第 33 卷 第 2 期 计 算 机 工 程 2007 年 1 月
Vol.33 No.2 Computer Engineering January 2007
·网络与通信· 文章编号：1000—3428(2007)02—0111—03 文献标识码：A 中图分类号：TP393

使用 CR-LDP 协议实现组播的研究

张 进
(西北工业大学软件与微电子学院，西安 710065)

摘 要：扩展 RSVP 从设计之初就考虑了支持组播技术，而 CR-LDP 没有提供支持组播的机制。虽然已经有了关于 MPLS 组播的规范草案，但其中只是对组播路由协议的选择以及与组播的各种可选项的关系进行了讨论，并没有提出如何使用 CR-LDP 实现组播机制。通过对 CR-LDP 的消息进行扩展，可以在 MPLS 网络中建立组播路径和 LSP，从而实现具有流量工程特征的组播路径。该文对如何使用 CR-LDP 实现组播进行了详细的论述。

关键词：组播；基于约束的标签分发协议；标签交换路径；多协议标签交换

Realization of Multicast Using CR-LDP

ZHANG Jin
(College of Software and Microelectronics, Northwestern Polytechnical University, Xi'an 710065)

【Abstract】 Compared with RSVP, CR-LDP dose not provide support for multicast, yet RSVP provide support for multicast at the beginning of design. Although there have been drafts of standardization about MPLS multicast architecture, it dose not describe how to realize multicast with CR-LDP in MPLS network and just discuss the choice of multicast route protocol and the relation between multicast and optional parts of LDP. A new method is proposed to build a shared multicast distribution tree centered at a rendezvous point using CR-LDP message(extensions). This paper discusses how to realize multicast with CR-LDP in MPLS network.

【Key words】 Multicast; CR-LDP; LSP; Multiprotocol label switching (MPLS)

多协议标签交换(MPLS)是 IP 通信领域中的一项新技术，是对传统 IP 网络技术的改进，是一种较为理想的 IP 骨干网络技术。使用这一技术能够实现许多崭新的功能，如流量工程、显式路由、VPN 等。

流量工程(Traffic Engineering，TE)是设计流量使之能够在网络上正常传输。一般来说，流量工程的核心就是把流量进行转移，从而使拥塞链路上的流量能够转移到那些没有被充分使用的链路上。目前，实现 MPLS 流量工程的信令协议有 2 种：一种是 CR-LDP(Constraint-Based Routed Label Distribution Protocol)；另一种是扩展 RSVP(Resource Reservation Protocol)。与扩展 RSVP 方案相比，CR-LDP 没有提供在 MPLS 技术中支持组播功能的机制。在本文中将对 CR-LDP 如何在 MPLS 网络中实现 IP 组播进行详细的论述。

1 使用 CR-LDP 实现组播功能

1.1 CR-LDP 实现组播功能基本原理

由于 CR-LDP 是一种基于目的树结构的控制协议，它将为每一个业务流建立一条 LSP 并进行资源预留。在 Internet 世界中，对于组播这样的业务，如果仅仅使用目的树结构的控制协议就会造成很大的资源浪费。通过对 CR-LDP 中信令系统与操作规程进行扩展，可以在 MPLS 网络中建立 IP 组播业务。为了更好地描述问题，可以把 MPLS 域和组播域统一在一起，即一个 MPLS 网络，同时又是一个组播域，在这个域中的标签交换路由器(Label Switch Router，LSR)运行相同的域内组播路由协议如距离矢量组播路由协议(Distance Vector Multicast Routing Protocol，DVMRP)、协议无关组播－密集模式(Protocol Independent Multicast-Dense Mode，PIM-DM)和协议无关组播－稀疏模式(Protocol Independent Multicast- Sparse Mode，PIM-SM)。应该考虑到组播源和组播数据接收者并不在同一个组播域内(域间组播)的这种情况。由于域间组播目前仍然处于研究和试验阶段，目前比较成熟的解决方案是下面 3 个协议的组合：组播边界网关协议(Multicast Border Gateway Protocol，MBGP)，用于在自治域之间交换组播路由信息；组播信源发现协议(Multicast Source Discovery Protocol，MSDP)，用于在 ISP 之间交换组播信源信息；PIM-SM，用作域内的组播路由协议。由于考虑到域间组播路由的复杂情况，因此在本解决方案中组合这 3 种协议实现组播技术。

