区分服务中分层视频组播报文测量和转发算法*

张明杰+, 朱培栋, 卢锡城

(国防科学技术大学 计算机学院,湖南 长沙 410073)

A Packet Metering and Forwarding Algorithm to Support Layered Video Multicast in Differentiated Services Networks

ZHANG Ming-Jie+, ZHU Pei-Dong, LU Xi-Cheng

(School of Computer, National University of Defense Technology, Changsha 410073, China)

+ Corresponding author: E-mail: skyws_hn@yahoo.com, http://www.nudt.edu.cn

Received 2002-12-25; Accepted 2003-06-04

Zhang MJ, Zhu PD, Lu XC. A packet metering and forwarding algorithm to support layered video multicast in differentiated services networks. *Journal of Software*, 2004,15(3):414~420.
http://www.jos.org.cn/1000-9825/15/414.htm

Abstract: Differentiated services (DiffServ) is a scalable architecture for supporting quality of services (QoS), and video multicast is the application which needs network to support QoS guarantee. To accommodate to heterogeneous network and host, it is a good idea to transmit video in a few layers. The paper proposes LVMM (layered video multicast meter) and LVMF (layered video multicast forwarder) algorithms for distribution of the layered video multicast in DiffServ networks. The method needs only one multicast address and its validity is verified using ns-2 simulator.

Key words: quality of services; differentiated services; layered video multicast; heterogeneousness; assured forwarding

摘 要: 区分服务是一种可扩展的服务质量支撑框架,视频组播是对服务质量有较高要求的应用.为了满足端系统的异构性要求,对视频进行分层传输是比较好的方法.研究了使用区服务中的确保服务进行分层视频组播传输的方法,提出了 LVMM(layered video multicast meter)测量算法和 LVMF(layered video multicast forwarder)转发算法.该方法只需要一个组播地址,其有效性通过 ns-2 模拟器进行了验证.

关键词: 服务质量;区分服务;分层视频组播;异构性;确保转发

中图法分类号: TP393 **文献标识码**: A

随着路由器技术和光传输技术的发展,Internet 上的应用种类越来越多,这些应用具有各种各样的服务质量要求.但是,当前的 best-effort 网络模型对应用不提供任何服务质量保证,端系统通过估计网络拥塞程度来调整

* Supported by the National Natural Science Foundation of China under Grant No.90204005 (国家自然科学基金)

作者简介: **张明杰**(1974－),男,天津人,博士生,主要研究领域为网络服务质量,拥塞控制;**朱培栋**(1971－),男,博士,副教授,主要研究领域为网络路由,组播技术,高性能路由器;**卢锡城**(1946－),男,教授,博士生导师,中国工程院院士,主要研究领域为先进网络技术,高性能计算,并行与分布处理.

发送速率.

视频组播对带宽有较高要求,典型的视频组播包括远程会议、视频点播、远程教学等.由于 best-effort 网络的特点以及端系统处理能力和网络本身的异构性,使得以单一速率进行视频组播发布不能满足所有用户的要求[1].为了解决这个问题,研究人员提出使用分层编码[1,2]进行视频组播发布:每一层是对前面层次的强化,不同层在不同的组播组中传输.接收方根据当前网络资源的可用情况加入/退出组播组.用户接收的层越多,视频解码的质量就越高.

虽然分层视频组播较好地解决了接收成员之间的异构性,但是由于 best-effort 网络不提供任何服务质量保证机制,因此,在其中进行视频组播传输必然存在接收质量不稳定、控制比较复杂、组播树变动比较频繁等问题,不适合对服务质量有较高要求的应用.

在网络中如何提供服务质量一直是研究的热点.为了解决 IntServ/RSVP 模型可扩展性差的问题,近几年研究人员提出了 DiffServ 模型[3],该模型把服务质量要求分成几大类,应用可以使用某一类服务传输数据.DiffServ 网络由边缘路由器和核心路由器组成,边缘路由器根据用户要求对其流量进行测量,然后根据测量的结果为报文打上不同的标记;核心路由器根据报文标记执行简单的缓冲管理和调度转发.由于 DiffServ 具有非常好的可扩展性,因此受到了广泛关注.

