

自相似网络环境中 TCP-Friendly 协议的研究

刘郁恒^{1), 2)} 胡 严³⁾ 张光昭¹⁾

¹⁾(中山大学电子与通信工程系 广州 510275)

²⁾(广东省电信规划设计院 广州 510630)

³⁾(中国科学院计算技术研究所 北京 100080)

摘 要 该文针对自相似的网络条件提出了一种单播的 TCP-Friendly 拥塞控制算法,称为 RAAR-MT. 该算法通过在线预测未来流量水平的变化趋势,并利用预测结果在接收端对业务流进行多时间尺度的速率控制. 仿真结果表明,在自相似的网络条件下该算法能够有效降低业务流的丢失率,即使在传统短相关环境中,其性能也不会出现严重衰减. 通过与 TFRC 协议进行对比研究发现,RAAR-MT 协议在对 TCP 的友好性、协议内的公平性以及速率的平滑性等主要的 TCP-Friendly 协议指标都具有更好的性能. 由于 RAAR-MT 不需要进行每包反馈,且主要功能在接收方实现,因此可方便地将该机制引入多媒体组播传输系统中.

关键词 RAAR-MT; TCP-Friendly; 自相似; 拥塞控制; 流量预测

中图法分类号 TP393

Study on TCP-Friendly Protocol in Self-Similar Traffic Network

LIU Yu-Heng^{1), 2)} HU Yan³⁾ ZHANG Guang-Zhao¹⁾

¹⁾(Department of Electronics & Communication Engineering, SUN Yat-Sen University, Guangzhou 510275)

²⁾(Guangdong Planning and Designing Institute of Telecommunications, Guangzhou 510630)

³⁾(Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100080)

Abstract Since the traffic in modern telecommunication networks always shows self-similarity or long-range dependence, a new unicast TCP-Friendly protocol using in self-similar traffic network, called RAAR-MT, is proposed in this paper. RAAR-MT on-line predicts the tendency of traffic level in the near future and use this predicted information to exert multiple time scale traffic control at receivers. Simulation shows that, comparing with RAAR protocol, RAAR-MT protocol can enhance the tracking ability to the network state and gain 10%~20% performance improvement in loss rate on self-similar traffic network. Even in traditional short-range dependent environment, the performance of this protocol will not degrade severely. Comparisons have been made between RAAR-MT and TFRC too. Results are that RAAR-MT has better performance in TCP-Friendliness, intra-protocol fairness and smoothness in typical self-similar network environment. RAAR-MT is also a promising avenue of development for congestion control of multicast multimedia traffic, since it is not a per-packet feedback and is mainly implemented at receivers.

Keywords RAAR-MT; TCP-Friendly; self-similar; congestion control; traffic prediction

1 引 言

众所周知, TCP 与 UDP 协议都不能很好地满

足连续媒体流传输的需要. TCP 速率减半的拥塞退避机制容易引起多媒体流过大的速率波动, 其每包确认机制也是多媒体传输中所不希望的. 而不具备拥塞退避机制的 UDP 流(被视为违规流)在拥塞的

网络环境中将大量抢占具有拥塞退避机制的协议流的带宽,同时自身的丢包也迅速增加,并带来系统拥塞崩溃(congestion collapse)的潜在危险^[1].为保证系统的稳定性和服务的公平性,业界正在研究各种在拥塞条件下能对违规流进行严厉惩罚的队列管理机制(如 SFB^[2], RED-PD^[3], CHOKe^[4], SCHOKe^[5]等),在这些系统中,使用 UDP 的多媒体流在遭遇拥塞时,将会遇到更为严重的性能衰减.

目前 Internet 中的传输业务主要是基于 TCP 的^[6],随着多媒体实时应用在 Internet 的迅速增长,研究一个适合于多媒体传输,具有拥塞退避机制,且能够与 TCP 公平分享带宽的传输协议,成为 Internet 传输的一个重要课题.如果一种通信协议与在同等条件下的 TCP 流具有近似相同的吞吐量,则称这种协议是 TCP 友好(TCP-Friendly)的. TCP-Friendly 协议的研究得到了业界的广泛重视,目前提出的 TCP-Friendly 协议大致可分为 2 类:一类是基于 AIMD 的,如 TEAR^[7], RAP^[8], RAAR^[19]等,另一类是基于数学模型的,如 TFRC^[9]等.其中,TFRC 已经成为了 IETF 工作组的正式草案^[10],协议的标准化进程正在启动.

近十年来,大量网络测量与分析证明,自相似在现代高速通信系统中是一种普遍存在的现象^[11].网络业务流自相似特性的发现,打破了原有网络流是短相关的基础性假设,自然会带来许多与原有假设不同的性能特征.自相似网络流量的长时突发性,通常对网络性能起破坏性的作用,如增加队列长度、时延以及数据包的丢失率等.但是,新的流量模型的引入也可能为网络流量的控制带来新的更加高效的解决方案.目前,对在自相似环境中 TCP-Friendly 协议的性能研究还属于一个崭新的研究课题.本文中,我们将讨论利用网络业务流具有自相似的特点,在 RAAR 协议的基础上,通过接收端在线预测业务流的未来变化趋势,在接收端实施多时间尺度的拥塞控制策略,增强系统对网络变化的跟踪能力,降低业务流的丢失率,提高传输质量,并在自相似的网络环境中保持良好的 TCP-Friendly 性能.

2 网络业务流的自相似性与自相似流的预测

2.1 自相似过程与长相关(long-range dependence)

从直观上来说,自相似的网络流量是指无论以哪个时间尺度作为基本测量单位时间去观察,网

络的流量都具有相似的特征.考察一个计数过程 $\{N_i(\Delta)\}$, $N_i(\Delta)$ 表示在第 i 个时间间隔(间隔长度为 Δ)内分组到达的数量.如果对于所有的扩张因子 $s > 0$,均有 $\{N_j(s\Delta)\}$ 的扩张过程的所有有限维分布 $\{s^{-H}N_j(s\Delta)\}_{j=1,2,\dots}$ ($0 < H < 1$),都与原来的过程 $\{N_i(\Delta)\}_{i=1,2,\dots}$ 相同,则称过程 $\{N_i(\Delta)\}$ 为自相似过程,并称 H 为其自相似参数(又称 Hurst 参数).由上面的定义可知,如果 $\{N_i(\Delta)\}$ 是自相似过程,那么 $\{N_i(\Delta)\}$ 与经过归一化后(归一化因子为 s^{-H})的时间扩展过程 $\{N_j(s\Delta)\}$ 具有相同的分布.因此,自相似过程具有时间扩展不变性(scale-invariance).随机过程的自相似特性也可以完全由其自相关结构来定义.

