

IP 网络组播路由研究^{*}

江 魁, 杨文玲

(深圳大学网络中心, 广东 深圳 518060)

摘 要: 阐述了 IP 组播的基本原理, 分析了各组播路由协议, 探讨了它们之间的联系和优缺点, 并对组播技术的未来发展作出了预测。

关键词: 组播; 组播路由; 网络性能

中图分类号: TP393.4 **文献标识码:** A **文章编号:** 0529-6579 (2002) S1-0042-03

1 组播的基本概念

传统网络通信模式主要有单播 (unicast) 和广播 (broadcast) 两种, 单播方式下数据只有一个接收者, 是点到点的通信; 广播方式下网络中所有主机都是数据接收者, 是点到多点的通信。随着网络应用的不断发展, 语音、视频等多媒体数据在网络中大量传送, 单播方式下数据流要重复传送多次, 浪费了网络带宽; 广播方式下会造成不希望接收数据的主机资源耗费。组播 (multicast) 克服了单播和广播通信方式缺点, 组播方式下数据只被同一个组播组成员接收, 非组播组成员不会接收到数据, 因此应用组播可以显著减少网络负载, 提高网络运行效率, 改善网络性能, 使得分布式网络应用更为有效地实现。

2 组播路由概述

常用路由协议如 RIP、OSPF、BGP 等适用于单播通信, 不适用组播通信。组播路由协议完成的是路由器间组播数据转发, 其构造思路是在组成员之间构造一棵分发树, 组播数据在这颗分发树上传播至所有接收者。分发树构造从结构上主要分为两类: 最短路径树和共享树。

2.1 最短路径树 SPT (shortest path trees)

最短路径树 SPT 又叫有源树 (source based trees), 以组播源为根构造从根到所有接收者路径最短的分发树。如果组中存在 M 个组播源, 必须为每个组播源构造一棵组播树, 需要生成 M 棵组播树。不同组播源发出的数据包被分散到各自的组播树上, 采用最短路径树有利于网络中数

据流量的均衡。由于从组播源到每个接收者的路径最短, 端到端时延性能较好, 有利于流量大、时延性能要求较高的实时媒体应用。其缺点主要是构造分发树开销较大, 消耗较多的路由器 CPU、内存等资源。

2.2 共享树 ST (shared trees)

共享树又叫有核树 (core based trees), 为每个组播组选定一个汇合点 RP (rendezvous point), 以 RP 为根建立组播树。组播源先将数据包单播到 RP, 再由 RP 向组播组成员转发。共享树不是对所有组播源都是最优, 但将最短路径树中 M 棵树的存储开销降到了一棵树, 在路由器存储空间和组播树的代价两个方面具有较好性能。当组规模较大, 每个成员发送的数据量较少时, 使用共享树比较合适。当通信量较大时, 使用共享树会使流量集中在 RP 附近, 带来额外的延迟, 降低网络性能。

3 域内组播路由协议

Internet 是由不同的自治系统 (automated system) 组成, 自治系统可以看作是采用同一路由策略的路由器集合^[1]。组播路由协议分为自治系统内的域内路由协议和自治系统间的域间路由协议。域内组播路由协议有密集模式 (dense mode) 和稀疏模式 (sparse mode) 两种工作模式。密集模式假设所有主机都接收组播数据, 将数据从路由器各个接口扩散出去, 当某个主机不接收组播数据时, 向上级 (upstream) 路由器发出剪枝消

* 收稿日期: 2002-03-21

作者简介: 江 魁 (1975-), 男, 助理工程师, E-mail: jiangkui@szu.edu.cn

息将其从分发树中删除;稀疏模式由期望接收组播数据的主机向上级路由器发出加入消息加入到分发树中。

域内组播路由协议主要有 4 种:距离向量组播路由协议 DVMRP (distance vector multicast routing protocol)、开放式组播最短路径优先协议 MOSPF (multicast open shortest path first)、协议无关组播密集模式 PIM-DM (protocol independent multicast dense mode)、协议无关组播稀疏模式 PIM-SM (protocol independent multicast sparse mode)。下面分析各域内组播路由协议。