在组播域内，通常可以通过静态方式配置一个路由器作为汇集点(Rendezvous Point，RP)，RP 还可以根据 PIM-SM 动态获得。RP 的作用是建立源和域内组播树节点之间的联系。在 PIM-SM 域中，运行 PIM-SM 的路由器周期性地发送 Hello 消息，用于发现邻接的 PIM 路由器，并负责在多路访问网络中选举指定路由器(Designated Router，DR)。DR 负责为与其直连的组成员向组播树根节点的方向发送“加入/剪枝”消息，或是将直接相连的组播源的数据发向组播分发树。当 DR 直接相连的网络中具有组播组 G 的成员活动时，DR 就要加入 RP 的共享树。建立组播 LSP 的具体过程如下：

(1)一个 LSR(DR)运行 IGMP 协议，查询到与它相连的某个主机想要加入一个组播组。

(2)需要加入组播树的 DR 向域中的共享树根结点(RP)发送 JOIN 消息，在该消息中将包含该 LSR 的 IP 地址、组播组地址以及业务

作者简介：张 进(1979－)，男，硕士生，主研方向：计算机网络，MPLS 技术，流量工程

收稿日期：2006-02-21 **E-mail：**dengtuzi_xa@mail.china.com

量参数等信息。

(3)收到 JOIN 消息的 RP 将向想加入共享树的 LSR 回复一个 ACK 消息,其中将包含 RP 的 LSR-ID 与 RP 为该业务分配的 LSP-ID，RP 记录组播地址与 LSP-ID 之间的一一对应关系。

(4)RP 的 MPLS 路径选择单元查找 TED(Traffic Enginee- ring Database)，依据有关业务量信息和某种路由算法，计算出从 RP 到某个加入节点的共享组播树路径。

(5)根节点将根据计算出的路径发送标记请求消息，在这一消息中将包含前面定义的组播 LSP-ID TLV 和显式路由 TLV 等。

(6)标记请求消息沿计算好的路径传输，收到标签请求消息的中间节点除了执行标准的 CR-LDP 规程之外，还在标记请求状态表中记录下组播 LSP-ID，RP 的 LSR-ID 和请求加入 LSR 的 LSR-ID。

(7)当请求加入组播组的 DR 收到标签请求消息时，该 LSR 将检查该请求消息中携带的 LSP-ID 和 RP 的 LSR-ID 等消息是否与先前收到的 ACK 消息中的内容一致，如果一致，则分配标签，向 RP 发回标签映射消息。

(8)收到标签映射消息的中间 LSR 首先检查是否已经为具有相同 LSP-ID 的标签请求消息进行过标签分配和标签绑定的分发。如果没有，则表明该标签映射消息来自于组播树的第一条下游分支，此时，该中间 LSR 将为该标签请求消息进行标签分配，并向上游节点发出标签映射消息，同时记录这次标签分配直到路由器明确组播树的拆除。沿途 LSR 都须记录相关的组播转发状态(路由表项)，建立组播标签转发表。

(9)如果中间 LSR 发现对于某个标签映射消息中的 LSP-ID 和 RP 的 LSR-ID，该节点过去已经为某一组播树分支进行了标签分配与标签绑定的分发的话，则不再进行标签分配。该中间 LSR 将检查前一次标签分配对应的资源预留是否大于新的标签映射消息中的资源预留，最终，应尽量使资源预留取二者中较大的数值。随后，该 LSR 清除标签请求状态消息库中具有与该标签映射消息完全相同的 LSP-ID TLV 的标签请求状态，将标签映射消息中的标签与收到该标签映射消息的接口号，在上次标签分配中已经建立起来组播标签转发表中的转发条目增加出标签与出接口列表，并向上游节点发出组播确认通知消息。

