# RWLM 协议在有线信道上的 TCP 友善性研究*

尹冬生，张丰沛，张光昭
（中山大学电子与通信工程系，广东 广州 510275）

摘 要：提出了一种基于接收方的无线多媒体分层组播拥塞控制协议——RWLM。详细描述了 RWLM 协议的拥塞判定机理、RWLM 协议的工作过程以及 RWLM 协议的 TCP-Friendly 拥塞控制机制。重点研究 RWLM 协议在有线信道上的 TCP 友善性。仿真实验显示，RWLM 协议在不同竞争流数目、不同分组大小及不同缓冲区长度的情况下，都具有比较良好的 TCP 友善性。

关键词：分层组播；TCP-Friendly；AIMD（加性加乘性减）；多媒体
中图分类号：TN919.85 文献标识码：A 文章编号：0529-6579（2005）01-0029-05

分层组播（layered multicast）技术[1,2]是目前公认在组播环境中解决网络异构性以及网络动态性的最佳方案，而拥塞控制机制是分层组播的核心问题。为此，本文提出了一种基于接收方的无线多媒体分层组播拥塞控制协议（RWLM），并重点研究了 RWLM 协议在有线信网络环境下 TCP 友善性。做这样的研究是基于以下两点考虑：①无线 Internet 通常以接入网的形式与有线 Internet 相连，无线部分其实是有线部分的补充和外延；②目前，TCP 业务仍然是 Internet 上的主要业务，能否在 Best-Effort 网络上与 TCP 业务流公平地分享网络带宽，是评价一种网络协议性能的重要指标。

## 1 RWLM 拥塞控制机制

### 1.1 RWLM 协议的拥塞判定机理

在无线网络中，分组的丢失有可能是因为拥塞造成的，也有可能是因为随机传输差错造成的。因此，不能将分组丢失作为拥塞发生的判定依据。在这样的情况下，必须通过其它技术手段来判断网络的拥塞水平。为了保证组播系统的可扩展性，RWLM 采用了接收端隐式计算的方法进行拥塞判定。由于端系统通常缺乏必要的技术手段直接观察网络的总体流量水平状况，因此，它只能通过观察与其传输会晤相关的流量参数来间接评估网络当前的拥塞水平。本文有一个基本假定：网络的传播时延以及网络系统的处理时延是基本固定的。这样，分组传输总时延的变化主要由分组在网络中间节点的排队时延来决定，而此排队时延则可以反映出当前网络系统的拥塞水平。因此，接收节点通过对分组排队时延的测量，能够在一定程度上判定拥塞的发生。为了尽量保证拥塞判定的准确性，RWLM 采用了双门限拥塞判定机制判定：基层分组到达间隔门限和 RTT 门限。

1.1.1 门限 1：基层分组到达间隔 RWLM 使用了基于速率的拥塞控制算法，其发送端在每个速率调整周期内均以恒定的速率发送数据分组。基础层是所有 RWLM 接收者都必须接收的一个组播层，因此，“到达间隔”的测量只是针对基层分组。若基层的发送速率为 $r_1$，分组大小为 pktSize，每个分组都带有一个序号，而且该序号随着发送时间的先后以 1 单调递增，那么，序列号相差为 $k$ 的两个基层分组的发送间隔为：

$$\mathrm{snd\_interval}_1(k) = (\mathrm{pktSize\_}\ *8/n)*k$$

假设 $\mathrm{Delay}_{S\to R}$ 为发送端到接收端的网络传播时延与系统处理时延总和（固定值），$\mathrm{Delay}_{\mathrm{queuing}}(i)$ 为第 $i$ 个分组在路由器中的排队时延。那么，第 $i$ 个分组从发送端到接收端的总时延为：

$$\mathrm{Delay}(i) = \mathrm{Delay}_{S\to R} + \mathrm{Delay}_{\mathrm{queuing}}(i) \quad (1)$$

基层分组的到达间隔 $\mathrm{arr\_interval}_1(k)$ 与发送间隔 $\mathrm{snd\_interval}_1(k)$ 的关系是：

$$\mathrm{arr\_interval}_1(k) = \mathrm{time}_R(i+k) - \mathrm{time}_R(i) = \mathrm{snd\_interval}_1(k) + (\mathrm{Delay}_{\mathrm{queuing}}(i+k) - \mathrm{Delay}_{\mathrm{queuing}}(i)) \quad (2)$$

