

MPLS 网络环境下组播应用研究^{*})

夏利¹ 杨宏¹ 刘淑霞² 王光兴¹

(东北大学计算机信息科学与工程学院 沈阳 110004)¹ (青岛理工大学计算机工程学院 青岛 266033)²

摘要 组播与 MPLS 是两种互补的技术,将二者结合起来,对提高网络性能、解决组播的可扩展性问题及控制开销等有着深远的意义。本文首先简要介绍了 MPLS 和组播的基本原理及优点,接着概括了 MPLS 网络环境下组播技术应用及实现技术,最后讨论了发展趋势和进一步研究的方向。

关键词 MPLS, 组播, 服务质量, MPLS 组播树, 组播标记交换路径

A Study of Multicast Application over MPLS Networks

XIA Li¹ YANG Hong¹ LIU Shu-Xia² WANG Guang-Xing¹

(College of Information Science and Engineering, Northeast University, Shenyang 110004)¹

(Qingdao Technological University, School of Computer Engineering, Qingdao 266033)²

Abstract Multicast and MPLS are two complementary technologies. Merging these two technologies will enhance performance and present an efficient solution for multicast scalability and control overhead problems. The paper first introduces the principle and advantages of MPLS and multicast, and surveys the multicast application and implementation over the MPLS networks. Finally, the trends and several issues for more research are presented.

Keywords MPLS, Multicast, QoS, MMT, Multicast LSP

1 引言

近年来,随着 Internet 的发展和普及,网络应用趋于多元化,同时对网络的带宽和多媒体化提出了越来越高的要求。带宽的增加使得许多高带宽的应用,如远程教学、视频会议、网络集体游戏、视频点播、分布式机群的数据同步等成为可能。这些应用又带来了新的带宽的急剧消耗,从而加剧了网络的拥塞。这些高带宽应用的典型特点就是高带宽和群发。解决这类问题的一个非常好的思路就是发展 IP 网络的组播技术,并且要在支持组播技术的网络平台上满足 QoS 要求,而 MPLS 正满足了这一要求^[1,2]。因此,如何在下一代骨干网的核心技术 MPLS 上实现组播业务是亟待解决的问题。

2 MPLS 与组播技术概述

2.1 多协议标签交换

多协议标签交换(MPLS)是一个可以在多种第二层媒质上进行标签交换的网络技术。它结合了第二层交换和第三层路由的特点,将第二层的基础设施和第三层的路由有机地结合起来,在 MPLS 网络的边缘对收到的分组进行分类,按照分类的结果给分组加上一个定长的标签该标签将对应于该分组的处理方式(使用的路由、资源预留、业务量参数、业务服务等级等)有关的一切信息。MPLS 网络不再查路由表,而是依据标签的内容进行转发并进行相关处理。MPLS 的价值在于支持多协议混合的网络,能够在提供 IP 业务时确保 QoS 和安全性,实现流量工程^[1~6]。因此, MPLS 技术将成为下一代骨干网的关键技术。

2.2 IP 组播

IP 组播技术主要由控制用户加入/离开组播组的 Internet 组管理协议(IGMP)和构造路由器间组播路径的组播路由协议来支持实现^[7]。IGMP 协议是一种注册协议,主要完成组播用户组的管理。利用 IGMP 协议,组播路由器可以判断在与它直接相连的网段中是否存在组播组的成员。组播路由协议主要用于实现组播路由器的组播路由信息传递,它有以下儿种:DVMRP(距离向量组播路由协议)、MOSP(组播开放最短路径优先)、CBT(基于核的树)和 PIM(协议无关的组播)。

在组播网络中,通过组播分布树来传播组播信息。组播分布树有两种:源树和共享树。源树采用记号(S,G)来标识,S 指源地址,G 反映组播组地址。它以组播源为树根,形成到各个接收端的组播分布树,每个单独的组播源都有各自独立的 SPT 树。共享树采用记号(*,G)标识,以某些集合点为根,形成到各个接收端的组播分布树。所有的组播源都使用相同的共享树。