虽然 DiffServ 网络能够提供比较好的服务质量保证,但是在 DiffServ 网络中进行视频组播传输同样存在异构性问题,异构性产生的原因包括:

(1) 组成员需求的异构性:不同的组成员希望付出不同的费用来获得不同的接收质量;

(2) 端系统处理能力和网络带宽资源的异构性.

基于以上两点,在 DiffServ 中进行视频组播传输同样需要对视频源进行分层编码,以满足接收者的异构性要求.

DiffServ 和组播相结合存在报文如何测量、标记的问题.DiffServ 中已有的标记算法 TSWTCM [4]只适合标记单播报文,不适合异构组播报文标记.因为在单播应用中,接收方只有一个,接收方的要求是确定的,这样就可以在报文的 DSCP(DiffServ code point)中标记报文所属的服务类别及报文的丢弃优先级.但是在组播应用中,由于接收者有多个且可以动态变化,同一个组播报文对一个接收者来说对应 A 服务,但是对另一个接收者来说对应 B 服务,因此边缘路由器无法使用已有的标记算法为组播报文打标记.

为了解决上述问题,本文提出了 LVMM(layered video multicast meter)算法和 LVMFLVMF(layered video multicast forwarder)算法.两个算法不仅较好地解决了异构组播在 DiffServ 网络中的传输问题,而且该方法只需要一个组播地址,既节约了存储开销,也避免了组播树的频繁变化问题.

本文第 1 节介绍 DiffServ 网络中进行分层视频组播传输的网络模型.第 2 节阐述 LVMM 和 LVMF 算法.第 3 节通过实验验证所提算法的有效性.第 4 节介绍相关工作及比较.第 5 节总结全文.

1 网络模型

DiffServ 网络中进行分层视频组播传输的网络模型如图 1 所示.

在 DiffServ 网络中,外部可观察的路由器对每一类报文的转发行为称为 PHB (per-hop-behavior).IETF 区分服务工作组定义了两类 PHB: 加速转发 (expedited forwarding, 简称 EF)PHB[5] 和确保转发 (assured forwarding,简称 AF)PHB[6].

AF PHB 分为 4 类,每一类具有 3 个丢弃优先级.使用 AF PHB 进行数据传输的用户首先向网络预约一定的带宽,边缘路由器根据

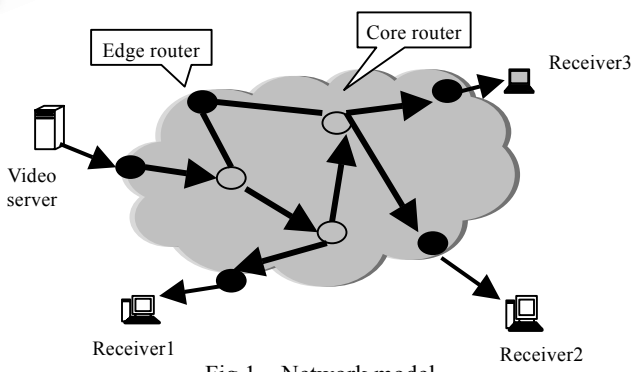

Fig.1 Network model

图 1 网络模型

用户当前的流量以及用户所预约的带宽给报文打上 green,yellow,red 三色,当核心网络发生拥塞时,报文的丢弃概率 $p_{green} \le p_{yellow} \le p_{red}$.

图 1 中的视频组播源具有如下特点:

分层视频编码流用 $(L, r_1, r_2, ..., r_L)$ 表示,其中 L 为编码器的编码层数, r_i 为第 i 层的平均编码速率,设 $R_{\max} = \sum_{i=1}^{L} r_i$.这些层通过一个组播组发送,而不像 best-effort 中使用单独的组播地址传输每一层.

在 DiffServ 体系结构中进行视频组播传输,需要如下技术支持:QoS 组播路由、资源预约以及报文标记和转发.组播 QoS 路由是一个研究热点,已有的组播 QoS 路由算法见文献[7,8],本文不讨论 DiffServ 中的 QoS 路由问题,而是研究在 QoS 路由找到一条满足用户要求的组播树之后,如何进行资源预约和报文标记转发.

