IPv6 中的 Anycast 的扩展性分析与研究

王晓楠　钱焕延
（南京理工大学　南京 210094）

摘　要　IPv6 以两种方式提供 Anycast 服务：一种是将 Anycast 组成员限制在共享一个地址前缀的特殊拓扑区内；另一个是将 Anycast 地址表示的共享某个特性的结点组分散在互联网的各个地方，这种方式使得路由表会随全球 Anycast 组数呈比例增长，从而构成了 Anycast 的可扩展性问题。在这种情况下，本文提出了建立在 BGP 和 ICMPv6 基础之上的一个可扩展的 Anycast 服务方案，并且深入分析了该方案的可行性以及它的性能表现。
关键词　IPv6，Anycast，BGP，ICMPv6，路由器

Analysis and Study of Anycast Scalability in IPv6

WANG Xiao-Nan　QIAN Huan-Yan
(Nanjing University of Science & Technology, Nanjing 210094)

Abstract　The existing designs for providing anycast services are either to confine each anycast group to a preconfigured topological region or to globally distribute routes to individual anycast groups which cause the routing tables to grow proportionally to the number of all global anycast groups in the entire Internet, both of which restrict and hinder the application and development of anycast services. This paper proposes a scalable architecture for global IP-anycast service on the basis of BGP and ICMPv6 in the simulating of IPV6, and analyzes and discusses the feasibility and validity of this architecture. According to the simulation, this paper further studys its error tolerance and its performance.
Keywords　IPv6, Anycast, BGP, ICMPv6, Router

1　前言

Anycast 是 IPv6 所提供的一种特殊网络服务，允许服务申请者访问共享同一 Anycast 地址所标识的一组接口中最近的一个（这里的最近是按路由协议的距离量度来计算）。图 1 说明了 Anycast 的这种功能。

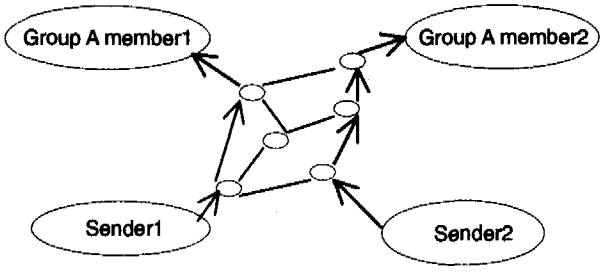

图 1　Anycast 服务

图中 Sender1 和 Sender2 都向同一个 Anycast 地址发出了请求包，但是包被网络转发到离发送者最近的一个组成员（接口）。

Anycast 有广泛的应用。例如，如果把一个 Anycast 地址分配给所有的域名服务器（DNS），当一个用户从一个网段移动到另一个网段后，他不需要重新配置其本地的 DNS，主机可以使用全局的 Anycast 地址访问任何地方的本地 DNS。

不难看出，Anycast 是一种非常有用的服务，它在许多应用领域发挥着重要的作用。随着网络新应用、新服务的不断涌现，对它的需求也在不断增长。但是，由于它的研究才刚刚起步，所以在许多方面还存在着制约这种服务实施的种种问题。Anycast 的扩展局限性就是其中之一。

2　Anycast 的扩展局限性

IPv6 中的 Anycast 地址是从 Unicast 地址空间中进行分配的，所以 Unicast 和 Anycast 地址从结构上没有任何区别。这样做的目的是为了缓解 Anycast 带来的路由表规模扩张问题，同时也能充分利用现有的路由资源，使得 Anycast 的路由问题变得更简单。传统的 Unicast 路由方法可以将所有共享同一个 Unicast 前缀的目的地址聚合为一条路由项，以达到减小路由表规模的目的，这样它可以在 Unicast 前缀允许的地址空间中扩展目的主机（或者子网）的数量。也就是说，Unicast 通过层次聚合的形式可以实现扩展性。但是，Anycast 却不能通过这种层次聚合来实现其扩展性，如图 2 所示。