其中，$\mathrm{time}_R(i)$ 为第 $i$ 个分组的到达时间。

* 收稿日期：2004-04-26
基金项目：国家自然科学基金资助项目（90304011）
作者简介：尹冬生（1976 年生），男，博士；通讯联系人：张光昭；E-mail：yyds@21cn.com

从式（2）可知：① 若网络处于非拥塞阶段，基层分组的到达间隔等于其发送间隔；② 若网络处于拥塞的开始阶段，基层分组的到达间隔大于其发送间隔；③ 若网络处于拥塞的结束阶段，基层分组的到达间隔小于其发送间隔；④ 若网络处于正在拥塞阶段，基层分组的到达间隔约等于其发送间隔。

因此，通过在接收端实时测量基层分组的到达间隔，并将其与发送间隔相比较，就能够较好地预测和判定拥塞的发生与结束。

1.1.2 门限 2：来回程时间值（RTT 值） 除了分组到达间隔之外，RTT 值也能够比较直接和有效的反映网络节点的排队时延变化。

（1）若网络处于非拥塞状态，来回程的分组都没有在缓冲区中排队，则

$$RTT = (Delay_{s\to R}) + Delay_{R\to S})$$

（2）若网络处于拥塞状态，来程或回程的分组将在缓冲区中排队，则

$$RTT = (Delay_{s\to R}) + Delay_{R\to S} + Delay_{queuing\_RTT})$$

因此，在网络拥塞时，端系统测量到的 RTT 值将远大于非拥塞时的 RTT 值。

本文使用了闭环与开环相结合的 RTT 测量方法[3]，既在一定程度上保持 RTT 测量的准确度，又能有效防止反馈内爆现象的发生，保证 MRAAR-MT 系统具有良好的可扩展性。

1.1.3 拥塞判定的实现 接收端以 Est_Interval = 12 * snd_interval$_1$ 为估算周期。每个估算周期内，通过测量分组到达间隔 arr_interval$_1$ 和 RTT 的平均值来判定当前网络的拥塞情况。

（1）到达间隔门限的设定

$$thres\_pkt(\delta) = snd\_interval_1 * (1 + \delta)$$

其中，snd_interval$_1$ 为基层分组发送间隔，$\delta$为到达间隔波动系数。若平均分组到达间隔 mean_pkt_interval_ > thres_pkt_ ($\delta$)，接收端认为网络可能发生了拥塞，并进行相应的处理。

（2）RTT 门限的设定

$$thres\_rtt(\gamma_{nLayer,pktSize}) =$$

$$min\_rtt\_ + (max\_rtt\_ - min\_rtt) * \gamma_{nLayer,pktSize}$$

其中，max_rtt_ 是在会晤中接收端测量过的最大 RTT 值，min_rtt_ 是在会晤中接收端测量过的最小 RTT 值，$\gamma_{nLayer,pktSize}$ 为 RTT 波动系数。若平均 RTT 值 mean_rtt_ > thres_rtt_ ($\gamma_{nLayer,pktSize}$)，接收端认为网络可能发生了拥塞。

到达间隔门限和 RTT 门限并不是相互孤立的，而是相互配合、相互补充的，这样才能有效地提高拥塞判定的准确度。其中，到达间隔波动系数 $\delta$ 和 RTT 波动系数 $\gamma_{nLayer,pktSize}$ 的相关设定参考文[3]。

1.2 RWLM 协议的基本工作过程

RWLM 的组播源节点功能非常简单，仅仅需要完成以下的 3 个主要的操作：

（1）对原始的信号进行分层编码（分为 $n$ 层），并以组播方式把各层数据流发送到网络上。

（2）周期性地向基层的数据流插入同步点（SP）信号。这里的SP 与 RLC[4] 中的 SP 略有不同，RLC 需要调节 SP 的出现周期来模拟 TCP 协议的性能，而 RWLM 仅仅利用 SP 点来同步接收节点的加入操作。所以 RWLM 协议只需要在基层加入同步点，而不需要象 RLC 那样在所有的组播层都引入同步点，减轻了组播源节点的负荷。