定义协方差平稳过程 $\{X_i\}$ 长度为 m 的聚集过程

(aggregated process) $X^{(m)}$ 为 $X_i^{(m)} = \frac{1}{m} \sum_{t=m(i-1)+1}^{im} X_t$.

设 $\gamma(k)$ 和 $\gamma^{(m)}(k)$ 分别为 X 和 $X^{(m)}$ 的自相关函数,如果对于所有 $m \geq 1, k \geq 1, \gamma^{(m)}(k) \equiv \gamma(k)$,则称 $X(t)$ 是一个精确二阶自相似过程;如果对于所有 $k \geq 1$,仅有 $\lim_{m \rightarrow \infty} \gamma^{(m)}(k) \rightarrow \gamma(k)$ 成立,则称 $X(t)$ 是一个渐近二阶自相似过程.可见,二阶自相似过程能够在在大时间尺度的聚集过程中保持自相关结构的精确相同或者渐近相同,即当 $m \rightarrow \infty$ 时,其聚集过程 $X^{(m)}$ 的自相关结构并不退化.而在传统的业务模型中,当 $m \rightarrow \infty$ 时,聚集过程 $X^{(m)}$ 的自相关结构将退化,趋向于白噪声,即 $\gamma^{(m)}(k) \rightarrow 0$,当 $m \rightarrow \infty, k \geq 1$.

如果协方差平稳过程 $\{X_i\}$ 的自相关函数 $r(k)$

不收敛,即 $\lim_{n \rightarrow \infty} \sum_{k=1}^n r(k) = \infty$,则称平稳过程 $\{X_i\}$ 具有长相关性(Long-Range Dependence, LRD),这种特性也称为约瑟夫效应(Joseph effect).在长相关过程中,距离较远的数据的相关性是不可忽略的.而传统基于 Markov 过程的模型均视长距离的相关性为零,即 $\lim_{n \rightarrow \infty} \sum_{k=1}^n r(k) < \infty$,只具有短相关特性.

可以证明,上面定义的渐近二阶自相似过程实际上是长相关过程的一个子集^[11],反之,当 LRD 过程的自相关系数的拖尾以幂函数形式衰减时,即 $r(k) \sim k^{-\beta}$,则该过程也将呈现渐近二阶自相似^[12].因此本文中,在不导致混淆的情况下,长相关与自相似(实际上是指渐近二阶自相似)的概念可互用.

2.2 自相似流量的预测

为了通过观察最近的过去网络的流量情况来对未来的网络流量水平作出预测,需要考察条件概率

$P\{X_{i+1}^{(m)} | X_i^{(m)}\}$ 的分布情况. 对于传统的短时相关的网络流, 因其只具有短时的相关性, 未来某个较大时间尺度的流量水平, 与过去网络流量观察值的相关性很小, 即 $P\{X_{i+1}^{(m)} | X_i^{(m)}\} \sim P\{X_{i+1}^{(m)}\}$, 所以无法在合适的时间尺度上根据过去的流量情况预测未来一段时间内的流量情况. 当网络的业务流具有长相关特性时, 网络流极端值的持续时期比较长, 体现为网络流偏离均值的持续时间较长. 对于在最近过去的某个时间尺度 m 内的两个观察值 $a, b > 0 (a \neq b)$, 将有 $P\{X_{i+1}^{(m)} | X_i^{(m)} = b\} \neq P\{X_{i+1}^{(m)} | X_i^{(m)} = a\}$, 因此, 利用在最近过去的某个时间尺度 m 对网络流量的观察信息, 可以较好地预测未来网络流量水平的变化趋势.

为了方便对未来流量的变化趋势作出预测, 可把网络流量量化成 h 个等级, 令 $x_i^{(m)} = L(X_i^{(m)})$. $x_i^{(m)} = 1$ 表示网络处于最低的流量水平; $x_i^{(m)} = h$ 表示网络处于最高的流量水平. 令 L_1 与 L_2 分别表示相邻两个时间段 (区间长度为 m) 的流量水平 $x_i^{(m)}$ 与 $x_{i+1}^{(m)}$, 研究表明^[13,14], 对于长相关业务流, 在较大的观察时间尺度下, 条件概率密度 $P\{L_2 | L_1 = m\}$ 的分布具有向 $L_2 = m$ 集中的趋势, 即如果某一区间内的平均流量值较大, 则下一区间内的平均流量值较大的可能性也较大; 反之, 如果某一区间内的平均流量值较小, 则下一区间的平均流量值很可能也较小.

在实际应用当中, 利用自相似业务流在大时间尺度下自相关结构保持不变的特性, 通过观察当前系统流量水平的状态, 根据条件概率密度分布矩阵 (实时测量) 可以较为准确地预测未来系统的流量水平状况, 从而为流量控制决策提供了高效的参考依据, 可望用于大幅提高当前基于反馈环 (或 RTT , 反馈周期约为 20~200ms) 的拥塞控制机制的性能.

3 多时间尺度的

RAAR 算法 (RAAR-MT)

RAAR-MT 算法是在 RAAR 算法^[19] 的基础上, 利用网络流量自相似的特点, 在更大的时间尺度对节点的发送速率加以控制, 使得网络节点的发送速率能够更好地适应网络背景流的变化, 从而获得更好的协议性能. 本文提出的多时间尺度的速率控制算法与 Park^[13,14] 等人提出的多时间尺度拥塞控制算法相比较, 有以下的特点:

(1) Park 等人的多时间尺度拥塞控制 (MTSC) 算法完全在发送端进行, 而 RAAR-MT 算法的流量

预测与速率控制工作主要由接收端完成. 因为在多媒体传输系统中, 发送方一般为多媒体服务器, 工作负荷比较重, 所以, 我们希望复杂的工作尽可能在接收端完成, 以减轻服务器的工作负担, 提高系统的扩展能力, 便于该机制将来向组播系统扩展.

(2) Park 等人采用的是非对称的速率控制算法, 称为选择性的贪婪控制 (Selective Aggressiveness Control, SAC)^[13], 通过迅速发现并占用空闲的带宽以获取尽可能高的吞吐量增益为最终目标. 而 RAAR-MT 作为 TCP-Friendly 协议, 需要与 TCP 尽可能公平地使用带宽, 因此采用了对称的控制算法, 以获得丢失率的性能增益为目标.

(3) 通过对网络系统总体流量的直接测量来预测未来网络流量的变化趋势尽管具有非常高的理论价值, 但在现有网络环境中很难实施. 我们在尽量不增加中间网络节点的复杂度以及系统信令开销的条件下, 采用单流的 RTT 估算值作为网络系统的拥塞状况的度量, 通过对未来网络拥塞概率的估算, 在比正常反馈周期 (RTT) 大 1~2 个数量级的更大时间尺度上对发送速率实施控制.

