一种基于分组对的分层组播拥塞控制机制*

任立勇**

(电子科技大学计算机科学与工程学院 成都 610054)

【摘要】分析了现有分层组播拥塞控制协议的基本原理，提出了一种基于分组对推测网络可用带宽的分层组播拥塞控制机制PLMCC。其中间节点和接收者利用逐级向下的分组对来推测本地可用带宽，实现对本地可用带宽的准确、快速推测。接收者只需向其父节点发送反馈报文，就可以在最短的时间内获得其允许的最佳接收速率。仿真试验证明，PLMCC不仅具有快速的收敛速度，同时具有良好的协议间公平性和可伸缩性。

关 键 词 分层组播； 分组对； 拥塞控制； 带宽推测

中图分类号 TP393.04

A Packet-Pair-Based Layered Multicast Congestion Control Schemes

Ren Liyong

(College of Computer Science and Engineering, UEST of China Chengdu 610054)

Abstract Layered transmission of data is the general recommended solution to the problem of varying bandwidth constraints in multicast applications. In this paper, several existing layered multicast congestion control protocols are analyzed and their inherent shortcomings are pointed out. Then a novel layered multicast congestion control, PLMCC that is based on packet-pair, is presented. In PLMCC, middle nodes and receivers infer local available bandwidth accurately and quickly by hop-by-hop downward packet-pair. Moreover, each receiver can obtain its optimal throughput in the least time just only sending feedback message to its parent. Simulation results show that PLMCC not only can converge fast, but also is TCP-friendly and scalable.

Key words layered multicast; packet-pair; congestion control; bandwidth inferring

组播技术是一种为节约网络带宽而提出的网络技术，通常应用在一对多(One-to-Many)工作模式的网络应用中。设计组播传输协议面临的根本问题是拥塞控制，分层组播拥塞控制协议将多媒体数据分割成多个层次并发送到不同的接收者。由于充分考虑接收者异构性，提高了网络带宽的利用率，故分层组播协议深受重视。

文献[1]提出了接收者驱动的累加分层组播拥塞控制RLM，将视频数据按其自然属性分割成多个累加层次，并通过不同的组播组发送到接收者。由于不同的接收者可根据其接收链路带宽状况不同而预定不同的层次数，以获取其“最佳”收视效果，因此具有较高的带宽利用率。但由于RLM采用定时器触发其状态的转移，因此RLM收敛到优化速率较慢。另外，RLM的协议间公平性较差，并且RLM的预定操作可能导致大量的分组丢失。文献[2]在RLM的基础上提出了一种新的累加分层

2002年7月1日收稿
* 信息产业部生产发展基金资助项目
** 男 31岁 博士

组播拥塞控制协议RLC。由于RLC采用了按指数分布层次来分割媒体数据(这种方式模拟了TCP的行为)，因此在一定程度上达到了TCP友好。但RLC仍然没有解决收敛速度慢的缺陷，同时，周期性地加入试验可导致大量数据丢失。文献[3]提出了一种基于分组对(Packet-pair)的分层组播拥塞控制策略PLM。PLM采用分组对推测链路带宽，接收者依据各自的推测带宽决定加入或离开某一层。因此PLM具有较快的收敛速度，并且公平性也较好，同时不会出现由于盲目的加入试验而导致分组丢失的情况。但由于端到端的分组对传输易出现丢失的情况，使带宽推测不及时和不准确，从而导致收敛速度较慢和错误的层次加入和离开操作。

下面将提出一种新的基于分组对推测带宽的分层组播拥塞控制机制PLMCC。PLMCC利用逐级(Hop-by-Hop)分组对来推测本地可用带宽，并与直接下游可用带宽汇聚，实现对可用带宽的准确、快速测量。不同的接收者根据其测量的可用带宽，快速、准确逐级向上加入或离开相应组播层次。仿真试验证明，PLMCC不仅具有比PLM更为快速的收敛速度，同时，接收者的吞吐量也更为稳定，不会出现由于误测而导致的错误加入或离开操作。另外，PLMCC还具有良好的协议内公平性和协议间公平性。