下面具体分析为什么 TSWTCM 算法不适合异构组播报文标记.在图 1 中,如果接收者 1 预约的层数为 2 层,接收者 2 预约的层数为 3 层,那么,第 3 层的报文对于接收者 1 而言是高丢弃优先级报文(在预约范围之外),而对接收者 2 而言却是低丢弃优先级报文(在预约范围之内).因此,与视频服务器相连的边缘路由器无法为报文打上丢弃优先级标记.

2 算 法

本文的方法由 3 部分组成:组成员预约过程、边缘路由器测量过程以及核心路由器转发过程.

2.1 预约过程

在 DiffServ 网络中,带宽资源必须在预约成功之后才能使用.由于组播接收者的动态变化和资源要求的异构性,使得基于发送方发起的资源预留不适合组播传输.如果使用传统的 RSVP 进行资源预留会带来 NRS (neglected reservation subtree)问题[9],本文的资源预留过程对文献[9]的方法进行了简单扩充.

当每个接收者在加入组播组时,根据具体情况决定自己要使用的服务种类和需要接收的层数 l 及 l 层所对应的接收速率 $R = \sum_{i=1}^{l} r_i$,然后向资源管理器提出资源预留请求.在资源请求被认可之前,接收者只能使用 LBE(limited best effort,是比 best-effort 级别更低的服务)PHB 接收数据;当接收者的资源请求得到资源管理器认可之后,接收者向分枝路由器传送 $\lambda = R / R_{\max}$ 信息,λ 为预约的接收比例.分枝路由器接收到该预约消息后,把 λ 与对应分枝接口记录在组播路由表项之中,然后把当前输出接口中最大的 $\max\{\lambda\}$ 和 $\max\{R\}$ 沿组播树向上游路由器传送.组播路由表项 *mEntry* 的结构如下所示:

$$(S, G, (OL_1, SL_1, R_1, \lambda_1), (OL_2, SL_2, R_2, \lambda_2), ..., (OL_n, SL_n, R_n, \lambda_n)).$$

其中组的源地址为 S ,组地址为 G, OL_i 标示一个输出接口, SL_i 为对应接口使用的服务级别, R_i 为对应接口上预约的带宽最大值, λ_i 为对应接口上预约的带宽比例最大值.

DiffServ 具有“核心简单”的特点,因此在 DiffServ 体系结构中核心路由器不维护资源预留信息;然而从路由器的角度看,组播与单播最大的区别在于组播是面向流的,核心路由器必须维护组播树信息,即使在 best-effort 网络中也是如此,那么在组播路由表项中加入上述信息之后,本质上没有增加系统的复杂性.从下面的 LVMF 算法中可以看到,路由表项中维护预约比例信息使得报文的重标记过程很简单.

2.2 边缘流量测量算法LVMM

LVMM 算法的思想如下:

组播源首先通知边缘路由器 $(L, r_1, r_2, ..., r_L)$ 信息,边缘路由器把该信息记录在策略表中.同时边缘路由器维护一个测量的速率数组 $r_{avg[i]}$ ($0 \le i \le L$), $r_{avg[i]}$ 记录了由所有小于等于 i 层的报文组成的报文流的平均速率,其中 $r_{avg[0]}$ =0.组播源在发送报文时,把该报文属于的层信息写在 IP 报文头的 TOS(DiffServ 中称为 DSCP)域中.当有组播报文到达时,边缘路由器首先取出层信息,然后更新 $r_{avg[i]}$ 数组.

LVMM 算法.

接收到一个组播报文

Setp 1:

读取报文对应的层数 layer;

for(i=layer;$i \le L$;i++){

计算平均速率 $r_{avg}[i]$;

}

Step 2:

在[$r_{avg[layer-1]}/R_{\max}$, $r_{avg[layer]}/R_{\max}$]取一个均匀分布的随机数 rand;

把 *rand* 记录在报文的 DSCP 域中

计算 $r_{avg[i]}$ 采用时间滑动窗口算法[10].

由于每个接收者所要求的服务类型和带宽比例信息已经记录在组播路由表项之中,这时DSCP失去了标记报文服务类型的作用,因此可以把 *rand* 记录在 DSCP 域中.

2.3 核心路由器转发算法LVMF

LVMF 算法.