2.3 MPLS 网络环境下组播应用面临的问题

在 MPLS 网络中实现组播业务是必须的,并且基于 MPLS 的流量工程也有能力为组播通信提供 QoS^[3~6]。然而,在 MPLS 域上提供组播应用,需要将第三层的组播树映射到第二层的标签交换路径上,由此产生下列问题^[6]:

(1) 泛洪/剪枝问题。由于组播路由协议的泛洪/剪枝特性,导致组播树的不稳定性和 L3 层路由的流量驱动性,这就需要与路由相关的 LSP 也以流量驱动的方式建立和拆除。如果一个 LSR 不支持 L3 层转发,这个流量驱动的属性甚至需要上游标签交换路由器发起创建 LSP。

(2) 汇聚问题。在 MPLS 域内,单播情况下不同的目的

^{*})基金项目:国家 863 计划(708-4-5)和辽宁省自然科学基金(项目编号 20042012)。夏利 副教授,博士研究生,主要研究方向:网络 QoS 理论及路由算法研究。王光兴 教授、博士生导师,主要研究方向:第三代移动通信,ATM 网络管理。

地址因具有局部共性而在入口边界路由器产生一个等价类,具有相同转发等价类 FEC 的单播流被映射到相同的 LSP,以此获得单播流的汇聚,而如何将具有不同“目的地址”的组播汇聚到一条 LSP 上,需要进一步地研究。

(3)RPF 检查问题。组播树是利用组播路由协议建立起来的,一些组播路由协议是基于组的源端与接收端的 RPF (反向路径树)建立起来的。但是,RPF 是基于网络中路径的对称性,当引入路径约束时,就无法保证利用链路的对称性,那么 RPF 转发将会导致数据在子优化路径上转发,甚至接收端无法从一些(或全部)源端接收数据。

(4) L2 层交换与 L3 层转发共存问题。有两种情况仅靠第二层的标签不能确定出口的标签:第一种情况是当某个中间 LSR 同时是一棵共享树和一棵源树的组播成员,它对某个目的组 G 就同时拥有两种转发状态(S,G)和(*,G);第二种情况是标签分配不当,比如某个组播流分配的标签和一个单播流的标签相同,那么在组播树的出口边界路由器上进行标签分离。我们发现,在这两种情况下,仅靠第二层的标签交换是不够的,需要涉及三层路由,而 MPLS 网络的重要特征是中间 LSR 只进行标签交换,不进行第三层路由。

(5)单向/双向树问题。双向共享树(CBT 协议)会产生大量的合并点,不利于建立 LSP。单向共享树中要使两个源能够发送到存在交叉方向的目的,必须使用隧道方式,将流量全部发送到共享树的根节点(RP)。

(6)组播数据的封装问题。单向共享树的源以及双向共享树的非成员源在往组播组成员发送数据时,要先把组播数据进行封装,然后发送给共享树的根节点,由根节点来进行解封装。这种封装和解封装的操作要 L3 来处理。所以,当存在这种操作时,是无法建立起端到端的 LSP 的。

(7)环路问题。依靠单播路由协议的组播路由协议和单播协议一样会遇到传输环路问题,并且在组播情况下,环路的影响将比在单播情况下更为严重。因为一旦发生环路,在每个分支节点会产生大量的环路数据,给网络带来更大的负载。

3 MPLS 环境下实现 IP 组播业务的方法

近年来,关于 MPLS 网络环境下组播应用技术,学者们已经提出大量解决方案^[8-18]。本文根据这些方案的侧重点不同,大致将其分为四类:(1)建立 LSP 的触发方式;(2)LSP 的设置;(3)组播流的汇聚问题;(4)组播树的创建。下面仅对其中典型的方案进行分析。