3.1 距离向量组播路由协议 DVMRP

DVMRP 是一种距离向量路由协议,动态生成最短路径树。由 4 个过程构成,第 1 个过程是邻居查找,即查找网络上具有 DVMRP 能力的路由器;第 2 个过程是组播数据传输,通过和互联网组管理协议 IGMP 的结合,向期望接收组播数据主机传输组播数据;最后两个过程是加入和剪枝,实现动态地从组播分发树中增加和删除接收组播数据的主机和路由器。为了使新加入的组播成员能及时收到组播数据,DVMRP 采用扩散机制,定时向路由器所有接口发送数据包,导致过大的网络流量,同时主机和路由器发出的大量加入和剪枝消息也带来了额外的开销;DVMRP 使用跳数作为路由量度,32 跳的上限限制了网络规模扩大;保存大量的组播路由状态信息和距离向量路由协议本身特性使得当网络规模扩大时路由表聚合时间过长,因此 DVMRP 不适合网络规模较大且对实时性要求较高环境,不适用于少量的接受者分散地处于组播源周围的网络^[2]。

3.2 开放式组播最短路径优先协议 MOSPF

MOSPF 是对单播路由协议 OSPF 的扩展,由 RFC 1584 定义,是一种链路状态路由协议,OSPF 中链路状态通告 LSA (link state advertisement) 类型 7 用来携带组播路由信息,实现组播路由的查找。每个路由器都必须维持一份最新的网络拓扑表,在此基础上用 Dijkstra 算法为每个组播源生成最短路径树。MOSPF 要先运行 OSPF,有三种级别的路由:①区内组播路由,用于各网段中的组播成员;②区间组播路由,用于汇总区内成员关系,并在区域 0 上发布组成员关系记录通告,实现区间组播包转发;③自治系统间组播路由,用于自治系统间的组播包转发。MOSPF 对网络拓扑改变能快速响应,由于需要重新计算,

消耗大量路由器 CPU、内存资源,不适用于网络拓扑不稳定的环境。

3.3 协议无关组播密集模式 PIM-DM

PIM-DM 和 DVMRP 类似,采用了扩散和剪枝机制。两者的主要区别是 DVMRP 使用内建组播路由协议,PIM-DM 则可用任一种单播路由协议。PIM-DM 优点是配置简单,易于排错,并且可以和 DVMRP 互连,不足之处在于扩散和剪枝影响网络性能,要采取复杂的断言 (assertion) 机制避免多个组播路由器重复发送数据包^[3]。PIM-DM 只支持最短路径树而不支持共享树,适用带宽较大的网络,在组播源和组较多环境下使用则可能因需维护过多组播路由信息导致路由器负载过重,降低网络性能。

3.4 协议无关组播稀疏模式 PIM-SM

PIM-SM 由 RFC 2362 定义,既支持最短路径树又支持共享树,基于显式加入模型。每个组播组有一个汇合点 RP (rendezvous point),组播源先向 RP 发送注册消息,接收端向 RP 主动发送加入消息,路由器只在某个组播组输出接口上转发数据包。组播源数据先沿最短路径树向 RP 发送数据,再由 RP 沿共享树将数据发送到各个接收端,随后直接建立组播源到接收端的最短路径树。其特点在于能从共享树向最短路径树转换,直接建立从组播源到接收端的最短路径树,绕过 RP 进行组播数据转发,从而提高了组播包转发效率,减少了网络延迟和 RP 上出现的瓶颈,改善网络性能。PIM-SM 适用于任何大小的网络,不论密集模式还是分散模式都可以取得良好的性能,因此当前组播网内普遍采用的是 PIM-SM 组播路由协议。

4 域间组播路由协议

DVMRP 需要频繁扩散、加入、剪枝和周期性路由表更新,引起的大量网络流量和其最大跳数限制使其不适用于作为 Internet 中组播路由协议;PIM-DM 会带来和 DVMRP 类似问题,而 PIM-SM 在互连多个稀疏模式域时会变得非常地不可预测,特别是把一个 PIM-SM 域当作在其他两个 PIM-SM 域间过渡网络时,会使整个网络退化到类似密集模式组播路由协议状况。由于自治系统内普遍采用 PIM-SM,接收方和发送方通过 RP 进行相互通信,每个自治系统不依赖其他自治系统 RP,这样自治系统间发送者和接受者相互通信

出现了问题, 需要有新的协议来解决这个问题。多协议边界网关协议 (multiprotocol border gateway protocol) 正是在这一背景下提出的:

4.1 多协议边界网关协议 MBGP

MBGP 由 RFC 2283 定义, 是对 BGP4 的扩展, 在其基础上能携带组播路由信息, MBGP 不传播组播状态信息也不建立组播分发树, 由 PIM-SM 来建立组播分发树和转发组播数据。采用 MBGP 后, BGP 仍然维护一个单播路由信息表 U-RIB (routing information base), MBGP 则维护一个组播路由信息表 M-RIB。MBGP 支持和 DVMRP 间的路由重分发, 可以和 DVMRP 组播网络相互学习组播路由信息, 利于实现传统 DVMRP 网向 PIM-SM 网平滑过渡。BGP 单播路由可转换成组播路由, 收到单播路由信息后, 创建相应组播路由传入到组播网络中, 能够实现非组播网到组播网转换。如果仅仅使用 MBGP 和 PIM-SM, 为实现各个域中 RP 信息共享要把 RP 放在连接各域的路由器上, 运行 MBGP 协议相互交换组播路由信息, 这种实现限制了 RP 放置位置, 且必须将各域 RP 放在同一 PIM-DM 网中来实现 RP 信息共享, 会形成过大的 PIM-DM 云, 降低网络性能。因此仅仅利用 PIM-SM 和 MBGP 不能完全解决域间组播路由问题, 要有一种机制使 RP 在域内存放位置具有很强灵活性, 并且每个域不依赖其它域的 RP。当前主要采用组播源发现协议 MSDP (multicast source discovery protocol) 来实现, 只要 RP 知道域中活动组播源, 就由接收者向组播源所在域 RP 发送加入消息。

4.2 边界网关组播协议 BGMP

尽管 PIM-SM, MSDP 和 MBGP 的结合能够解决域间组播路由问题, 但 IETF 工作组正在定义新的域间组播路由协议以便组播能在 Internet 中进一步扩展, 其中最主要的是边界网关组播协议 BGMP (border gateway multicast protocol)。BGMP 利用了动态组播地址分配中的组播地址设置一声

明 MASC (multicast address set claim) 来实现地址分配^①。BGMP 使用双向共享树来连接多个域, 其中一个域被指定为根域, 其他域向根域发送显示加入和剪枝消息, 使用 MASC 为各个域分配一个连续范围组播地址空间。BGMP 和 MASC 实现非常复杂, 当前利用还不是很成熟, 协议本身仍处于 Internet 草案阶段, 属于未来域间组播路由解决方案。

5 前景展望

随着组播应用范围逐渐扩展, 带来一些新问题, 未来组播研究重点将集中在以下几个方面: 流量工程 (traffic engineering): 流量工程是下一代 Internet 研究中的重点之一, 将流量工程引入组播路由中可以有效地提高组播网络的性能和扩展性。

组播管理: 组播网络管理所要实现的功能和单播网络管理类似, 包括组播网络结点的拓扑发现、组播网络状态监控、组播网络流量统计、组播计费、组播故障检测等功能。

组播拥塞控制: 组播是基于 UDP 协议实现的, 是一种尽力而为 (best effort) 的传送方式, 无法提供稳定的数据传输, 可能存在数据丢失, 缺少 TCP 协议中窗口机制来避免网络拥塞, 需要对组播拥塞控制作进一步的研究。

此外还有组播性能分析、安全性问题等都将未来组播研究中的重点。随着组播技术的研究深入, 这些问题最终将得到圆满的解决, 组播与组播路由必将在整个 Internet 中得到更广泛的应用。

参考文献:

- [1] DOYLE J. Routing TCP/IP (II) [M]. America: Cisco Press, 1998: 128.
- [2] 田红鹏. IP 组播技术综述[J]. 计算机世界报, 2001, 6: 32.
- [3] WILLIAMSON. Developing IP Multicast Network[M]. America: Cisco Press, 2000: 328.

① 龚俊. IP 组播服务及其实践考虑. 中国教育科研网 2001 年会报告

Research on IP Network Multicast Routing

JIANG Kui, YANG Wen-ling

(Network Center, Shenzhen University, Guangdong, Shenzhen 518060, China)

Abstract: The basic concepts of the IP multicast network are described. All kinds of multicast routing protocol are analyzed. Meanwhile, the relationships among them are introduced and the information about the advantages and disadvantages of each protocol are presented. The future development of the multicast technology is forecasted finally.

Key words: multicast; multicast routing; network performance