（3）配合接收节点进行频率密度较低的闭环 RTT 测量，以提高 RWLM 协议的拥塞判定精度。

RWLM 协议绝大部分的拥塞控制操作都由接收节点来完成，从而有效地保证了 RWLM 系统的扩展性。

在开始接收组播会晤时，RWLM 的接收节点首先进入慢启动（Slow-Start）状态，在此状态中接收节点通过加层操作随时间以指数方式增加接收速率，以快速探寻可供合理利用的带宽。当 mean_pkt_interval_ > thres_pkt_ (2$\delta$) 时，接收端才认为网络发生了拥塞。若接收节点检测到网络可能出现拥塞后，则进入拥塞避免阶段的 Steady 状态。

进入 Steady 状态后，系统启动一个长度为 Est_Interval 的计数器，当计数器超时后系统进入拥塞评估（Congestion Estimate）状态，以分析最近的这次估算周期内的网络情况。接收节点算出该周期内基层分组到达间隔及 RTT 的平均值。若它们均小于其相应的门限值，则认为网络没有出现拥塞，系统返回到 Steady 状态；否则，系统将 is_congestion_ 置为 1，但并不立即进行拥塞退避，而是作进一步的分析以提高拥塞判定的准确性。进一步的分析是基于以下两个方面的考虑：①虽然对于高误码率的无线网络来说分组丢失并不意味着拥塞，但是在低误码率无线网络或有线网络中，分组丢失是十分重要的拥塞标志；②Internet 中存在着大量的TCP 流，由于 TCP 是一种基于流控窗口的拥塞控制协议，必然导致 TCP 业务的高突发性。这种高突发性有可能造成路由器缓冲区队列及 RTT 值的暂时增加。因此，系统进一步判断的策略是：

若该估算周期内没有发生分组丢失或者 mean_rtt_ < thres_rtt_ (2 * $\gamma_{nLayer, pktSize}$)（注意：这里的 RTT 门限作了一定的提升，RTT 波动系数设置为正常情况下的 2 倍，即，2 * $\gamma_{nLayer, pktSize}$），系统认为网络的拥塞可能是临时的现象，马上进入拥塞坚持（Congestion Persist）状态；否则，系统认为网络出现了比较严重的、非临时性拥塞，马上进入退层（Drop-Layer）状态，通过离开一个或多个组播层，减缓网络的拥塞。当进入了拥塞坚持（Congestion Persist）状态后，系统将重新启动一个新的估算周期，当 est_interval 计数器超时后，系统进入坚持评估（Persist Estimate）状态并再次分析此估算周期内的分组丢失以及平均 RTT 值。若没有分组丢失或 mean_rtt_ < thres_rtt_ (2 * $\gamma_{nLayer, pktSize}$)，系统确认刚才的拥塞是临时性的，并返回到 Steady 状态；否则，系统进入退层（Drop-Layer）状态。由于离开组播组的操作受 IGMP 离开时延[5] 的影响，因此，系统进入 Drop-Layer 状态后马上转到 Drop-Deaf 状态。在 Drop-Deaf 状态中，系统不对拥塞作出响应。当计时器 $T_{DD}$ 超时时，系统返回 Steady 状态。

当 SP 到来时，处于 Steady 状态的接收节点考察 is_congestion_ 标致位的值以及最近 3 次估算周期内的分组丢失情况。若最近的一次估算周期内网络出现了拥塞（is_congestion_==1）并且在最近的 3 次估算周期中曾经出现过分组丢失，系统进入 Join-Deaf 状态；否则，系统进入加层评估（Join Estimate）状态。在 Join Estimate 状态中，系统仿效 TCP 的 AIMD[6] 控制原则计算出一个评估速率 rate_exp，并与加层速率门限 $Join_{th}$ 作比较。若 rate_Exp ≤$Join_{th}$，系统马上进入 Join-Deaf 状态；否则，系统进入加层尝试（Join-Trying）状态。