3.1 建立 LSP 的触发方式

MPLS 中,建立 LSP 的触发方式主要有拓扑驱动、流量驱动、请求驱动三种模式^[6]。所谓流量驱动是指 LSR 在收到第一个数据报后启动建立 LSP 过程。对于采用泛洪/剪枝模式的组播路由协议,路由是流量驱动的,所以 LSP 的建立也需要采取流量驱动的模式。尽管有的方案采用请求驱动的方式^[13],但都是以建立额外的管理实体为代价的。更重要的是,采用流量驱动建立组播 LSP 还有三个好处:可以减少建立组播 LSP 的控制开销;由于仅当有数据流时才需要标签,所以可以节省标签,这对于标签空间有限的 LSP 来说非常重要;由于流量驱动时数据流的特性会被第三层协议看见并分析,易于实现组播流量工程和区分服务。

文[15]提出了流量驱动分配组播标签的方案,该方案分为两类:上游主动模式和下游按需模式。在上游主动模式下,当支持组播的路由器检测到一个新的组播 FEC(没有分配标

签或带有未绑定标签)或者在流量驱动之后又增加了 L3 组播树分枝时,启动 L3 路由来决定出端口,然后为每个出端口选择一个标签与组播树绑定,并在转发表中增加一个由该组播 FEC 和所选择的所有标签以及端口构成的下一跳标签转发条目。这样,在第一个组播分组到达各出口边界 LSR 时就建立了完整的 LSP 树,以后的组播数据分组都通过标签进行转发。标签回收时的方法与分配时基本一致,但必须保证下游 LSR 先释放标签。该方案可以快速建立起 LSP,缺点是标签分配和回收方案复杂,很难与单播的标签方案取得一致。

在下游按需模式下,当数据分组在沿着 L3 层组播树到达 LSR 之后,LSR 开始向上游分配标签。在收到第一个数据分组后,如果采用独立 LSP 控制模式,LSR 可以立即向上游分配标签;如果采用有序的 LSP 控制模式,LSR 必须等到至少有一个组播树上的下游 LSR 向它分配标签之后再向上游分配标签。上游主动模式相对于下游按需模式来说,L3 层转发时间短。标签分配协议可采用 LDP 的子集。标签回收不需要信令消息来进一步协调,L3 层的泛洪/剪枝消息维护了 L3 层的组播树,从而也维护了对应于 L2 层的 LSP 组播树。该方案容易与单播的标签分配方案一致,标签回收简单,缺点是 LSP 建立的滞后时延长。文[13]便是采用下游按需模式标签的分配,由接收端路由器发起 LSP 的建立。

3.2 MPLS 网络环境下组播 LSP 的设置

组播树结构要求建立点对多点(P2MP)的 LSP,甚至是多点对多点(MP2MP)的 LSP,而在原 MPLS 框架中,只是解决了点对点(P2P)的 LSP^[1,2]。为了在 MPLS 网络中实现组播业务,需要为 P2MP 和 MP2MP 设置 LSP。而在文[9]提出的方案中,通过边界路由器组播机制,即通过将组播组的源和分支节点限制在 MPLS 网络的边界路由器上,将组播流在边界路由器处转变为多个单播流,从而简化 LSP 的建立。由于树的分支节点在边缘 LSR 上,所以不必创建和管理点对多点(P2MP)、多点对多点(MP2MP)的 LSP。相反,树可分解并映射成多个 P2P 的 LSP。该方案的缺点是必须将组播的源及分支节点限制在 MPLS 域的边界路由器上,丧失了通用性。

(1)P2MP 的 LSP 设置。对于 P2MP 的 LSP 设置方法的研究到目前为止还没有很好的方案提出,不过文[14]中详细说明了将 P2MP 应用于 MPLS 流量工程构架下解决方案的功能性需求,给出了设置 P2MP 流量工程(TE)/MPLS LSP 需要满足的条件。尽管它没有给出建立 P2MP LSP 的明确细节,不过定义了建立 P2MP LSP 规范,为今后的研究指明了方向 and 标准。

(2)MP2MP 的 LSP 设置。文[13]给出了为建立单向组播共享树而设置的 MP2MP LSP 框架,提议在网络中采用集合点 RP(功能与 PIM-SM 中的 RP 类似)方法,组播会话的发送端必须向 RP 注册且与 RP 建立单播 LSP,加入会话的接收端必须向 RP 发送加入请求消息。该方案利用多个 RP 来实现 RP 故障恢复和提供负载均衡,同时只需要在 RP 处添加一定的管理功能,就可利用现存的 MPLS 技术和路由协议来实现 MPLS LSP 的设置。将该方案加以扩展,还能用于 MPLS 域间 LSP 的设置。