对于每一个分枝接口:

If (rand$\le \lambda_i$)

报文送入 SL_i 的 green 队列;

else if (rand$\le C_{SL}/SumR * R_i$)

报文送入 SL_i 的 yellow 队列;

else

报文送入 SL_i 的 red 队列;

其中,C_{SL} 为链路分配给对应服务类的带宽,*SumR* 为对应服务被预约的带宽总量.*rand* 记录在报文的 DSCP 域中,R_i,λ_i 记录在路由表项中,$C_{SL}/SumR$ 也是预计算好的.

算法中报文进入 yellow 队列的比例与一个组播组在该输出链路上预约的带宽成正比,从而保证了对剩余带宽的公平共享.

2.4 算法性能分析

对于 LVMM 算法,复杂性主要来自于 for 循环的次数.为了分析方便,设第 i 层在单位时间内到达的报文数为 n_i,则单位时间内 for 循环的次数为

$$N_m = \sum_{i=1}^{L}(L+1-i)*n_i \tag{1}$$

由于分层视频编码的模式各不相同,下面具体分析指数递增分层编码模式下 LVMM 算法的复杂性.指数递增分层编码模式[1,2]具有如下特征:

$$n_i = 2^{i-1}*n_1 \quad (1 \le i \le L) \tag{2}$$

把式(2)带入式(1)

$$\begin{aligned} N_m &= \sum_{i=1}^{L}(L+1-i)*2^{i-1}n_1 \\ &= (1+2+4+...+2^{L-1})*L*n_1 - (2+2*4+3*8+...+(L-1)*2^{L-1})*n_1 \\ &= (2^L-1)*L*n_1 - 2*[(L-2)*2^{L-1}+1]*n_1 \\ &= (2^{L+1}-L-2)*n_1 \end{aligned} \tag{3}$$

考虑一个单播应用流,如果在单位时间内到达的报文数为 $\sum_{i=1}^{L}n_i$,那么,TSWTCM 需要计算 $\sum_{i=1}^{L}n_i$ 次平均速率,而且有

$$N_u = \sum_{i=1}^{L}n_i = \sum_{i=1}^{L}2^{i-1}*n_i = (2^L-1)*n_1 \tag{4}$$

比较式(3)和式(4)

$$\eta = \frac{N_m}{N_u} = \frac{(2^{L+1} - L - 2) * n_1}{(2^L - 1) * n_1} < 2 .$$

从上面的分析得知,LVMM 算法对于指数递增的层次编码测量平均速率的次数小于具有相同流量的单播测量平均速率次数的 2 倍.

需要指出的是,LVMM 算法在最接近视频源的路由器上运行,由于这样的路由器中流的数目比核心路由器中少得多,因此该算法不会带来可扩展性问题.

因为 LVMF 算法要在核心路由器中实现,因此算法的复杂性要求比 LVMM 算法要苛刻.在单播情况下,报文进入队列(green,yellow,red)由 DSCP 域决定,LVMF 算法相对于单播报文操作只是在把报文放入相应队列之前增加至多两次比较操作和一次乘法操作,这种操作既可以用 VLSI 高效实现,也可以使用高性能网络处理器来实现.除了运算开销之外,核心路由器转发算法需要在组播路由表项中为每个输出接口增加 6 个字节的存储开销,用来表示对应接口的服务类别(1 个字节)、预约的比例(1 个字节)以及预约的速率(4 个字节).由于存储器的容量越做越大并且价格不断下降,因此 6 个字节的开销是可以接受的,而且本文方法只需要一个组播地址,节省了 $L * size(mEntry)$ 倍存储开销.

3 ns-2 模拟实现及验证

本文提出的方法在 ns-2[11]上进行了验证.

3.1 模拟拓扑

模拟拓扑如图 2 所示.其中 us 为单播源(unicast sender),ms 为组播源(multicast sender);E 为边缘路由器,C1,C2 为核心路由器;R1~R3 为接收点.