3.1 RAAR 算法

RAAR 本质上是一种 GAIMD (General Additive Increase Multiplicative Decrease)^[15] 控制机制, 但速率控制主要在接收方完成. 接收方依据 GAIMD 的原理评估合适的发送速率, 并通过反馈 (ACK) 分组控制发送方的发送速率. 发送端接收到一个 ACK 分组后, 就从 ACK 分组中取出速率字段中的速率, 马上按这个速率发送数据分组, 这时表明一个新的 round (如图 1 所示) 开始; 直到发送端接收到新的 ACK 分组, 这个 round 才结束, 同时开始下一个 round. 在以下两种情况之一出现时,

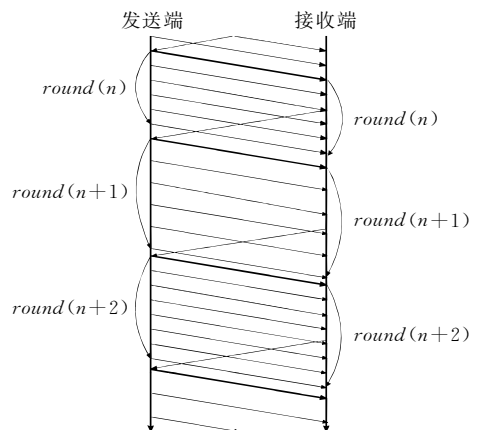


图 1 RAAR 协议

需要立即发送 ACK 分组,通知发送端更新发送速率;(1)接收端最新估计的发送速率小于当前的发送速率;(2)接收端的 RTT 计时器超时。

在 ACK 分组无丢失的情况下,发送方每隔一个时间段 $t \leq 2RTT$ 就应该接收到一个 ACK 分组.当接收方在 $2RTO$ 后还没有收到新的 ACK 分组,则发送方需要主动降低发送速率(超时减速功能),把速率调整为 $\beta \times rate_-$. RTO 的估算也在接收方中进行,并通过 ACK 分组反馈给发送方。

在每包确认的 GAIMD 机制中,发送方每收到一个表明无分组丢失的 ACK 报文,则发送窗口增加 α ; 如果检测到分组丢失事件,则把窗口减少为原来的 β 倍,其中 $0 < \beta < 1$, $\alpha > 0$. 对于 TCP, $\alpha = 1$, $\beta = 0.5$. 要获得一个与 TCP 吞吐量相当的 GAIMD, 需要满足^[15]

$$\alpha = \frac{4(1-\beta^2)}{3} \quad (1)$$

RAAR 抛弃了每包确认机制(每个 $round$ 只发送一个 ACK 报文),并对 GAIMD 机制进行修改,使之适合于在接收端实现.当接收端接收到第一个数据分组后就进入慢启动(SLOW-START)状态.此后,接收端每按序收到一个分组就把目前的速率 $rate_-$ 修改为 $rate_- + pktSize_- / RTT$,以尽快地发现可合理利用的带宽.如果在接收端探测到了一个丢分组事件,则 RAAR 进入拥塞退避状态.在这个阶段中,RAAR 利用了修改的 GAIMD 算法对速率进行调整。

$$\begin{cases} rate_- \leftarrow rate_- + \frac{\alpha \cdot k_0}{wnd_-} \times \frac{pktSize_-}{RTT} \times FA_MT, \\ \quad \text{如果没有发生丢包事件} \\ rate_- \leftarrow rate_- \times \beta \times FM_MT, \\ \quad \text{如果发生了丢包事件} \end{cases} \quad (2)$$

其中, wnd_- 是在接收端估计的在一个 $round$ 内接收分组的数目, k_0 是一个通过实验测量的修正因子.通过仿真发现, k_0 取 $1/6$ 协议性能较佳, k_0 的最佳取值问题还需继续研究. FA_MT 与 FM_MT 是在较大时间尺度下的速率控制因子(详见下文的叙述). (α, β) 是满足式(1)的常数对,我们以 $(0.3125, 0.875)$ 代替 TCP 的 $(1, 0.5)$,以减少发送速率的波动。

端系统对一个拥塞现象从探测到作出反应约需要一个 RTT 的时间,在这个 RTT 内就会出现多个分组丢失,但这些分组丢失实际上是由同一个网络拥塞引起的.因此以分组丢失事件来代替个别分组

的丢失作为速率调整的依据是合理的.接收端根据分组的序列号容易判断分组的丢失.当发生第一个分组丢失时,就可认为一个分组丢失事件发生了,同时启动一个长度为 RTT 的定时器 T ,在这个定时器超时前的所有分组丢失都可认为是属于相同的一个分组丢失事件(由相同的拥塞引起).RAAR 系统通过分组丢失事件来判断网络的拥塞变化,即,在上一个成功接收的分组和最新成功接收的分组之间没有发生分组丢失事件时增加发送速率,反之,减少发送速率。

3.2 网络流量的在线实时预测

RAAR-MT 对网络流量的在线实时预测也在接收端完成.接收端对 $[t - Ts, t]$ 时间段内的网络系统流量状况进行评估,将其量化为流量水平 L_1 (设流量水平分为 m 级, $L_1 = l, 1 \leq l \leq m$),则接收端根据在线评估的流量条件概率分布,计算在下一个时间间隔 $[t, t + Ts]$ 内的流量水平 L_2 的期望值: $\bar{l}' = E(L_2 | L_1 = l) = \sum_{l'=1}^m l' \times P(L_2 = l' | L_1 = l)$,其中,流量概率分布的估算由一个概率矩阵 $Condprob[m][m+1]$ 的实时更新来实现.我们把时间轴分成长度为 Ts 的连续子块. $Condprob[l][l']$ ($1 \leq l \leq m, 1 \leq l' \leq m$) 记录在 $[t - Ts, t]$ 内平均流量水平为 $L_1 = l$ 的条件下,在其后续的 $[t, t + Ts]$ 时间间隔内平均流量水平为 l' 的子块数目 $h_{l'}$, $Condprob[l][m+1]$ 记录满足 $L_1 = l$ 的子块的总数 h_l ,则条件概率 $P(L_2 = l' | L_1 = l)$ 可直接由 $h_{l'} / h_l$ 得到。

具体地,假设 $[(n-1)Ts, nTs]$ 的流量水平 $L_1 = l$,在下一个子块 $[nTs, (n+1)Ts]$,若流量水平 $L_2 = l'$,则令 $Condprob[L_1][L_2]$ 及 $Condprob[L_1][m+1]$ 的计数值分别加 1,同时修改在当前流量水平为 L_1 的条件下,下一时间间隔流量水平的期望值 $arv_condprob[L_1]$.最后,令 $L_1 \leftarrow L_2$,继续观察下一时间间隔 $[(n+1)Ts, (n+2)Ts]$ 的流量水平.上述在线流量预测操作在每个测量间隔内的复杂度仅为 $O(1)$,并不会对网络节点带来过重的负担。