(10)收到组播确认消息的中间节点依据消息中的请求加入的 LSR ID 和 LSP-ID TLV，清除标签请求状态信息库中具有与该消息中完全相同的 LSP-ID TLV 的标签请求状态,并向上游节点继续发送组播确认消息。

(11)当 RP 收到标签映射消息或组播确认消息时，新的组播树分支的建立即告结束。RP 将利用收到的标签映射消息或组播确认消息中的标签为相应的业务建立标签转发表。

1.2 对 CR-LDP 消息的扩展

要实现上述功能，只需对 CR-LDP 协议进行以下扩展：在 CR-LDP 中，一个重要的机制就是 LSP-ID TLV (Type-Length-Value)。在标签请求和标签映射消息中，除了需要包含 CR-LDP 规范中规定的内容之外，还将包含组播 LSP-ID TLV，组播 LSP-ID TLV 是组播专用的 TLV。LSP-ID TLV 中包含一个 LSP-ID，一个入口 LSR-ID，一个出口 LSR-ID。该 TLV 是 MPLS 组播专用的 TLV，其编码如图 1 所示。

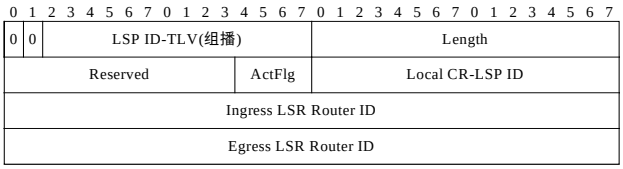

图 1 LSP-ID TLV 的编码

其中：ActFlg 标记的语义与标准的 CR-LDP 中 LSP-ID TLV 中的定义相同，利用这一标记，可以实现对 LSP 的重新路由；Local CR-LSP ID 是 RP 节点分配的本地唯一的一个 LSP-ID；Ingress LSR Router ID 是组播共享树起点的 LSR 标记；Egress LSR Router ID 是 MPLS 请求加入组播树 DR 的 LSR 的标记。

根据上文的研究,如果 LSR 发现对于某个标签映射消息，该节点过去已经为和它属于同一组播树的分支进行过标签分配的话，即标签映射消息发生反向合并之后，合并点 LSR 需要向 RP 发出组播确认通知消息，这一消息的编码如图 2 所示。

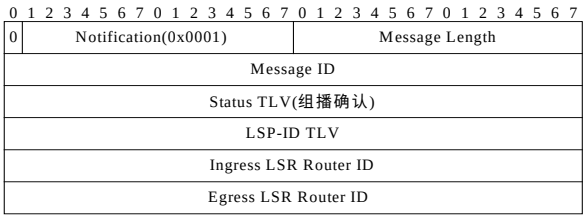

图 2 组播确认消息编码

中间节点收到这一消息时，将取消本地状态信息库中具有与该消息中 LSP-ID TLV 中相同的 LSP-ID 值与 RP 的 LSR-ID 值的标签请求状态，同时，继续向上游节点发送这一消息。RP 节点收到这一消息时，将使用该 LSP-ID 已经从下游节点获得的标签进行业务分组的传送。另外，在各个节点上，需要建立具有一对多功能的标签转发表，即一个入标签可能对应多个出标签。

1.3 组播树的剪枝

当某个 DR 发现与之相邻的网络上最后一个组播数据的接收端发送 IGMP 离开消息时，则查找组播标签转发表，找到组 G 的组播数据的入接口，并向该接口邻接路由器发送 PIM-SM 剪枝消息。收到剪枝消息的路由器同样也要查找组播标签转发表，清除组播标签转发表中组 G 的组播数据出接口列表中相应接口数据，释放标签资源等，并查找还有其他出接口还有组 G 数据。如果没有，该路由器也要向组 G 的入接口邻接路由器发送剪枝消息。依次地，一直到 RP，拆除组播树和释放相应资源(标签等)。