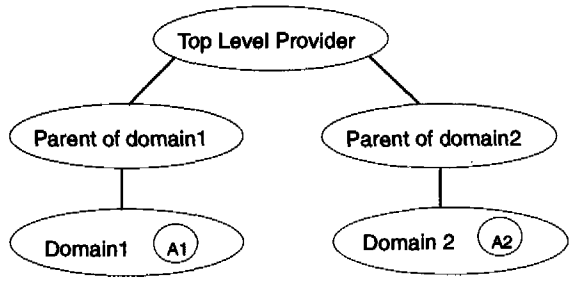

图 2　Anycast 拒绝层次聚合

一个 Anycast 地址表示的是共享某个特性的结点组，这些结点可能分散在互联网的各个地方。如果用 Unicast 路由协议路由 Anycast，则每个全球 Anycast 地址必须作为独立的路由表项处理。这种要求使得路由表会随全球 Anycast 组数呈比例增长，从而构成了 Anycast 的可扩展性问题。

王晓楠　博士研究生，研究方向为网络安全应用；**钱焕延**　教授、博士导师，研究方向为网络安全与应用。

IPv6 将每个 Anycast 组成员限制在共享一个地址前缀的特殊拓扑区内。在这个拓扑区域中，Anycast 地址在单播路由系统中是独立的表项；在该拓扑区域外，Anycast 地址被汇聚到其所在区域的地址前缀的路由项中传播。通过把 Anycast 组限制在一个预定义的区域内，IPv6 减轻了 Anycast 的可扩展性问题，但是并没有根本解决它。因为全球 Anycast 地址仍然需要作为独立的路由项在互联网中传播。而对某些应用来说，全球 Anycast 地址是必需的。

在这种情况下，本文提出在应用层实现 Anycast 服务。在应用层实现 Anycast 服务有很多优点：首先，可以比较容易地实现路由配置，而无需把每个 Anycast 地址作为独立的路由表项处理(IP 层的处理方式)；其次，可以使用应用层独特的距离度量方式，例如服务器的负载度量；最后，可以有效地解决 Anycast 的扩展问题。

下面就此方案给予具体的分析和讨论。

3 应用层实现 Anycast 服务

本方案是在 Multicast 和 Unicast 分层结构网络域间路由的基础上提出来的，它采用 BGP4 和 ICMPv6 协议来实现 Anycast 的地址路由。

3.1 地址问题

IPv6 中的 Anycast 地址模型与 RFC1546 最初建议的完全不同，前者提出在 Unicast 地址空间中分配 Anycast 地址，这样 Unicast 和 Anycast 地址从结构上没有任何区别；后者则推荐使用独立的地址模型。本方案采用后者的观点，使用 Anycast 的独立地址模型。

IPv6 的地址格式与 IPv4 不同，一个 IPv6 的 IP 地址由 8 个地址节组成，每节包含 16 个地址位，除了 128 位的地址空间，IPv6 还为点对点通信设计了一种具有分级结构的地址，其分级结构划分如下所示：

3	13	8	24	16	64
FP	TLA ID	RES	NLA ID	SLA ID	Interface ID

其中，FP 是可聚合全局地址的格式前缀(例如，001 代表单播地址)；TLA ID 为顶级聚合标识符；RES 为将来使用而保留；NLA ID 是次级聚合标识符；SLA ID 是站点级聚合标识符；Interface ID 为接口标识符。IPv6 全局单播地址的分配方式如下：顶级地址聚合机构 TLA(即大的 ISP 或地址管理机构)获得大块地址，负责给次级地址聚合机构 NLA(中小规模 ISP)分配地址，NLA 给站点级地址聚合机构 SLA(子网)和网络用户分配地址。

