) + (E[Y] + E[W^S]) \quad (27)$$

$$B_j \approx 1 / \left\{ E[RTT_j] \sqrt{\frac{2bp_j}{3}} + T_0^j \min \left(1, 3\sqrt{\frac{3bp_j}{8}} \right) p_j (1 + 32p_j^2) \right\} \quad (28)$$

通常, 难于求解方程组 (27) 式和 (28) 式. 考虑如下的两种情况: (1) 接收主机具有相同的端到端时延, 但是丢失率 p_j 不同. (2) 接收主机具有相同的丢失率, 但时延 τ_j 不同.

在假设 (1) 的假设下, (27) 式中只有 $E[W^{R-j}]$ 项与 j 有关. 因为发送端的负载远远高于接收端, 则 $E[W^{R-j}] \ll E[W^S] < E[RTT_j]$, $E[RTT_j]$ 几乎与 j 无关, 则近似地, $E[RTT_j] \approx E[RTT_i]$, $i \neq j$, 则方程组 (27) 式和 (28) 式易解. 我们考虑接收主机具有两种丢失率: $N(1-v)$ 个低丢包率接收主机 ($p = 0.001$), Nv 个高丢包率接收主机 ($p = 0.1$), 可以利用参数 v 来调整高丢包率接收者的比例. 这种方法曾在文献 [8] 中使用. 图 8 中画出了假设 (1) 的数值解. 具有高丢包率的接收者对应的 TCP 吞吐量远远低于低丢包率的接收者, 但两种接收者的比例关系对 TCP 的吞吐量影响不大 (图 8 中上方 3 条实线分别是在 5%, 15%, 25% 的比例下计算得到的低丢包率主机的吞吐量, 图 8 下方 3 条虚线为高丢包率主机的吞吐量). 在仿真时, 在发送主机和接收主机间建立 10, 20, ..., 60 个连接, 随机选定 10% 的连接设置为高丢包率连接, 在 Router0 中对选定的连接以 0.1 的概率进行丢弃, 对其他连接以 0.001 的概率进行丢弃, 其他路由器处理丢弃.