由于端系统较难直接观察整个网络的流量状况,因此只能通过观察相关的流量参数来间接评估网络当前以及未来的拥塞水平.方便接收端直接测量且能间接反映网络系统总体流量的系统参数主要包括业务流的平均到达速率、分组平均丢失率以及端一端 RTT 的估算值.但考虑到前 2 个参数与底层的拥塞控制算法紧密耦合,不一定能真实反映网络当前的状况,因此我们把端一端 RTT 的估算值

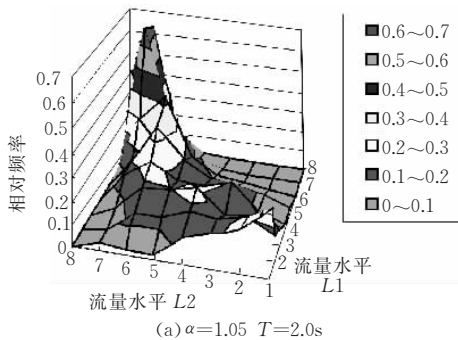
作为网络当前流量状况的度量. 当网络的传播时延与网络系统对分组处理的时延固定时, RTT 大小的变化主要由分组在网络节点的排队时延来决定, 而分组在网络节点的排队时延则可以反映当前网络系统的拥塞水平. 当 RTT 相对比较大时, 我们认为当前网络流量处于高水平, 当 RTT 相对比较低时, 则认为网络流量处于低水平.

令 $RTT_{Am} = RTT_{max} - RTT_{min}$, $RTT_{var} = (RTT_{now} - RTT_{min}) / RTT_{Am}$. 其中, RTT_{max} 与 RTT_{min} 分别为传输过程中测量得到的最大与最小的 RTT 值; RTT_{now} 为当前的时间间隔 T_s 内 RTT 的估值; $RTT_{var} \in [0, 1]$ 为当前 RTT 估值相对大小的一个度量. 根据 RTT_{var} 的取值, 我们把网络的流量水平分为 8 个等级 (如表 1 所示). 等级越高, 表明网络的拥塞程度越高; 等级越低, 表明网络越空闲.

表 1 流量水平的划分

流量水平	RTT_{var} 值	流量水平	RTT_{var} 值
1	[0, 0.05)	5	[0.50, 0.65)
2	[0.05, 0.20)	6	[0.65, 0.80)
3	[0.20, 0.35)	7	[0.80, 0.95)
4	[0.35, 0.50)	8	[0.95, 1]

图 2 是根据上述的流量预测算法以及流量水平等级划分标准通过在线评估获得的流量水平条件概率分布图 (预测时间间隔 $T_s = 2s$). 其中图 2(a) 的背景流为自相似背景流, 背景流的重尾指数 $\alpha = 1.05$ ($H = 0.975$), 条件概率的谱线向对角线集中的趋势



明显, 即下一时间间隔的流量水平与当前流量水平比较接近的概率较大; 图 2(b) 的背景流为传统的短相关背景流, 背景流的重尾指数 $\alpha = 1.95$ ($H = 0.525$), 未来的流量水平分布与当前的流量水平基本无关, 主要在均值附近抖动, 即 $P(L2 | L1) \approx P(L2)$.

由此可见, 我们测出来的条件概率分布矩阵与自相似流量模型的可预测性分析 (见 2.2 节) 相当一致, 因此我们的流量水平的划分标准与在线流量水平预测算法是比较准确而有效的. 对于持续时间较长的业务流, 运用这种在线流量预测算法, 我们可以较为准确预测未来较大的一个时间尺度上网络系统的流量变化状况, 为流量控制措施的实施提供了较为可靠的依据, 可望大幅提高当前基于反馈环 (或 RTT , 反馈周期约为 20~200ms) 的拥塞控制机制的性能. 尽管需要引入额外的存储空间与系统处理时间, 但这种流量预测算法主要在负担较轻的接收端执行, 预测操作的复杂度在每个测量间隔内仅为 $O(1)$, 在合理规划测量时间间隔与流量水平等级数的条件下, 不会明显加重系统的存储与处理的负担. 然而, 由于条件概率分布矩阵的准确生成需要足够数量的测量点, 对于持续时间很短的流不一定能获得准确的条件概率分布矩阵. 为此, 我们把条件概率分布矩阵按照传统的短相关模型进行初始化, 使持续时间短的流基本可以按照传统模型进行流量控制.

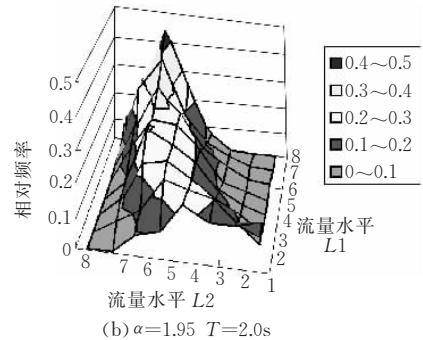


图 2 条件概率密度分布

3.3 大时间尺度的速率控制

利用在线流量预测的结果, 在更大的时间尺度上对系统在传统小时间尺度的拥塞控制实施调制, 可在一定程度上弥补小时间尺度的反馈控制难以反映网络流量长期变化趋势的缺陷. 我们在 RAAR-MT 中的具体做法是根据未来流量水平的预测期望值 \bar{l}' , 调整式 (2) 中的加性控制因子 FA_MT 与乘性控制因子 FM_MT (调整周期与流量预测间隔相同), 其控制方程如下:

$$FA_MT(\bar{l}') = (17 - 2 \times \bar{l}') / 8 \quad (3)$$

$$FM_MT(\bar{l}') = \begin{cases} 15.0/14.0, & [\bar{l}'] = 1 \\ 36.0/35.0, & [\bar{l}'] = 2, 3 \\ 34.0/35.0, & [\bar{l}'] = 4, 5 \\ 32.0/35.0, & [\bar{l}'] = 6, 7 \\ 6.0/7.0, & [\bar{l}'] = 8 \end{cases} \quad (4)$$

调整的基本原则是: 当预测网络的流量水平较高时, FA_MT 与 FM_MT 的取值较小, 以减缓拥塞爆发的程度, 降低数据的丢失率; 当预测网络的流量

水平较低时, FA_MT 与 FM_MT 的取值较大, 以更有效地获取空闲的带宽, 提高系统的吞吐量. 上述两个控制方程使控制因子 FA_MT 与 FM_MT 基本上以 1 为中心对称波动.