从以上分析看出，IPv6 的地址是分层的。由于本方案是在应用层上实现 Anycast 服务，所以 Anycast 地址是域相关的，而每个域又和一个分层的 IPv6 地址相关联，所以得出如下的 Anycast 地址格式：

3	13	8	24	16	64
Anycast Prefix	Home Domain				Interface ID

其中，地址的前三位是 Anycast 的地址前缀，其取值范围可以是除了 001(可聚合全局单播地址的格式前缀)以外的任何值。而随后的 TLA ID、RES、NLA ID 和 SLA ID 的值则和该 Anycast 所在的主域的 Unicast 地址空间相同。这里的主域必须包括该 Anycast 地址所确定的一个组成员。这种地址分配结构使得 Anycast 地址结构仍然具有全球性，可以把 Anycast 地址分配到 Internet 的每台机器上。

3.2 加入 Anycast 组

Anycast 所在的主域里都设置一个控制员(该控制员也是 Anycast 的一个组成员)。如果一台主机要加入 Anycast 组，它必须知道该控制员本身的 Unicast 地址以及它所在的 Anycast 地址。一台主机加入 Anycast 组的过程描述如下：

①该主机向控制员(Unicast 地址)提出加入 Anycast 申请(Anycast 地址)；

②控制员接到该申请后，如果允许它加入，就发送允许加入的消息，否则发送拒绝加入的消息；

③主机接收到控制员的消息后，如果是允许加入，那么就发消息给本域的内部路由器以及边界路由器，通知其新的身份，否则放弃该消息。

上述过程可以通过在 BGP(域间)或者 IGMP(域内)中增加新的消息类型来实现。这里，为了防止被恶意攻击，主机和路由器之间的信息交互要采用一些安全措施。

3.3 路由分析

在分析本方案的路由之前，先明确本文所提到的几个概念。

域：指一系列路由器和网络或者一个自治系统。一个自治系统的所有部分都是保持连通的，也就是说自治系统内部的所有路由器都是互联的，且这些路由器通过不断交换路由信息来维持相互的内部连通性。每个自治系统都被赋予一个自治系统编号，该号码由负责分配 Internet 网络地址的管理机构来分发。

对等体：为了交换路由信息，在任意两个路由之间建立连接，并进行网络可达信息的交换，则这两个路由器称作对等体，也称相邻体。

域邻居：以本域为中心，半径为 R 的区域里的所有对等体。这里的 R 指 Domain Hop 的次数。

根据以上的概念，可以认为整个网络就是多个域的集合。而在整个网络中，为了准确地把数据包从源端传送到目的端，又需要很多路由器进行路由。这些路由器根据各自功能的不同，又划分为内部路由器、主干路由器和域边界路由器。内部路由器直接连接的网络都属于同一域；主干路由器不属于任何域的网络，它们形成了域间主干网；域边界路由器是与多个域相连的路由器。

3.3.1 域内路由

Anycast 域内路由方法和传统的 Unicast 域内路由方法基本相同，这里不再赘述。Anycast 域内路由可以采用多种算法(例如 RIP)，如果一个域内具有多个 Anycast 组成员，那么由路径最短的 Anycast 组成员来提供 Anycast 服务。

下面着重介绍 Anycast 的域间路由。

3.3.2 域间路由

一般来说，Anycast 报文在两个域间的路由可分成三段连续的路径：源域内的一段路径，即从源主机到本域边界路由器；主干部分的一段路径，从源域边界路由器到目标域边界路由器；目标域内的一段路径，即从目标域边界路由器到目标主机。这三段连续的路径路由需要通过内部路由器、主干路由器和域边界路由器之间相互交换路由信息来实现，具体步骤如下：①域边界路由器汇总本域的 Anycast 路由信息并发布给主干路由器，这些汇总信息提供本域内包含哪些 Anycast 组，以及该域边界路由器到这些 Anycast 组成员的距离等信息；②域边界路由器计算主干的最短路径优先树，从而得出到其它域边界路由器的距离；③通过比较其它域边界路由器广播的域汇总信息，域边界路由器可以得出去往所有本域以外的其它网络的距离，然后在本域内广播这些距离，同时根据链