1　PLMCC拥塞控制模型

如图1所示，PLMCC采用了接收者驱动模型。PLMCC主要由发送方、中间节点和接收者三种节点组成。其中发送方负责对原始媒体数据进行编码，形成若干个数据层次，并以不同的组播组(Multicast group)向外发送。同时，发送方还定期背对背(back-to-back)地发送分组对(packet-pair)，该分组对包含原始数据的分层信息(如层次数目，每层速率等)。中间节点一般由路由器组成，因此，应具有转发组播报文的功能。同时，中间节点还负责依据接收的分组对信息，推测其上游链路可用带宽，并汇聚下游链路可用带宽以确定本地最佳可用带宽。另外，中间节点还维护下游节点预定的层次信息。接收者依据接收的分组对信息，推测本地可用链路带宽，并加入或离开某一组播层次。

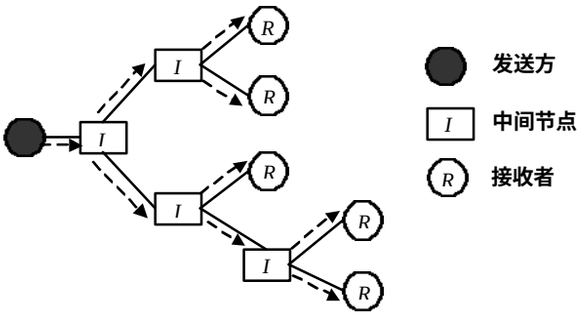

图1　PLMCC拥塞控制模型

PLMCC在PLM的基础上发展而成，但其区别在于，PLM采用端到端的分组对推测链路可用带宽，该方法实现简单，但由于网络链路状况复杂多变，可能导致分组对丢失其中一个或全部丢失，使接收者不能及时判断网络状况，收敛速度降低。另外，对网络带宽的误测而引起错误的加入或离开操作将导致接收质量的振荡。与此同时，PLM要求组播树沿路由器必须实现公平调度，这在目前的Internet中是不现实的。PLM要求在每一个组播层次组中发送探测分组对，不仅是对网络带宽的浪费，而且也给接收者合并不同层次的分组对来推测带宽带来了极大的实现难度。

PLMCC采用了逐级(hop-by-hop)发送分组对的方法，不仅能准确推测本地链路可用带宽，而且逐级反馈链路带宽状况避免了反馈爆炸问题，并能快速收敛。同时，发送方在独立的组播组中发送包含分层信息的探测分组对，不仅节约了网络资源，同时也无需路由器实现公平调度算法。另外，接收者只需向其父节点发送反馈报文，就能快速收敛到其最佳接收速率。

2　基于分组对的分层组播拥塞控制PLMCC

PLMCC在发送方将原始媒体数据编码成多个累加的数据层次，并通过不同的组播组向外发送，中间节点(路由器)和接收者根据接收的分组对PP推测本地链路可用带宽，并据此决定加入(join)或离

开(leave)某一组播层。PLMCC假设路由器除具有组播功能、维护下游节点层次预定信息外，还应能重新定时发送分组对PP。另外，PLMCC对下层的组播路由协议(DVMRP、PIM等)没有特殊要求，只要求到达同一接收者的所有组播层必须经过同一条路径[4]。