### 1.3 TCP 友善的速率控制机制

**1.3.1 RWLM 的速率增加机制** 在 Join Estimate 状态中，系统希望其速率的增加不要超过 TCP 的 AIMD 速率控制原则。因此在 RWLM 中按照 AIMD 原则所评估的 TCP-Friendly 发送速率为：

$$rate\_Exp = B_{nLayerr} + \frac{pktSize}{RTT} * \frac{(T_{now} - T_{last\_rate\_Adj})}{RTT} = B_{nLayer} + \frac{pktSize}{RTT} * \frac{T_{Adj\_interal}}{RTT} \quad (3)$$

其中，$T_{now}$ 为当前的时刻，$T_{last\_rate\_Adj}$ 为上次进行速率调整的时刻，$T_{Adj\_interval}$ 为 2 次速率调整的间隔，$n$Layer 为接收节点当前所接收的层数，$B_{nLayer}$ 为接收 $n$Layer 层数据流的总带宽。

在 RWLM 协议中，加层速率门限被设定为：

$$Join_{th} = \frac{B_{nLayer} + B_{nLayer+1}}{2}$$

其中，$B_{nLayer+1}$ 为增加一层码流后的总带宽。当 rate_Exp > $Join_{th}$ 时，接收节点可以尝试加入更高的一个组播层。而当 rate_Exp ≤ $Join_{th}$ 时，接收节点不能加入更高的一层。

**1.3.2 RWLM 的速率减小机制** 当网络出现拥塞，系统希望能够作出较迅速的反应，通过离开组播层的操作减小接收速率，缓解网络的拥塞状况。

无论系统处于拥塞评估（Congestion Estimate）状态或者坚持评估（Persist Estimate）状态，只要发现最近的一次估算周期内 mean_rtt_ > thres_rtt_ (2 * $\gamma_{nLayer, pktSize}$) 并且发生了分组丢失（is_loss==1），系统就认为网络出现了比较严重的、非临时性的拥塞，马上进入退层（Drop-Layer）状态。在 drop-layer 状态中，RWLM 会针对 2 种不同的情况进行处理：

①当接收节点加入了一个新的组播层后，有一个加层尝试时期。若拥塞发生在加层尝试时期之内，RWLM 认为拥塞是由于加入新的组播层所一起的，系统采取的退避策略是离开刚刚加入的那个组播层，恢复到以前的接收速率。

②若拥塞发生在加层尝试时期之外，系统希望其对拥塞的反应幅度不超过 TCP 的 AIMD 拥塞控制机制的反应幅度。因此，在 RWLM 中取 rate_Exp 为当前速率的 β 倍（β = 0.5），即 rate_Exp = $B_{nLayer}/2$，并令接收节点退出其订阅的最高的 $k$ 个组播层，使得退层后的总接收带宽 $B_{nLayer-k}$ 仅小于 rate_Exp，数学表述为：

$$B_{nLayer-k} <= rate\_Exp,$$

$$B_{nLayer-k+1} > rate\_Exp$$

## 2 仿真实验

RWLM 协议在 NS2.1b8a[7] 仿真平台上进行测试，以评估协议在有线网络环境下的性能。

RWLM 协议对各组播层的速率并没有特别的要求，但鉴于多媒体传输对平滑性的需要，实验中，均采用以下的方式分配各组播层的发送速率：

$$r_i = L_0 + i * \Delta,\ 1 \leqslant i \leqslant Max\_Layer\_ID$$

其中取 $L_0$ = 50 kbps，Δ = 50 kbps，最大的静态层数

Max_Layer_ID＝10。即，基础层的发送速率为 $r_1$ ＝100 kbps，每层速率以 50 kbps 递增，最高层速率为 $r_{10}$＝550 kbps。

RWLM 对组播系统中的路由器所使用的组播路由协议没有特定的要求，实验中使用了 PIM－DM[8] 作为组播路由协议。组播系统的离开时延[9]（leave latency）为 2 s。仿真实验所使用的拓扑结构如图 1 所示。