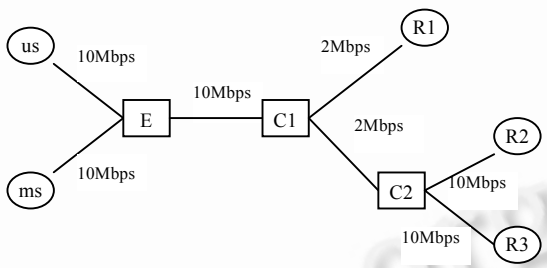

Fig.2 Network topology

图 2 网络拓扑

3.2 模拟过程描述

0s~20.0s:从 us 向 R1 发送流量为 1.6M 的 CBR 报文,预约带宽为 1.5M;

0s~20.0s:从 us 向 R3 发送流量为 1.1M 的 CBR 报文,预约带宽为 1.0M;从 us 发送的 CBR 作为背景流量存在.

0s~20.0s:从 ms 发出的分层视频组播流分为 5 层,每层的发送速率为 2^m×32Kbps(m=0…4),编码模式与文献[1]相同;由于 LVMM 使用基于时间滑动窗口的速率测量方法,因此可以平滑视频流中发生的突发.

3.0s~20.0s:R2 加入组播组,可以接收 5 层信息,预约比例为 100%;

5.0s~20.0s:R1 加入组播组,由于带宽不足,因此 R1 只能接收 4 层信息,预约比例为 50%.

3.3 模拟结果

图 3~图 6 分别显示了 R1 和 R2 获得的带宽随时间变化的情况.从图 3 和图 4 可以看出,R1 成功接收了 1~4 层的信息,第 5 层的视频信息被丢弃.从图 5 和图 6 可以看出,R2 成功接收了 1~5 层的视频信息.综合上面的结果,每个接收者都得到了其预约的带宽,满足了异构性要求.

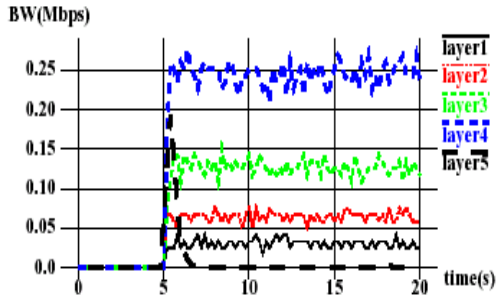

Fig.3 Bandwidth of each layer received by R1

图 3 R1 获得的各层带宽

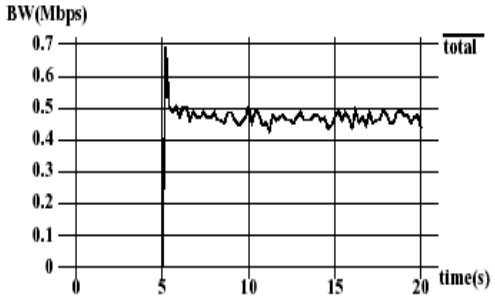

Fig.4 Total bandwidth received by R1

图 4 R1 获得的总带宽

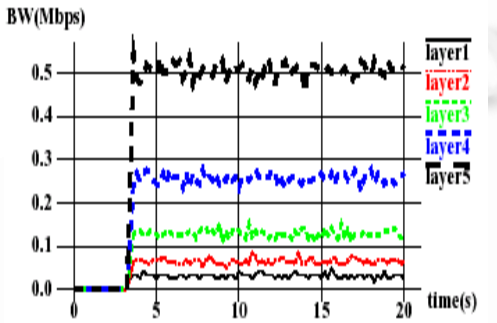

Fig.5 Bandwidth of each layer received by R2

图 5 R2 获得的各层带宽

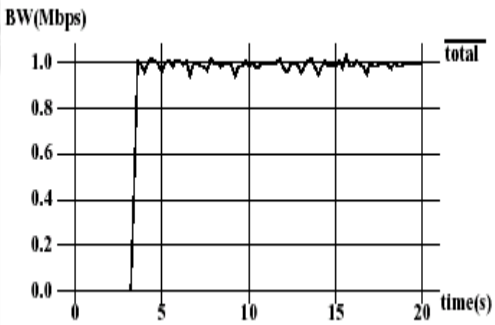

Fig.6 Total bandwidth received by R2

图 6 R2 获得的总带宽

4 相关工作及比较

文献[9]最先针对区分服务组播提出了异构组播问题,并给出如下解决方法:

在组播路由表中存储分枝接口的服务级别,当报文复制时,取出该服务值复制到报文的 DSCP 中,这样,不同的接收者可以接收不同的服务级别.文中提到了具体服务质量参数的不同更增加了问题的复杂性,但文中没有给出解决办法.本文研究具体服务质量参数(预约带宽)的异构性问题,该问题更有实际意义.