3.3 MPLS 网络环境下组播汇聚

对于汇聚问题,Aiguo Fei 等人提出了汇聚组播方式^[9]。该方案引入了汇聚树的概念,在核心网络,多个组播会话可以共享一个汇聚树,而不是为每个组播会话分别建立组播树,从

而实现在核心网络处减少组播状态和控制开销。它通过隧道技术、转发状态汇聚和利用网络传输组播的透明性等技术将组播的复杂性放在端点处解决。如果通过某个域的不同组的源和出口节点都是汇聚树上的节点,那么来自特定组的数据分组在源节点被封装。在出口节点被解封装,对于核心路由器来说,只需为汇聚树管理一个单一的转发实体,而无需考虑有多少组共享。该方法用树管理器来处理树的管理和组播组与汇聚树之间的匹配问题。树管理器拥有网络中建立汇聚树的相关信息并且在需要时建立汇聚树,一旦它决定了哪个汇聚树用于该组,树的管理器在 MPLS 域的入口边界路由器装载相应的状态即分配相应的标签绑定。域内汇聚树的构建可以利用现成的组播路由协议来实现。文[8]提出的类似方案中,组播的每个分支可以和其他共享入口和出口边界交换路由器的单播流聚合,从而很好地实现了单播流与组播流的汇聚问题。

3.4 MPLS 环境下组播树的创建

(1)利用 MPLS 的 LDP 构造 MPLS 组播树。利用 MPLS LDP 构造 MPLS 组播树有两种方式:发送端发起的树建立方式和接收端发起的树建立方式^[15,16]。发送端发起方式适用于接收端数目有限并且加入/剪枝消息很少的情况下,第一跳路由器基于源端的通告来计算组播树。接收端发起方式适用于接收端数目较多且加入/剪枝消息频繁的情况下,接收端作为根来计算组播树。每个接收端路由器根据接收端的预留,分别计算满足 QoS 要求的路径。这个路径的计算可以只采用单播路由协议,也可以结合组播特定流的状态信息。文[11,14,17]介绍了如何利用 PIM 协议给组播路由器分发 MPLS 标签,提出了使用捎带(piggybacking)的方法给稀疏模式的组播树分发标签,即将 PIM-SM 的加入消息进行扩展,使其捎带下游 LSR 分配的标签。该方案的缺点是没有解决 RPF 检查问题。

(2)创建基于类的组播树。创建基于类的组播树的方法是通过组播路由器创建并管理一定数量组播转发实体的方式来达到减少组播转发状态的目的,并实现组播的流量工程^[12]。该方法引入实体(C,G)(C 代表 QoS 类型,G 为组地址)和 MTCA(Multicast Tree Class Agent),MCTA 存储了域内的拓扑结构和每个组播组相应的组播树,其作用是将组播的控制消息传送到相应的分支路由器。仅 MCTA 有管理组播状态信息,因此每个 LSR 只执行 L2 层交换功能。考虑到 MCTA 必须管理所有的控制流而导致的拥塞,可采用多个 MCTA 对不同类型的树加以管理。

(3)MPLS 组播树。关于组播树的建立更具意义的是 MPLS 组播树^[18]。在核心网络,为了构建组播树,不需要在所有的路由器上为每个组播会话保存转发状态信息。只有分支节点路由器才需要保存组播会话的转发状态,从而减少了组播转发状态,树的管理开销也随之减少。在该方案中,当一个组播分组到达 MPLS 域的入口边界路由器时,路由器根据它的组播 IP 头进行分析,决定其下一跳分支节点路由器,并基于该信息产生分组的多个拷贝,同时 MPLS 标签被压栈到对应每个下一分支节点路由器。接着,打了标签的组播分组就与单播分组没有区别。当到达下一跳分支节点路由器时,标签被取下来,并重复上述过程。这个过程一直重复下去,直