2.1 分组对推测网络带宽

对网络拥塞作出正确响应的前提是必须先推测网络的状况(如可用带宽)。现有的方法主要分为两类：一类是根据网络的丢包情况来增加和降低发送速率(如TCP)；另一类是收集接收方的反馈信息(如丢失率，往返时延等)来计算网络的可用带宽(实时媒体单播)。上述方法能比较好地应用在单播(Unicast)环境下，如应用在组播模式下，则有可能导致反馈爆炸[5]。文献[6]提出了发送方利用分组对推测网络带宽的理论。这种基于发送方的分组对在单播模式下能有效推测网络可用带宽，但却存在两个问题：1) 接收者收到分组对后原样返回给发送方，发送方对收到的分组对进行处理，因此在组播模式下基于发送方的分组对可能导致反馈爆炸；2) 现有的网络链路具有不对称性(如卫星链路、xDSL等)，因此这种端到端双向测量方法不能准确反映从发送方到接收者间的可用链路带宽。为此，PLMCC采用单向的基于接收者分组对推测网络带宽的方法，即发送方逐级向下定期发送分组对，接收者收到完整的分组对后，推测本地可用链路带宽为

$$W=\frac{S}{\Delta t} \tag{1}$$

式中 S为分组对的报文长度；Δt表示分组对中两个分组接收时间间隔。这种基于接收者的分组对方法既能过滤双向测量带来的估算噪声，同时也能在路由器发生拥塞丢包前推测出网络的拥塞情况。需要注意的是，式(1)推测出的可用带宽可能会因为偶然原因(如TCP突发业务流等)而造成振荡现象，因此，可采用一个低通滤波器将带宽估算中的高频值滤掉。一种常用的低通滤波器是指数加权滑动平均(*EWMA*)为

$$W_{\text{new}}=hW_{\text{estimate}}+(1-h)W_{\text{old}} \tag{2}$$

2.2 媒体数据分层

在组播应用中，如果发送者以单速率发送数据给所有接收者，在这种情况下，发送者没有考虑接收者的异构特性(处理能力、接收链路等)，不仅不能有效利用网络带宽，而且对某些接收者也不公平(协议内公平问题，Intra-fairness)。由于媒体数据具有自然分层的特点，故提出了在发送方将媒体数据按其自然属性分割若干互不冗余层(layer)，并以不同的组播组发送，而接收者可根据自身的处理能力和链路状况预定若干层，从而达到最佳的收视效果。

假设将媒体数据分割成n层$\{L_1,L_2,\cdots,L_n\}$，其中各层间互不冗余，L_1包含媒体数据里最重要的信息，称为Base layer，而其余各层包含增强上一层质量的信息，即L_i在$L_{i-1}(2\le i\le n)$基础上增强，称为Enhancement Layer。因此，不同的接收者如果预定相同层次数的数据$\{L_1,L_2,\cdots,L_i\}_{i\le n}$，则会有相同的收视效果，并且，预定的层次越多，接收质量就越高。关于媒体数据分层编码算法本文不作讨论，只采用固定层次的分层策略。下面就分发送方、中间节点和接收者三个点来介绍PLMCC的实现细节。

2.3 PLMCC中的发送方行为

PLMCC中的发送方完成将原始媒体数据分层组播和发送探测分组对的任务。发送方并不承担任何拥塞控制的任务，发送方主要有以下功能：

1) 按指定的层次数，利用某一分割算法将原始媒体数据进行分割，并以每层不同组播地址向外组播；

2) 维护一个固定大小的表，记录每一个数据层是否有用户接收，当接收到其直接子节点反馈报文，更新该表(发送方根据该表发送组播报文)；

3) 定时发送探测分组对(Packet-Pair)，定时器间隔D可配置。该分组对记录数据层次数，以及

每层的速率。因此，探测分组对不仅能起到推测网络可用带宽的功能，同时也可供新加入的接收者决定加入的层次信息。

2.4 PLMCC中的接收者行为

PLMCC中，接收者根据接收的分组对信息，推测目前可用链路带宽，然后与当前的接收速率比较，决定是否加入或离开某一个层，主要描述如下：

1) 根据接收的分组对信息(长度、时间)，由式(1)、(2)推测目前链路可用带宽B_e，并计算当前的接收带宽$B_n=\sum_{i=1}^{k}T_i$，其中k是当前接收的最高层次，T_i是第i层的速率；