3.4 RTT 的估计以及发送速率的调整

当接收端向发送端发送一个 ACK 分组(设为 P_ACK_n)时, 记录下这个发送时间 t_0 , 发送方将根据 P_ACK_n 调整发送速率并进入下一个 $round$ (设为 $round_n$); 当接收端收到属于 $round_n$ 的分组 P_DATA_n 时, 将记录接收时间 t_1 , 则 $t_1 - t_0 - OFFTs$ 为一个 RTT 的取样值 $sampleRTT$, $OFFTs$ 为 P_DATA_n 的发送时间与发送方收到 P_ACK_n 的时间的差值. $OFFTs$ 的信息可由 P_DATA_n 的包头携带. 为了避免单个分组的取样值的波动影响系统的性能, 我们采用 EWMA (Exponential Weighted Moving Average) 过滤器对 RTT 样值进行平滑得到 $SRTT$, 即 $SRTT \leftarrow \alpha \times SRTT + (1 - \alpha) \times sampleRTT$.

当 α 取值较大时, 系统可以获得比较平滑的性能, 但却难以对 RTT 所反映出的网络拥塞状况作出及时反应; 当 α 取值较小时, 可以提高系统对拥塞的反应速度, 但同时会给系统的性能带来较大的波动. 为了获得一个折中方案, 我们取 $\alpha = 0.95$, 并对由 GAIMD 算法直接得到的速率 $rate$ 进行调整. 令 $Rate \leftarrow rate \times \sqrt{SRTT/sampleRTT}$, 目的是当网络趋向于拥塞时 ($sampleRTT > SRTT$) 适当降低发送速率; 当网络趋向于空闲时 ($sampleRTT < SRTT$), 则适当增加发送速率.

调整后的速率 $Rate$ 将通过 ACK 分组通知发

送方, 但在进行新一轮的 GAIMD 计算时, 并不用 $Rate$ 来取代 $rate$, 否则将会影响 RAAR-MT 的 TCP-Friendly 性能.

4 仿真实验

我们以 NS2^① 作为仿真平台, 对 RAAR-MT 协议在具有自相似特性的网络环境中的性能进行综合分析. 我们首先研究对 RAAR 算法使用了多时间尺度的拥塞控制算法后的性能增益, 主要观察 RAAR-MT 算法在丢失率方面以及在网络带宽跟踪能力方面的性能改善. 接着, 我们通过与 TFRC 协议在自相似的网络环境中进行性能对比, 研究 RAAR-MT 在 TCP-Friendly 方面的主要性能指标. TFRC 协议是由发送方实现的基于数学模型的一种比较成熟的 TCP-Friendly 控制机制, 它已被接纳为 IETF 的协议草案, 并向 IESG 推荐为 TCP-Friendly 控制机制的正式标准. 对比实验表明, 在自相似的网络环境中, RAAR-MT 在对 TCP 的友好性、协议内的公平性以及速率波动的平滑性等方面都具有更好的性能.

4.1 仿真环境的配置

实验仿真的拓扑结构如图 3 所示. $R1$ 和 $R2$ 两个路由器之间的链路为瓶颈链路, 带宽为 15Mbps, 其它链路的带宽均为 100Mbps. 瓶颈链路的固定传输时延为 50ms, 两侧接入链路的传输时延之和为 5ms. $R1$ 是唯一可能拥塞的节点, 分组在 $R1$ 的排队时延的变动是引起 RTT 变化的决定因素, 因此 RTT 可以作为衡量系统拥塞程度的度量.

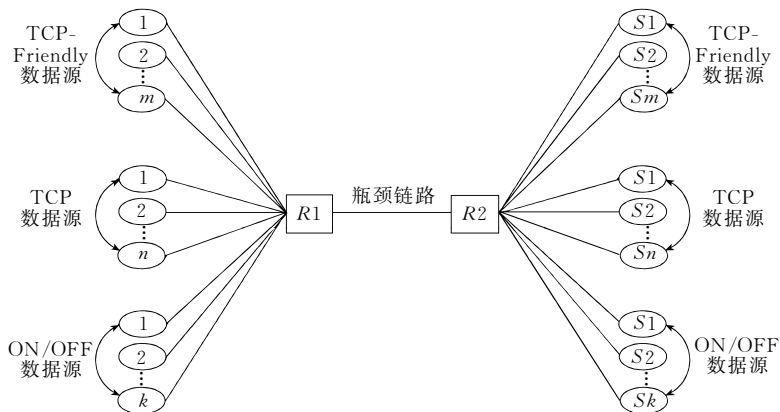


图 3 实验仿真的网络拓扑

我们使用了 64 个 ON/OFF 数据源构造网络系统的背景流, 其 ON/OFF 时间间隔分布均服从参数为 α 的 Pareto 分布, 设 ON 与 OFF 平均持续时间的比值为 1 : 1. Willinger 等人的分析证明^[16], 多个这

种具有重尾分布特性的 ON/OFF 数据源相叠加, 将

① UC Berkeley, LBL, USC/ISI and Xerox PARC. The ns manual (formerly ns notes and documentation), <http://www.isi.edu/nsnam/ns/ns-documentation.html>, 2001.

可以构造出具有自相似流量特征的网络环境,其自相似系数 H 与重尾指数 α 的关系为 $H = (3 - \alpha) / 2$. 假设这个背景流占用网络的大部分带宽,从而使网络的拥塞状况主要受自相似的网络背景流控制. 在仿真实验中,设 ON 期间数据源的发送速率 $On_Rate = 500\text{Kbps}$.

本实验测量的是协议的稳态性能,实验结果为仿真实验后 2/3 时间段内的统计数据. 所有协议流(包括 TCP 流与各种 TCP-Friendly 流)的分组大小都设为 1Kbyte,其开始发送数据时间都从 5~10s 之间均匀分布的随机变量中提取. 一次仿真实验的时间长度为 6000s,取 10 次实验的均值作为最后的实验结果.

4.2 多时间尺度拥塞控制对性能改善

我们分别观察一条 RAAR-MT(或 RAAR)协议流在不同重尾指数 α 的网络流量背景下的丢失率状况. 设 RAAR-MT 与 RAAR 的丢失率分别为 $Loss_{RAAR-MT}$ 与 $Loss_{RAAR}$,定义 RAAR-MT 在丢失率上的性能增益为 $(Loss_{RAAR} - Loss_{RAAR-MT}) / Loss_{RAAR-MT}$,我们分别考察在网络节点使用 Droptail 以及 RED 队列管理策略的情况下,RAAR-MT 在丢失率上的性能增益. 仿真时,取 Droptail 的缓冲区长度为 100 个 packets,对于 RED,其 min_thresh 与 max_thresh 的取值分别为 20 与 100.

从图 4 可以看出,网络背景流的自相似程度越高 ($\alpha \rightarrow 1$),性能增益越大,并且 RAAR-MT 在 Droptail 队列管理机制下,可以获得更大的性能增益. 在 $\alpha = 1.5$ 的典型网络自相似环境下,RAAR-MT 在 Droptail 与 RED 上的性能增益分别约为 20% 与 12%,即使在传统的短相关业务模型下 ($\alpha = 1.95$),RAAR-MT 的性能也不会严重恶化,依然可以取得一定的性能增益. 由于目前业界已公认 RED 队列管理策略的性能远远优于 Droptail,因此,在下面的仿真实验中,我们仅研究系统使用 RED 队列管理策略时的性能.