到分组到达它的目的端。当到达一个 LAN 时,没有标记标签的分组通过传统的组播协议根据 IGMP 信息传送。由此可见,组播分组在非分支节点被作为标记过标签的单播分组一样处理,不需要为组播分组提供额外的管理开销。

结束语 随着网络的发展和多媒体应用的兴起,组播已成为 Internet 应用的迫切需要,组播 QoS 和流量工程也日益显得重要。实现 MPLS 网络环境下组播业务,能够很好地解决组播所面临的问题。

由于各种应用采用的组播路由协议不同,解决 MPLS 网络环境下面临的问题也有所不同。针对不同的需要,本文总结了当前各种不同的解决方案。从今后发展趋势来看,应该综合各种方案的优点,在减少标签数、维护开销及快速的故障恢复等方面做进一步研究,以实现 MPLS 技术与组播技术完美结合,为组播提供更好的 QoS 和流量工程。

尽管组播技术在 MPLS 网络环境下的实现还不完善,但可以预见,在不久的将来,一个 MPLS 网络支持的组播将会给网络应用带来更大的发展空间。

参考文献

- 1 Rosen, et al, MPLS Architecture, RFC 3031[S]. 2001
- 2 Alwayn V. Advanced MPLS Design and Implementation. Cisco Press, 2002
- 3 Yang Baijia, Mohapatra P. Multicast in MPLS domains. Computer Communications, 2004, 27: 162~170
- 4 Matrawy A, et al. On the Scalability of using MPLS in Multicast Shared Trees. In: Workshop on the Quest to Control Next Generation Transport Networks, 2004
- 5 Agarwal A, Wang Kang-Bin. Supporting Quality of Service in IP Multicast Networks. Computer Communications, 2003, 26: 1533~1540
- 6 Ooms D, et al. Overview of IP Multicast in a Multi Protocol Label Switching Environment [S]. RFC3353, 2002
- 7 Almeroth K C. The Evolution of Multicast: From the Mbone to Inter-domain Multicast to Internet2 Deployment. IEEE Network, 2000
- 8 Fei A, et al. Aggregated Multicast: an Approach to Reduce Multicast State [R]. [UCLA CSD Technical Report # 010012]. 2001
- 9 Yang B, Mohapatra P. Edge Router Multicasting With MPLS Traffic Engineering [A]. ICON[C], 2002
- 10 Boudani A, Cousin B. A New Approach to Construct Multicast Trees in MPLS Networks [A]. In: Seventh IEEE Symposium on Computers and Communications [C], 2002. 913~919
- 11 Aggarwal R, Pusateri T, Farinacci D, et al. IP Multicast With PIM-SM Over a MPLS Traffic Engineered Core [EB/OL]. draft-ggawra-pim-sm-mpls-te-00. txt. 2004
- 12 Yeon Hunje, et al. Multicast Performance Enhancement with Class Based Tree Setup in MPLS Environment. ITC-CSCC2004, 2004
- 13 Matrawy A, Lung C H, Lambadaris I. A Framework for MPLS Path Setup in Uni- Directional Multicast Shared Trees. In: Proc. of SPIE ITCOM 2004, the Conference on Performance, QoS, and Control of Next-Generation Communication Networks, 2004
- 14 Aggarwal R, et al. Establishing Point to Multipoint MPLS TE LSPs. draft-ietf-mpls-p2mp-requirement-04. txt. September 2004
- 15 Ooms, Hoebeke R, Cheval P, et al. MPLS Multicast Traffic Engineering, draft-ooms-mpls-multicast-te-01. txt, February 2002
- 16 Yang Baijia, Mohapatra P. Multicast in MPLS domains. Computer Communications, 2004, 27: 162~170
- 17 Farinacci, et al. Using PIM to Distribute MPLS Labels for Multicast Routes [EB/OL]. draft-farinacci-mpls-multicast-03. txt, 2000-11
- 18 Boudani A. An Effective Solution for Multicast Scalability, the MPLS Multicast Tree (MMT) draft-boudani-mpls-multi-cast-tree-06. txt April 2005