2) 执行图2所示的伪代码，决定是否加入或离开组播层。因耗时因素，伪代码中的加入和离开操作并不需要执行真正的组播用户加入和离开；

3) 如果上一步中的modified=true，接收者构造反馈报文。其中包含该接收者的编号ID(由父节点分配)、可用带宽$B=B_e$和其预定的层次数$K=k$，然后向其直接父节点发送反馈报文。

```
boolean modified:=false
if Be<Bn then  /*显示网络将出现拥塞*/
    do{
        drop layer k    /*离开预定的最高层*/
        k:=k–1
        modified:=true
    }until Be≥Bn
else
    while Be≥Bn+Tk+1
    {
        add layer k    /*加入下一个组播层*/
        Bn:=Bn+Tk+1
        k:=k+1
        modified:=true
    }
```

图2　接收者的加入或离开伪代码

需要说明的是，上述方法可能导致接收者的可用带宽不能完全利用，尽管可以通过增加层次粒度来解决该问题，但实际上并不能显著提高接收质量，并且上述方法没有用完的带宽可以节约下来供其他业务流(如TCP)使用。另外，当一个接收者新加入该媒体组播时，需要先加入探测分组对所在的组播组。当收到第一个分组对时，推测可用带宽，然后尽可能多地加入若干个组播层。

2.5 PLMCC的中间节点行为

PLMCC的中间节点一般由组播树中路由器组成，因此应具有组播和路由功能。PLMCC对分组调度算法没有特殊要求，既可以采用公平调度，也可以采用FIFO。另外，中间节点还应实现如下功能：

1) 中间节点维护一个带宽预定状态信息表，并记录以下内容：当前本地预定带宽B、预定的组播层次数K，每个下游直接子节点(或接收者)编号ID_i、预定带宽B_i、预定层次数K_i，当接收到下游的反馈报文，更新相应的记录；

2) 接收父节点传来的探测分组对，采用2.4节中方法1)计算B_e和B_n，采用2.4节方法2)计算合适的接收层次K_e。需要注意的是，并不是收到分组对就开始计算，而是让分组对进入正常的发送对列，当发送该分组对的时侯才计算，目的是更准确计算分组对的时间间隔；

3) 计算本地可用链路带宽$B=\min\{B_e,\max_{i=1}^{M}\{B_i\}\}$，本地接收层次$K=\min\{K_e,\max_{i=1}^{M}\{K_i\}\}$，更新带宽预定状态信息表中的记录：$K_i=\min\{K_i, K\}$，其中$M$为下游节点个数；

4) 根据3)计算结果，构造2.4节3)描述的反馈报文，并向上一级发送，其中ID为其上一级节点为其分配的编号；

5) 当收到组播数据报文，按其下游节点预定层次转发报文；

6) 当遇到转发队列中探测分组对中第2个探测分组时(第1个探测分组已被丢弃)，再拷贝一个探测分组，并连续(back-to-back)组播该探测分组对。

通过上述中间节点的带宽计算和汇聚运算，不仅能准确计算本地的可用链路带宽，保证其下游节点能最大限度利用各自网络资源，同时反馈的聚集(Feedback Aggregation)算法能保证避免组播中常见的反馈爆炸(Feedback Implosion)。

3 仿真试验

PLMCC的设计目的是快速收敛、TCP友好和高可伸缩。为验证这些设计目标，本文在网络仿真器NS2中实现PLMCC的原型，并在各种网络环境下进行了大量的仿真试验，限于篇幅的原因，仅给出其中有代表意义的3个试验：仿真试验拓扑结构1如图3所示，用于比较在相同条件下PLMCC和PLM的收敛速度；仿真试验拓扑结构2如图4所示，验证当与TCP流竞争瓶颈链路带宽时对拥塞的响应；仿真试验拓扑结构3如图5所示，验证PLMCC的可伸缩性。