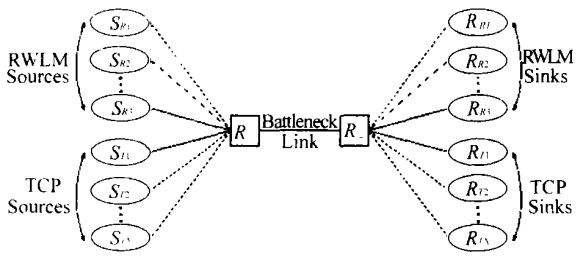


图 1 仿真实验的拓扑结构

Fig. 1 Topology of simulation experiment

$R_1$ 和 $R_2$ 为 2 个组播路由器，它们之间的链路为瓶颈链路，瓶颈链路带宽为 $N$ *450 kbps，其中 $N$ 为系统中协议流的总数（即 RWLM 流与 TCP 流数量之和），其余的链路均被分配了足够的带宽（100 Mbps）。瓶颈链路的固定传输时延为 50 ms，左、右两侧固定传输时延之和为 5 ms。

各链路的误码率均为 0。实验中使用了 SACK 版 TCP，它是 IETF 大力推荐的 TCP 版本。

系统中同时存在 RWLM 与 TCP（SACK TCP）的两种数据流竞争瓶颈链路带宽，所有协议流开始发送数据的时间都在 [5.0, 10.0] 之间均匀分布的随机变量中提取，以减少数据源之间的共振，以求尽快达到稳定状态。为了公平起见，TCP 流的数目与其竞争流的数目相同。为减轻瞬态过程（Transient Phase）对仿真结果的影响，取后 2/3 时间段内的实验数据，仿真实验的时间长度为 1 000 s。

实验 A：实验中，分组大小 pktSize＝500 bytes。RWLM 流的数目 $n$ 设定为 1～10，即，网络中 TCP 流与 RWLM 流的总数 $N$ 为 2～20，以考察在不同竞争流数目下，RWLM 协议的 TCP 友善性。RED 队列的 min_thresh 和 max_thresh[9] 为 5 *$N$ 和 25 *$N$（单位为包）。

图 2a，图 2b 分别表示 RWLM 流对 TCP 流的归一化吞吐量以及 RWLM 对 TCP 的协议间公平指数 $F_P^{inter}$。图 2a 中两者的归一化吞吐量越接近，说明公平性越好；图 2b 中公平指数值越接近 0.5，就说明协议间的公平性越好。实验结果表明，在不同竞争流数目下，RWLM 协议流与 TCP 协议流都能够比较公平地分享瓶颈链路的带宽，表现出比较好的 TCP 公平性。

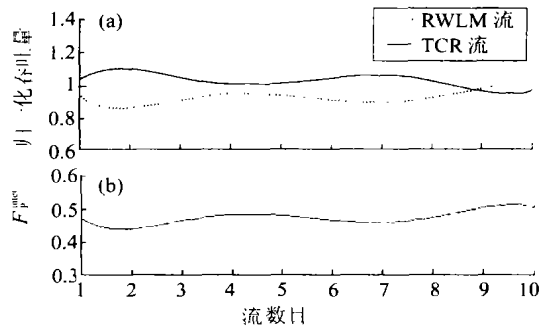


图 2 RWLM 协议流的 TCP 友善性－A

Fig. 2 TCP-friendliness of RWLM-A

实验 B：实验中，RED 队列的 min_thresh 和 max_thresh[9] 为 50 与 250（单位为包），RWLM 流和 TCP 流的数目都为 5。分组大小 pktSize 分别设定为 100、200、300、400、500、600、700、800、900、1 000、1 500、2 000 bytes（RWLM 与 TCP 取相同的分组大小），以考察在不同分组大小情况下，RWLM 协议的 TCP 友善性。

图 3 表明，在不同的分组尺寸情况下，RWLM 协议都能与 TCP 协议流公平地分享瓶颈链路的带宽，表现出比较好的 TCP 公平性。

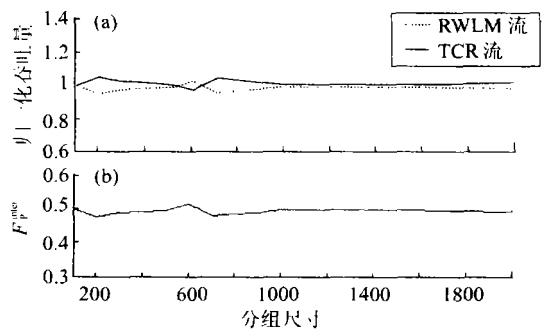


图 3 RWLM 协议流的 TCP 友善性-B

Fig. 3 TCP-friendliness of RWLM-B

实验 C：实验中，分组大小 pktSize 被设定为 500 bytes，RWLM 流和 TCP 流的数目都为 5。RED 队列的 max_thresh 被分别设定为 50、100、200、300、400、500、600、700、800、900 和 1 000（单位为包），且 min_thresh＝0.2 *max_thresh。以考察在不同缓冲区大小的情况下，RWLM 协议的 TCP 友善性。