2 建立组播源到 RP 之间的路径

如果组播源和组播数据接受者在同一域内，那么就可以在组播源和 RP 之间建立 LSP。组播数据首先经过源到 RP 的 LSP 被传送 RP，再由 RP 中转，沿组播 LSP 发送到组播数据接收者。如果组播源在组播域之外，建立组播源到 RP 的 LSP 的过程就比较复杂。本方案提出的 MBGP、MSDP、PIM-SM 这 3 种协议的组合，在信源到各域中的 RP 会形成一个跨域的组播树。考虑到域之外可能不是 MPLS 网络，以及域间组播路由的复杂性，所以在实现时，从域外的发送端到域内 RP 之间使用普通 IP 组播树路径，不建立 LSP。各 RP 直接向域外的组播源发送 JOIN 消息，加入到源的最短路径树。此处的最短路径树是依据传统 IP 基于度量的路由产生的，不具有流量工程的特性。

3 组播路径的选择

在以上步骤(4)中，提到由 RP 依据有关业务量信息和路由算法来为想加入组播组的 DR 计算路径。在 IP 网的组播路由选择问题可以归结为建立一个从中心点到多个节点的组播路由选择树的问题。其中，一个最优的选择树就是具有最小链路代价和的树。寻找最小链路代价和树的问题被称为 Steiner 树问题。但是，在现有的实现中，没有一个组播路由算法是基于这个算法的。最主要的一个原因是为了保持最小

代价树，每当一个链路代价发生变化时，必须重新运行算法来计算树。在 IP 组播中实现较多的是基于逆向路径转发(reverse path forwarding，RPF)算法实现的组播路由树。

在 MPLS TE 中，一般使用约束最短路径优先(Constraint-based Shortest Path First，CSPF)算法来决定每条 LSP 的物理路径。由于在 MPLS TE 中的路径选择要考虑到多种约束条件以及网络的动态变化，因此计算路由时，不能采用传统 IP 组播中的算法来构造路由树。可以把构造组播树的过程分解为对树的每一个节点计算单播 LSP 路径的过程。在本方案中采用一种特殊的 CSPF 算法来为树的每一节点计算路径。组播路径如果使用已建立且属于同一组播组的链路的话，那么在该链路上不需要再额外分配带宽和消耗资源。所以该算法在计算路径时对于已建立且属于同一组播组的链路认为带宽始终满足约束条件，并且运行 CSPF 算法时，将该链路的代价和跳数忽略。这样做的目的是尽量使已建立且属于同一组播组的链路在选择时权重占优，使新节点始终能嫁接到满足约束条件的最短路径树上。这主要是基于在这部分链路上不需要再额外分配带宽和消耗资源考虑的。

4 共享树到基于源的组播树之间的切换

在 PIM-SM 中，可以将组播数据从基于 RP 的共享树切换到基于源的最短路径树上。在本方案中，也可以将基于 RP 的组播 LSP 切换到基于源的路径上。如果组播源在域内，那么在切换之前可以建立基于源的组播 LSP，建立 LSP 的过程与上文的步骤相同。当数据切换到基于源的树之后，各节点向 RP 发送剪枝消息，并拆除从 RP 到各节点的组播 LSP。如果组播源在域外，考虑到域间组播的特殊性以及域外无法建立 LSP，所以源与各节点之间不建立 LSP，节点直接加入基于源的最短路径树，但是切换之后，应向 RP 发送剪枝消息，并拆除组播 LSP。

5 结论

本文详细论述了一种在 MPLS 网络中基于 CR-LDP 实现组播技术的解决方案。这种方案依赖于 CR-LDP 现有的各种消息与 TLV，需要对个别消息和 TLV 进行扩展，可以与现有的各种 MPLS 操作规程紧密地结合在一起，故而不会对标准的 MPLS 造成很大的影响。CR-LDP 本身是实现 MPLS TE 的很好途径，因此使用该方案可以把组播和 MPLS 的优势结合起来，有效地保证组播流量的传输质量，很好地节省了网络资源。然而由于考虑到目前使用不同协议的域间组播技术上的不成熟及 MPLS 域外无法实施 MPLS TE，本文只考虑了 MBGP，MSDP 和 PIM-SM 三者组合的这一种解决方案，并只考虑在域内建立组播 LSP。

参考文献

1 Network Working Group. Request for Comments: 3212, Constraint-based LSP Setup Using LDP[Z]. 2002-01.

2 吴　江, 赵慧玲. 下一代的 IP 骨干网络技术——多协议标记交换[M]. 北京: 人民邮电出版社, 2001-01 .