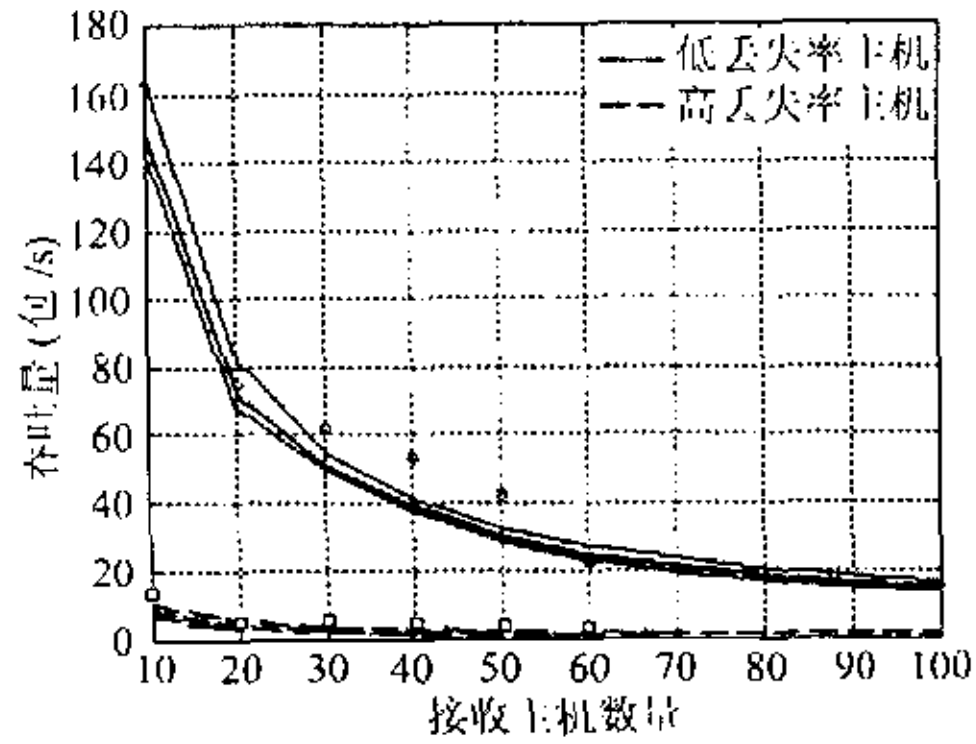


图 8 不同丢失率接收主机吞吐量比较

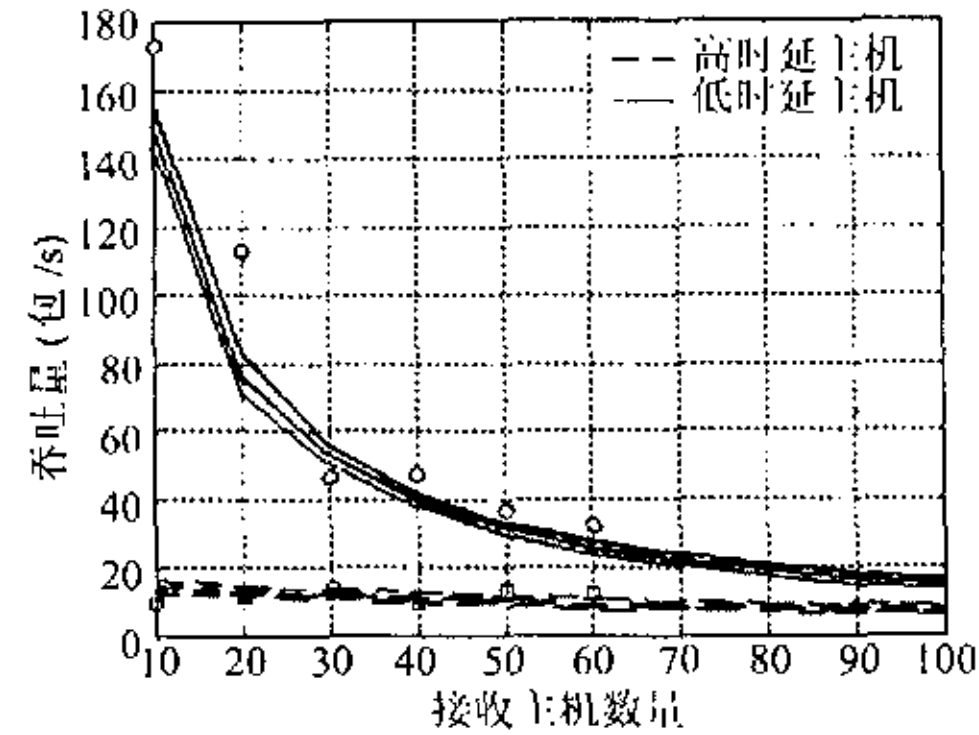


图 9 不同端到端时延接收主机吞吐量比较

在假设 (2) 下, 可以假定 $E[RTT_j] - 2\tau_j \approx E[RTT_i] - 2\tau_i, i \neq j$, 则方程组 (27) 式和 (28) 式易解. 类似地, 考虑具有低时延 ($\tau = 30\text{ms}$) 和高时延 ($\tau = 250\text{ms}$) 两种用户. 图 9 画出了高时延接收主机占不同比例时的吞吐量随用户数量的变化曲线 ($\rho = 0.01$). 高时延接收者的吞吐量远远低于低时延的接收者, 但两种接收者的比例关系对 TCP 的吞吐量影响不大. (图 9 中上方 3 条实线分别是在 5%, 15%, 25% 的比例下计算得到的低端到端时延主机的吞吐量, 图 9 下方 3 条虚线为高端到端主机的吞吐量). 仿真时, 在发送主机和接收主机间建立 10, 20, ..., 60 个连接, 随机选定接收者中 10% 的连接, 在 Router0 对 IP 包以 0.01 的概率进行丢弃, 并对选定连接上的 IP 包额外延迟 220ms.

4 结束语

本文提出了一种新的多 TCP 建模的方法, 并通过仿真结果验证分析的有效性. 关于在本文模型下最大窗口对多 TCP 连接吞吐量的影响, 将进一步进行研究.

参 考 文 献

- [1] J. Padhye, V. Firoin, D. Towsley, J. Kurose, Modeling TCP throughput: A simple model and its empirical validation, *IEEE/ACM Trans. on Networking*, 2000, 8(2), 133-145.
- [2] N. Cardwell, S. Savage, T. Anderson, Modeling TCP latency, *INFOCOM 2000, Israel, Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies*, 2000, Vol.3, 1742-1751.
- [3] C. Casetti, M. Meo, A new approach to model the stationary behavior of TCP connections, *ICC 2001, Helsinki, Finland*, 2000, Vol.1, 146-150.
- [4] A. A. Abouzeid, S. Roy, M. Azizoglu, Stochastic modeling of TCP over lossy links, *INFOCOM 2000, Israel, Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies*, 2000, Vol.3, 1724-1733.
- [5] R. Zorzi, A. Chockalingam, R. R. Rao, Throughput analysis of TCP on channels with memory, *IEEE J. on Selected Areas in Communications*, 2000, 18(7), 1289-1300.
- [6] L. Kleinrock, *Queueing Systems: Theory*, Wiley-Interscience, 1974, 191-192.
- [7] M. Yamamoto, J. Kurose, D. Towsley, H. Ikeda, A delay analysis of sender-initiated and receiver-initiated reliable multicast protocols, *INFOCOM 1997, San Francisco*, 1997, Vol. 2, 480-488.

- [8] J. Nonnenmacher, E. Biersack, D. Towsley, Parity-based loss recovery for reliable multicast transmission, *IEEE/ACM Trans. on Networking*, 1998, 6(4), 349-361.

A MODELING ANALYSIS OF MULTIPLE TCP CONNECTIONS

Lin Yu Cheng Shiduan

(National Lab of Switching Tech. & Telecom. Networks, BUPT, Beijing 100876, China)

Abstract Lots of research fruits have been presented in the modeling and improving methods of single TCP connection, while few in multiple TCP connections. This paper presents multiple TCP connections' modeling through theoretical analysis under two cases: receivers with same characteristics and receivers with different characteristics, and obtains the RTT and long-term TCP throughput formulas related to number of receivers, packet loss rate, and end-to-end delay. The simulation results confirm their effectiveness.

Key words TCP, RTT, Throughput

林 宇: 男, 1976 年生, 博士生, 研究方向包括 IP 多播、移动多播和网络调度、TCP 建模。

程时端: 女, 1940 年生, 主持过十余项重大科研项目, 发表文章 110 余篇, 专著 2 册, 译著 3 册, 1992-1999 年间, 任 863 网络与交换专家组组长, 目前研究方向包括 TCP/IP, 协议工程, 业务量工程, 宽带网络性能和服务质量等。