到目前为止,区分服务中支持组播的研究工作不多,除了文献[9]的工作,文献[12]提出了 DSMCast 区分服务组播框架,该框架让边缘路由器维护路由信息,核心路由器只进行报文复制和转发,并且针对该框架提出了组成员加入/退出算法.DSMCast 方法让报文携带组播树和服务级别信息,类似于源路由,该方法的缺点是,当组播树很大时(如密集模式),或者报文比较小时,每个报文携带的额外信息相对太多造成了有效带宽的利用率低.

文献[9,12]的工作只是针对接收者预约的服务级别不同,对于接收方预约带宽异构的情况没有考虑或没有给出解决方法,而预约带宽不同是很实际的问题.本文具体分析了具有带宽异构性要求的应用实例——分层视频组播如何在 DiffServ 网络中进行高效传输问题.

由于用户预约的带宽不一样,因此必须解决报文测量和转发问题.本文借鉴了文献[9]的思想并且对其进行了扩充,在组播路由表项中不仅记录了用户预约的服务级别,而且记录了接收者预约的带宽信息,从而满足了用户的带宽异构性要求.

5 结 语

本文提出了在 DiffServ 网络中进行分层视频组播传输的机制.该机制主要由 LVMM 和 LVMF 两个算法组成.该方法简单、易于实现,较好地解决了 DiffServ 网络中进行分层视频组播的异构性问题.而且与 best-effort 中使用的方法比较,本文的方法还有一个特点,就是只使用了一个组播地址.这样既节省了路由表存储空间,又避免了组成员加入/退出组播组带来的处理开销.

最后指出一点,本文的方法同样适用于非分层的组播应用,对于非分层的组播应用而言,相当于 L=1 的情况,此时接收者向网络预约某种服务并且指出其预约的接收比例 λ_l =1.

References:

[1] McCanne S, Jacobson V. Receiver-Driven layered multicast. In: Proc. of the ACM SIGCOMM'96. New York: ACM Press, 1996. 117~130.

[2] Gopalakrishnan R, Griffioen J, Hjálmtýsson G, Sreenan CJ, Wen S. A simple loss differentiation approach for layered multicast. In: Sidi M, ed. Proc. of the IEEE INFOCOM. Tel Aviv: IEEE Communications Society, 2000. 461~469.

[3] Blake S, Black D, Carlson M, Davies E, Wang Z, Weiss W. An architecture for differentiated services. RFC 2475, Internet Engineering Task Force, 1998.

[4] Fang W, Seddigh N, Nandy B. A time sliding window three colour marker (TSWTCM). RFC 2859, Internet Engineering Task Force, 2000.

[5] Jacobson V, Nichols K, Poduri K. An expedited forwarding PHB. RFC 2598, Internet Engineering Task Force, 1999.

[6] Heinanen J, Baker F, Weiss W, Wroclawski J. Assured forwarding PHB group. RFC 2597, Internet Engineering Task Force, 1999.

[7] Faloutsos M, Banerjea A, Pankaj R. QoSMIC: Quality of service sensitive multicast Internet protocol. In: Proc. of the ACM SIGCOMM'98. New York: ACM Press, 1998. 144~153.

[8] Chen S, Nahrstedt K, Shavitt Y. A QoS-Aware multicast routing protocol. In: Sidi M, ed. Proc. of the IEEE INFOCOM. Tel Aviv: IEEE Communications Society, 2000. 1594~1603.

[9] Bless R, Wehrle K. IP multicast in differentiated services networks. Internet Draft, Internet Engineering Task Force, 2000.

[10] Clark D, Fang W. Explicit allocation of best effort packet delivery service. IEEE/ACM Trans. on Networking, 1998,6(4):362~373.

[11] Ns-2 Network simulator. http://www.isi.edu/nsnam/ns

[12] Striegel A, Manimaran G. A scalable approach to diffServ multicasting. In: Neuvo Y, ed. Proc. of the IEEE Int'l Conf. on Communications (ICC). Helsinki: IEEE Communications Society, 2001. 2327~2331.