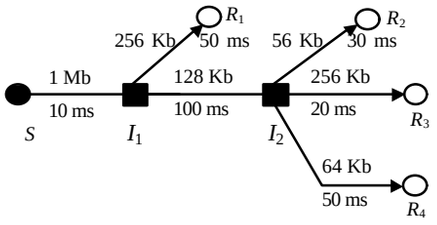

图3 PLMCC与PLM收敛速度拓扑图

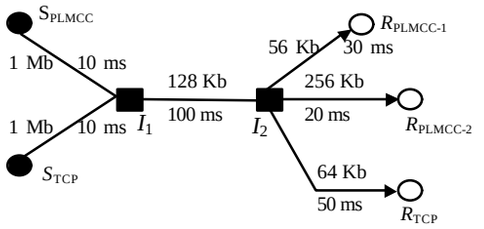

图4 PLMCC的TCP-friendly试验拓扑图

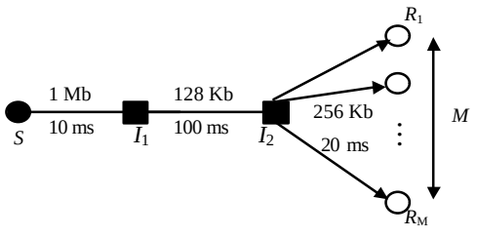

图5 PLMCC可伸缩性试验拓扑图

为简单起见，假定将原始媒体数据分割成30层，每层带宽为10 Kb/s，所有协议的报文长度500 b，PLMCC中分组对发送间隔定时器D=1，PLM的检查定时器C=1。

在试验1中，先启动接收者R_1、R_2、R_3，并在10 s 时启动第4个接收者R_4，试验持续20 s。图6和图7分别为协议环境为PLMCC和PLM情况下的接收者收敛速度对比。从图中看出，PLMCC的接收者能在端到端单向RTT加上从接收者到其直接父节点的传播时间内达到最佳的接收层次，甚至更快，如中间节点I_2在R_4加入组播前已经预定了12层数据，当R_4于10 s加入时，仅需发送一个反馈报文到中间节点I_2，就可以预定其最佳允许接收层次6。相反，如图7所示，尽管PLM也采用探测分组对推测网络带宽，其收敛速度远快于RLM和RLC[1,2]，但由于端到端的分组对传输不仅延迟大，而且PLM的接收者从最低层开始，每次只加入一层，同时，PLM的每次加入和离开层次的操作都是端到端的，这些都是导致PLM收敛速度慢的原因。

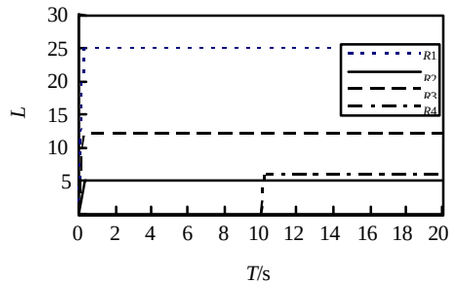

图6 PLMCC的收敛速度

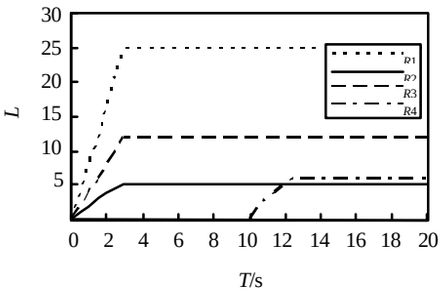

图7 PLM的收敛速度

试验2测试PLMCC对拥塞的响应及TCP友好特性，实验中分别有一个PLMCC发送方和2个PLMCC接收者，在试验开始启动。试验于6 s时加入一个TCP干扰流，假设该TCP流为无限数据流，