图 4 的实验结果显示，在不同的缓冲区大小情

况下，RWLM 协议流都能与 TCP 协议流相当公平地分享瓶颈链路的带宽，表现出比较好的 TCP 公平性。

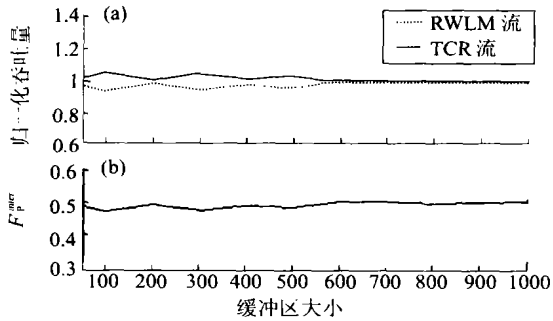


图 4　RWLM 协议流的 TCP 友善性—C

Fig. 4　TCP-friendliness of RWLM-C

## 3　结　论

文中详细描述了 RWLM 协议的拥塞判定机理，接收端状态机，TCP-friendly 的拥塞控制算法等问题。实验结果表明：RWLM 协议在不同竞争流数目、不同分组大小及不同缓冲区长度的情况下，都具有良好的 TCP 友善性。

参考文献：

[1] DEERING S E. Host Extensions for IP Multicasting[C]. Internet Engineering Task Force, 1989: 1112.

[2] DEERING S E. Internet multicast routing: State of the art and open research issues[C]. MICE Seminar, The Swedish Institute of Computer Science, Stockholm, Sweden, Oct. 1993.

[3] YIN D S, ZHANG F P, ZHANG G Z. A wireless layered multicast congestion control protocol for multimedia[C]. Proceedings of the 11th International Multimedia Modelling Conference (MMM2005), 2005(??): 298—303.

[4] VICISANO L, CROWCROFT J, RIZZO L. TCP-like congestion control for layered multicast data transfer[C]. Proc IEEE INFOCOM'98, 1998(3): 996—1003.

[5] CAIM B, DEERING S E, KOUVELAS I, et al. Internet Group Management Protocol, Version 3[C]. RFC 3376, Internet Engineering Task Force, 2002.

[6] YANG Y R, LAM S S. General AIMD Congestion Control [C]. In Proceedings ICNP, 2000.

[7] UC Berkeley, LBL, USC/ISI, and Xerox PARC, The ns Manual (formerly ns Notes and Documentation), Nov. 2001.

[8] ADAMS A, NICHOLAS J, SIADAK W. Protocol Independent Multicast — Dense Mode (PIM — DM): Protocol Specification (Revised)[C]. Internet Draft, Internet Engineering Task Force, draft-ietf-pim-dm-new-v2-03.txt, 2003.

[9] FLOYD S, JACOBSON V. Random early detection gateways for congestion avoidance[J]. IEEE/ACM Transactions on Networking, 1993, 1(4): 397—413.

## Study on TCP-Friendliness of RWLM Protocol over Wired Channel

*YIN Dong-sheng, ZHANG Feng-pei, ZHANG Guang-zhao*

(Department of Electronics &Communication Engineering, Sun Yat-sen University, Guangzhou 510275, China)

**Abstract**: A receiver-based wireless layered multicast congestion control protocol, called as RWLM, is proposed. The congestion detection mechanism, receiver state machine and TCP-friendly congestion control algorithm are described in detail. The TCP-friendliness of RWLM over wired channel is the keystone of this paper. The experiments indicate that under different conditions (e. g. different amount of competitive flows, packet size or buffer size), RWLM achieves excellent performance in TCP-friendliness over wired channel.

**Key words**: layered multicast; TCP-friendly; AIMD (additive increase multiplicative decrease); multimedia