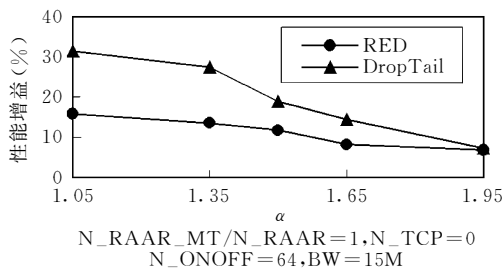


图 4 RAAR-MT 的性能增益

RAAR-MT 在丢失率上的性能增益是通过增强端节点对网络系统可用带宽的跟踪能力而获得

的. 图 5 示出了在不同重尾分布指数 α 的网络流量背景下,RAAR-MT/RAAR 协议流的吞吐量与背景流吞吐量之间的相关系数,测量时间粒度为 1s,使用 RED 队列管理策略. 由于 RAAR-MT/RAAR 协议流与背景流的吞吐量互为消涨关系,当背景流的吞吐量低时,意味着网络的可用带宽较多,被观察的协议流的吞吐量应随之增加,反之亦然. 因此,对于理想的反馈系统,其相关系数为 -1. 从图 5 可以看出,RAAR-MT 对背景流的相关系数比 RAAR 更接近 -1,表明引入多时间尺度的拥塞控制机制后,可以改善端系统对网络带宽的跟踪能力.

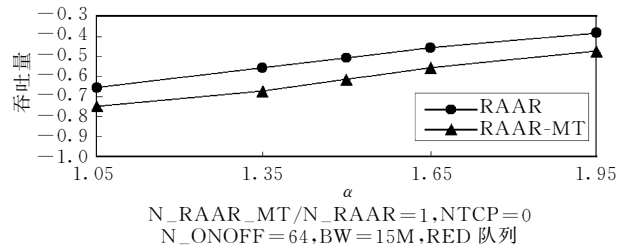


图 5 跟踪能力

4.3 TCP 的友好性

由于 RAAR-MT 是作为一种 TCP-Friendly 协议使用的,这里我们对 RAAR-MT 与 TCP(SACK-TCP)协议公平分享带宽的能力进行研究,网络节点使用 RED 队列管理机制.

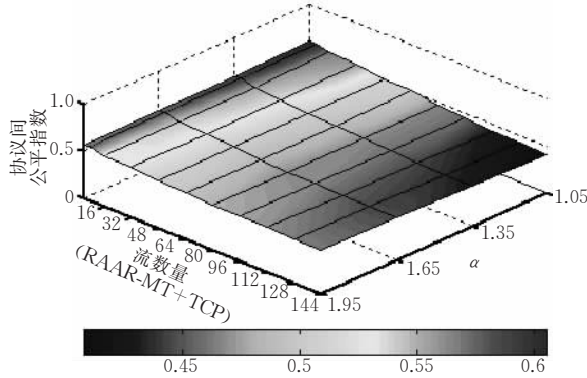
图 6(a)以协议间公平指数 F_p^{inter} 作为两种协议流吞吐量长时(long-duration)公平性的度量,示出了在不同特征背景流、不同数目竞争流的情况下 RAAR-MT 与 TCP 协议流的公平性. F_p^{inter} 定义为 $F_p^{inter} = \frac{R_{H \setminus P}}{R_P + R_{H \setminus P}} = 1 - \frac{R_P}{R_P + R_{H \setminus P}}$,其中 R_P 为待评估的协议流 P 的归一化平均吞吐量, $R_{H \setminus P}$ 为其它与之竞争带宽的协议流的归一化平均吞吐量,当 $F_p^{inter} = 0.5$ 时,我们称它是一种理想的公平(ideal fairness). 由图 6(a)可见,RAAR-MT 对 TCP 的协议间公平指数基本落在 $z = 0.5$ 的理想公平平面附近,因此 RAAR-MT 具有良好的 TCP 友好性.

图 6(b)对在典型的自相似环境中 ($\alpha = 1.5$) RAAR-MT/TFRC 协议对 TCP 的协议间公平指数进行了对比. RAAR-MT 的协议间公平指数在一个大的测量范围内基本上位于理想公平状态附近;而随着竞争流数目的增多,竞争程度的加剧,TFRC 逐渐偏离理想公平的状态.

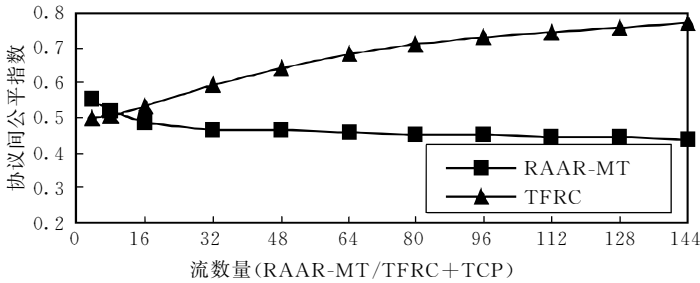
图 6(c)则是利用两个协议流在较小测量时间粒度中的吞吐量等价比(equivalence ratio)值^[18]来

衡量两个协议流(既可以是相同协议,也可以是不同协议)短时的吞吐量公平性.两个协议流 a 与 b 在时间尺度 δ 的吞吐量等价比值函数定义为 $e_{\delta,a,b}(t) = \min\left(\frac{R_{a,\delta}(t)}{R_{b,\delta}(t)}, \frac{R_{b,\delta}(t)}{R_{a,\delta}(t)}\right)$,其中 $R_{f,\delta}(t)$ 表示协议流 f 在时间间隔 δ 内的吞吐量.只有当 $R_{a,\delta}(t)$ 与 $R_{b,\delta}(t)$ 不全为 0 时, $e_{\delta,a,b}(t)$ 才有意义,因此需要去掉两个协议流吞吐量同时为 0 的时间片.时间序列 $\{e_{\delta,a,b}(t_0 + i \cdot \delta)\}_{i=0}^n$ 中各元素的均值 $e_{-r_{\delta,a,b}}$ 即为 a, b 两个协议流

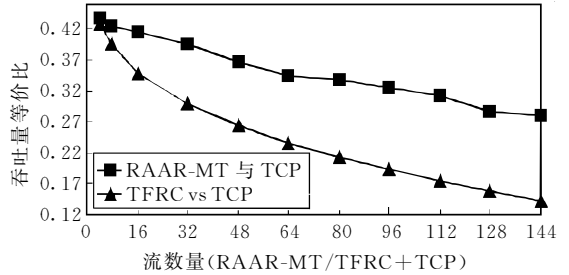
的吞吐量等价比值, $e_{-r_{\delta,a,b}} \in [0, 1]$. $e_{-r_{\delta,a,b}}$ 越接近 1, 表明两个协议流在时间尺度 δ 下越公平.图 6(c) 中的等价比值是 1 条 RAAR-MT(或 TFRC)流与各 TCP 竞争流的吞吐量等价比值取平均的结果,测量时间粒度 $\delta=1s$.从中我们可以看到类似的结果.当竞争流数目较少时,RAAR-MT 与 TFRC 均与 TCP 取得类似的公平性;但随着竞争的加剧,丢失率的增加,TFRC 与 TCP 的公平性迅速下降.