并于10 s时结束。如图8所示，可以看出PLMCC能及时响应网络的拥塞状况，同时，当拥塞解除后，能快速恢复其最佳的接收速率。需要注意的，尽管PLMCC流基本上能与TCP流公平共享瓶颈链路带宽，但由于PLMCC接收者能在网络发生拥塞丢包前作出响应，而TCP相反，因此，TCP流略微比PLMCC流多占用少量带宽。

图9给出了PLMCC的可伸缩性试验结果。试验开始时，共有20个接收者，以后每隔10 s加入10个接收者，试验共持续80 s。图中实线是PLMCC的吞吐量变化曲线，虚线是吞吐量的对数趋势图。可以看出，随着接收者个数的急剧增加，PLMCC流的吞吐量变化相当平滑，可见PLMCC的可伸缩性较强。需要指出的是，当接收者个数为320时(曲线为红色)，吞吐量不降反升，这种情况可能是组播树发生了变化，增加的中间节点分担了原有中间节点的开销，从而增加了整个组播会话的吞吐量。

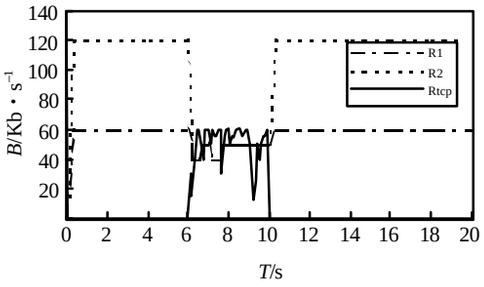

图8　PLMCC的TCP友好特性

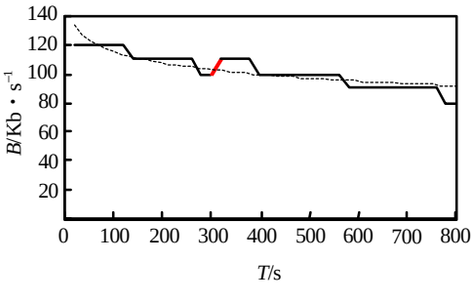

图9　PLMCC的可伸缩性

4　结 束 语

累加分层组播技术在解决组播接收者异构问题方面具有其独特的优势。本文分析了部分典型的分层组播拥塞控制协议的基本原理、存在的问题，提出了一种新的基于分组对探测网络可用带宽的分层组播拥塞控制机制PLMCC，采用逐级探测分组对推测本地网络可用带宽方法，避免了因分组对丢失而造成带宽推测错误的缺点。同时，接收者只需向其父节点发送反馈报文，就可以在最短的时间内收敛到其允许的最佳接收速率。试验结果表明，与PLMCC的设计目标一致。需要指出，层次粒度的选择对组播效率极为关键。另外，当网络发生拥塞时，优先丢弃重要程度较低的高层数据对保证重要数据的正确投递至关重要。

参 考 文 献

1　McCanne S, Jacobson V, Vetterli M. Receiver-driven layered multicast. In SIGCOMM' 96, 1996, 117-130

2　Vicisano L, Rizzo L, Crowcroft J. TCP-like congestion control for layered multicast data transfer. In Proceedings of IEEE INFOCOM, San Francisco, CA, USA, 1998

3　Legout A, Biersack E W. PLM: fast convergence for cumulative layered multicast transmission schemes. In Proceedings of ACM SIGMETRICS' 2000, Santa Clara, California, USA, 2000

4　Celio A, Brett V, Tatsuya S. An end-to-end source-adaptive multi-layered multicast (SAMM) algorithm. In 9th International Packet Video Workshop, New York, USA, 1999

5　任立勇, 卢显良. 一种基于速率组播拥塞控制机制. 电子科技大学学报, 2001, 30(6)：585-589

6　Keshav S. Congestion control in computer networks. PhD thesis, EECS, University of Berkeley, CA 94720, USA, September 1991