3 Osborne E. 基于 MPLS 的流量工程[M]. 张　辉, 卢　炜, 译. 北京: 人民邮电出版社, 2003-07.

4 Parkhurst W R. Cisco 组播路由与交换技术[M]. 京京工作室, 译. 北京: 机械工业出版社, 1999-11.

(上接第 98 页)

传统的结构化对等网络支持基于关键字的查询，但不支持复杂查询，如多属性查询和并行查询。而 UDDI 程序员规范称在主要的查询接口中支持复杂查询，所以传统的结构化对等网络系统不能直接应用来构建分布式的注册网络。但按照如下步骤，基于单关键字的查询机制已经足够实现精确定位正确的注册中心。

第 1 步　采用一个基于 Chord 协议的结构化对等网络系统。每个注册中心必须声明负责对应统一分类体系划分的某个子树。子树的根节点取值被散列到一个唯一的数字关键字 key。每个注册中心必须通过 put（key，object）操作将其负责的子树发布到结构化对等网络系统中，object 包含注册中心访问入口，它可以是任意的数据结构。

第 2 步　如果服务提供方没有注册中心分布及其职责范围的任何先验知识，服务发布消息可以被送到任意一个注册中心。注册中心在收到发布消息后，从中获取分类信息并提取出分类子树的根节点，然后通过同样的哈希函数将根节点的值散列到一个唯一的数字关键字上。通过 lookup（key）操作可以将发布消息正确转发给负责存储的注册中心。

第 3 步　如果没有注册中心分布及其职责范围的先验信息，服务查询消息能够被送到任意一个注册中心。注册中心接收到查询消息后，从中提取出子树的根节点，然后由相同哈希函数生成根节点的散列值。通过 lookup（key）操作可以找到正确的注册中心，然后注册中心将查询消息提交给所有的候选服务注册，并融合不同注册中心的响应作为最后的响应反馈给服务请求方。

3 UDDI 对等网络原型

我们在网络环境下安装了多套 Windows 2003 Server 操作系统中集成的标准 UDDI。在这些独立部署的 UDDI 注册中心的基础上，配置自主开发的查询转发和集成模块，从而实现可配置的、具有 Gnutella 和 Random Walk 协议特征的非结构化标准 UDDI 对等网络，并使用网格监控服务提供的信息，保障非结构化 UDDI 对等网络的查询输出结果具有很高的可用性。该原型系统已经在国家地质信息网格项目中得到应用和推广。

4 结论

本文提出了两种扩展性很好的分布式机制来克服面向传统集中式注册中心的缺点。首先在完全自治且独立部署的众多标准 UDDI 的基础上提出了非结构化 UDDI 对等网络，然后在可以彼此协同的众多标准 UDDI 的基础上提出了结构化 UDDI 对等网络。这两种 UDDI 对等网络在支持复杂查询协议的同时无损同类产品之间的互操作级别，并且都是扩展传统服务发现协议的不错方案。

参考文献

1 Gnutella[Z]. http://rfc-gnutella.sourceforge.net/src/rfc-06-draft.html.

2 Yang B, Garcia-Molina H. Improving Search in Peer-to-peer Networks[C]. Proc. of the 22nd IEEE International Conference on Distributed Computing, 2002.

3 Crespo A, Garcia-Molina H. Routing Indices for Peer-to-peer Systems[C]. Proc. of the 22nd International Conference on Distributed Computing, 2002.

4 Bloom B. Space/time Tradeoffs in Hash Coding with Allowable Errors[J]. Commun. of ACM, 1970, 13(7): 422-426.

5 Stoica I, Morris R, Karger D, et al. Chord: A Scalable Peer-to-peer Lookup Service for Internet Applications[C]. Proc. of ACM SIGCOMM, 2001.