(a) 50%RAAR-MT 与 50% TCP, $N_{ONOFF}=64, BW=16M, RED$ 队列



(b) 50%RAAR-MT(or TFRC)与50% TCP, $N_{ONOFF}=64, BW=15M, RED$ 队列



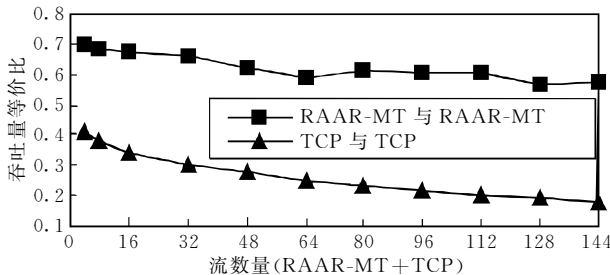
(c) $N_{RAAR_MT}/N_{RAAR}=1, N_{TCP}=0$, $N_{ONOFF}=64, BW=15M, RED$ 队列

图 6 TCP 的友好性

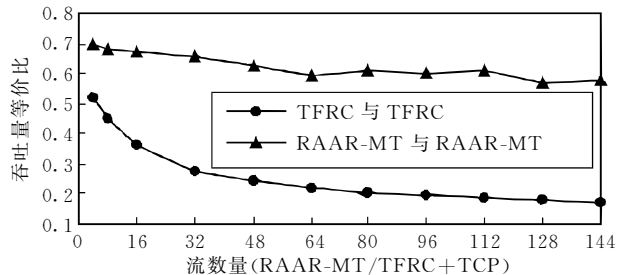
4.4 协议内的公平性(intra-protocol fairness)与平滑性

协议内的公平性与平滑性是传输协议的两个重要性能指标.图 7 用吞吐量的等价比值对 RAAR-MT, TFRC 与 TCP 的协议内公平性进行测量,其等

价比值是 1 条协议流与其它同类协议流的吞吐量等价比取平均的结果;图 8 用吞吐量的方差系数(Coefficient of Variation, CoV)^[17]对协议流的速率平滑性进行评估.协议流 f 吞吐量的 CoV 定义为吞吐量标准方差与均值的比值,即



(a) 50%RAAR-MT 与 50% TCP, $N_{ONOFF}=64, BW=15M, RED$ 队列



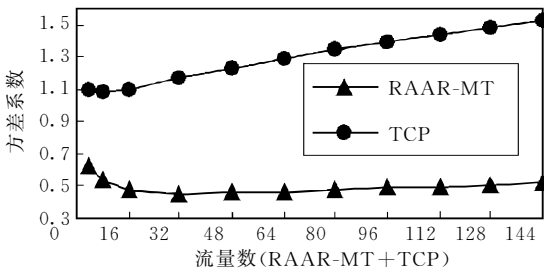
(b) 50%RAAR-MT(or TFRC)与50% TCP, $N_{ONOFF}=64, BW=15M, RED$ 队列

图 7 协议内的公平性

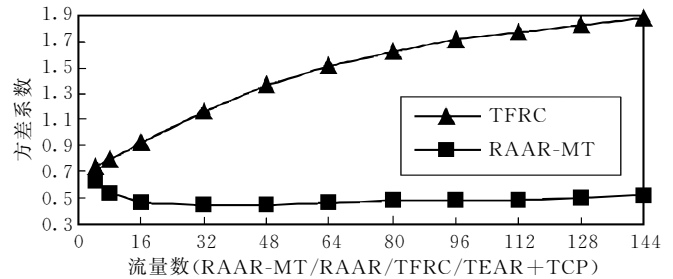
$$CoV_{f,\delta} = \frac{\sqrt{\frac{1}{(T-t_0)/\delta} \sum_{i=1}^{(T-t_0)/\delta} (R_{f,\delta}(t_0 + \delta \cdot i) - R_f)^2}}{R_f}$$

CoV 越小,表示协议流的速率越平滑.所取的测量时间粒度都为 $\delta=1s$.

图 7(a)与图 8(a)表明,在同等环境下 RAAR-MT 比 TCP 具有更好的协议内公平性与速率平滑



(a) 50%RAAR-MT 与 50% TCP, $N_ONOFF=64$, $BW=15M$, RED 队列



(b) 50%RAAR-MT(or RAAR, TFRC, TEAR)与 50% TCP, $N_ONOFF=64$, $BW=15M$, RED 队列

图 8 协议内的平滑性

5 结 论

本文针对自相似的网络条件提出了一种单播的 TCP-Friendly 拥塞控制算法,称为 RAAR-MT.该算法通过在线预测未来流量水平的变化趋势,并利用预测结果在接收端对业务流进行多时间尺度的速率控制.仿真结果表明,在自相似的网络条件下该算法能够增强系统对网络变化的跟踪能力,有效降低业务流的丢失率,即使在传统短相关环境中,其性能也不会出现严重衰减.通过与 TFRC 协议进行对比研究发现,RAAR-MT 协议在对 TCP 的友好性、协议内的公平性以及速率的平滑性等主要的 TCP-Friendly 协议指标都具有更好的性能.在拥塞比较高的自相似环境中,RAAR-MT 的优势更为明显. TFRC 在拥塞程度比较高时的性能缺陷在文献[6, 7]中也有同样的报道,我们认为这与 TFRC 协议本身所选用的 TCP 吞吐量公式有关,事实上要建立一个完善的,适合各种网络条件的 TCP 吞吐量数学模型并不是一件容易的事.由于 RAAR-MT 算法抛弃了每包确认的机制,大大减少了反馈包发送的数量,特别适用于像 ADSL、HFC、卫星系统、无线网络等非对称的通信信道.这种反馈机制有效地防止了反馈内爆(feedback implosion)的发生,便于将来把这种机制应用到组播系统当中.

尽管 NS2 是一个功能强大、配置灵活的通用网络系统仿真平台,但毕竟与实际的网络环境有相当

性,因此 RAAR-MT 比 TCP 更适合作为多媒体的传输协议.图 7(b)与图 8(b)对比了 RAAR-MT 和 TFRC 的协议内公平性与平滑性.结果表明,在相当大的测量范围内,RAAR-MT 的协议内公平性与平滑性均优于 TFRC.在竞争激烈的环境中,RAAR-MT 的优势更为明显.

的差别.我们正在进行将 RAAR 以及 RAAR-MT 协议在 SUN Solaris 系统主机上进行实现的工作,以便在真实的网络应用环境中进一步进行深入的性能测试.此外,目前对网络拥塞程度的可预测性以及多时间尺度下的速率调整的应用范围目前尚处于研究的初步阶段,这种预测与控制机制的实际应用环境以及可适用的业务类型,也有待于通过在真实网络环境进行各种的网络应用实验,收集大量的数据继续进行深入的分析研究,以进一步检验其有效性与合理性.

参 考 文 献

- 1 Floyd S., Fall K.. Promoting the use of end-to-end congestion control. *IEEE/ACM Transactions on Networking*, 1999, 17 (4): 458~472
- 2 Feng W., Kandlur D., Saha D. *et al.*. Stochastic fair blue: A queue management algorithm for enforcing fairness. In: *Proceedings of IEEE INFOCOM 2001*, Anchorage, Alaska, 2001. 1520~1529
- 3 Mahajan R., Floyd S., Wetherall D.. Controlling high-bandwidth flows at the congested router. In: *Proceedings of the 9th ACM International Conference on Network Protocols (ICNP)*, Riverside, California, 2001. 192~201
- 4 Pan R., Prabhakar B., Psounis K.. CHOKe — A stateless active queue management scheme for approximating fair bandwidth allocation. In: *Proceedings of IEEE INFOCOM 2000*, Tel Aviv, Israel, 2000. 942~951
- 5 Hu Yan., Zhang Guang-Zhao.. A stateless active queue management scheme for approximating fair bandwidth allocation and stabilized buffer occupation. In: *Proceedings of the 2nd*

- IEEE Pacific-Rim Conference on Multimedia, Beijing, 2001, 566~573
- 6 Yang Y. R. , Kim M. S. , Lam S. S. . Transient behaviors of TCP-Friendly congestion control protocols. Networking Research Laboratory, Department of Computer Sciences, the University of Texas at Austin; Technical Report TR-2000-14, 2000
 - 7 Rhee J. I. , Ozdemir V. , Yi Y. . TEAR: TCP emulation at receivers - flow control for multimedia streaming. North Carolina State University; Technical Report, 2000
 - 8 Rejaie R. , Handley M. , Estrin D. . RAP: An end-to-end rate-based congestion control mechanism for realtime streams in the Internet. In: Proceedings of IEEE INFOCOM 1999, New York, 1999. 1337~1345
 - 9 Floyd S. , Handley M. , Padhye J. *et al.* . Equation-based congestion control for unicast applications. Computer Communication Review, 2000, 30(4): 43~56
 - 10 Handley M. , Floyd S. , Padhye J. *et al.* . TCP Friendly Rate Control (TFRC): Protocol Specification. IETF RFC3448, 2003
 - 11 Park K. , Willinger W. . Self-Similar Network Traffic and Performance Evaluation. Hoboken, NJ: Wiley-Interscience, 2000
 - 12 Schwefel H. P. . Performance analysis of intermediate systems serving aggregated on/off traffic with long-range dependent properties [Ph. D. dissertation]. Institut für Informatik, Technische Universität München, Germany, 2000
 - 13 Tuan T. , Park K. . Multiple time scale congestion control for self-similar network traffic. Performance Evaluation, 1999, 36/37(1~4): 359~386
 - 14 Park K. , Tuan T. . Performance evaluation of multiple time scale TCP under self-similar traffic conditions. ACM Transactions on Modeling and Computer Simulation, 2000, 10(2): 152~177
 - 15 Yang Y. R. , Lam S. S. . General AIMD congestion control. In: Proceedings of ICNP 2000, Osaka, Japan, 2000. 187~198
 - 16 Willinger W. , Paxson V. , Taqqu M. S. . Self-similar and heavy tails; Structural modeling of network traffic. In: Adler R. J. , Feldman R. E. , Taqqu M. S. eds. . A Practical Guide To Heavy Tails; Statistical Techniques and Applications, Boston; Birkhauser, 1998. 27~53
 - 17 Widmer J. . Equation-based congestion control[Diploma Thesis]. Department of Mathematics and Computer Science, University of Mannheim, Mannheim, Germany, 2000
 - 18 Padhye J. . Towards a comprehensive congestion control framework for continuous media flows in best effort networks[Ph D dissertation]. University of Massachusetts Amherst, Amherst Massachusetts, 2000
 - 19 Hu Yan, Zhang Guang-Zhao, Tu Wang-Qing. RAAR: A TCP-Friendly congestion control mechanism of transporting multimedia traffic in internet. Chinese Journal of Computers, 2003, 26(4): 427~437(in Chinese)
(胡 严,张光昭,张国清. RAAR: 多媒体流在 Internet 上传输的一种 TCP-Friendly 拥塞控制机制. 计算机学报, 2003, 26(4): 427~437)



LIU Yu-Heng, born in 1975, Ph. D. . His research interests include multimedia network communication technology and network traffic control technology.

HU Yan, born in 1970, Ph. D. . His research interests are in network congestion control.

ZHANG Guang-Zhao, born in 1938, professor, Ph. D. supervisor. His research interests are in computer communication networks.

Background

This paper focuses on the study of TCP-Friendly congestion control policy in Internet. In Best Effort network system such as Internet, the congestion control in end system is needed to guarantee the fair service being provided and prevent the congestion collapse due to overload traffic exhausting the network bandwidth resource. Multimedia real-time applications will become the main traffic consuming most available bandwidth of Internet in the near future. However, neither TCP nor UDP can meet the requirements of multimedia real-time transport applications well. Since the dominant Internet traffic is TCP-based, it's very critical to develop new end-to-end congestion control schemes which are suit to multimedia applications and can share bandwidth fairly with TCP flows. TCP-Friendly congestion control policy that can meet the requirements of multimedia real-time applications well is the

major subject in this study group.

This study group has presented several unicast and multicast protocols for TCP-Friendly traffic control in Internet. RAAR, RAAR+ and RAAR-MT are proposed by authors unicast TCP-Friendly congestion control protocols proposed by authors for multimedia applications. MRAAR is layered multicast TCP-Friendly congestion control protocol proposed by authors for multimedia applications. Authors also develop a simulation test platform for TCP-Friendly congestion control protocols in NS2. This paper describes a unicast TCP-Friendly congestion control protocol called RAAR-MT in details, which is based on RAAR protocol and exert multiple time scale traffic control in self-similar network environment to improve traffic control efficiency.