路状态计算自己的路由表;④内部路由器根据域边界路由器广播的信息,选择最佳的域边界路由器。

综上所述不难看出,域间路由的核心内容就是获取到达Anycast组的最短路径。一般情况下,域边界路由器会周期性地发出Anycast最短路径查询消息到其相邻对等体,然后比较相邻对等体的应答信息与自己路由表中的最短路径信息,如果前者的路径比后者的路径更短,那么就更新自己的路由表。但是,域边界路由器里最原始的Anycast最短路径信息是从哪里来的呢?这就需要路由学习。

这里的路由学习过程是建立在BGP路由器与其对等体所建立的TCP连接基础之上的。为了达到路由学习的目的,在BGP中增加了两种消息:查询消息和应答消息。下面详细分析路由学习的过程。

①当域边界路由器接收到一个Anycast数据包时,它首先查看自己的路由表。如果没有关于这个Anycast组的路由信息,它会发送一个查询消息给它的相邻对等体。从这可以看出,查询消息的传播实际上就是域到域的广播。为了有效地获取Anycast组的最短路径,在查询消息中设置了两个字段:一个是路径属性字段,此字段记录了查询消息所跨越的所有域,这样可以防止路由查询消息时出现环路;第二个字段就是TTL字段,它用来控制查询消息的路由范围,TTL开始的时候被初始化为该查询消息所能通过的Domain Hop的最大值,每通过一个域,TTL的值递减1。

②当一个域边界路由器接收到一个查询消息时,它会根据不同的情况做出不同的处理。

·如果是本域内的查询消息,那么域边界路由器不作任何处理,将其转发到其他相邻的对等体;

·如果是从外部域边界路由器发送来的查询消息,那么首先检查本域内是否有到达该Anycast的最短路由信息。这里分成两种情况:第一种是本域中就包括这个Anycast组的成员,这种情况下,路由器就将本域的自治系统号附加到查询消息中的路径属性字段中,然后把这个路径属性字段作为应答消息的一部分直接发送给该查询消息的发起者;第二种是本域中不包括这个Anycast组的任何成员,但是它已经获取了到达该Anycast组的最短路径,那么它就从路由表里获取这个最短路径并附加在查询消息里的路径属性字段之后,然后把这个路径属性字段作为应答消息的一部分直接发送给该查询消息的发起者。如果路由器中没有到达该Anycast的路由信息,那么将查询消息中的TTL字段减1,检查此时的TTL值是否为零,并且检查路径属性字段是否构成环路。如果不为零并且没有构成环路,就继续将该查询消息发送给该路由器的相邻对等体。否则,丢弃该查询消息。

③当发起者发送查询消息以后,它会设置一个时钟来等待回应消息。然后,检查所有在时钟指定的时间内接收到的回应信息(路径属性字段),选择一个最短路径,将其保存到自己的路由表中,并将该路径发送给它的所有相邻对等体。

需要注意的是,以上描述的过程会出现很多意外情况,所以需要对这些意外情况加以分析和讨论。这里,最重要的是保证路由的正确性。

3.4 容错分析

在本方案中,维护路由信息的正确性是非常重要的。下面就几种情况来分析如何维护路由信息的正确性。

①如果域边界路由器对Anycast的最短路径组成员不可达,那么它会收到BGP发来的一个取消消息,这个消息促使该路由器放弃已经学习到的Anycast最短路由,并且重新发起一个查询该Anycast最短路径的一个查询消息。

②如果Anycast最短路径成员宕机或者离开该Anycast组,那么当Anycast数据包到达目的域时,该域的边界路由器发现该Anycast最短路径成员不可达或者不存在,则它会将该Anycast数据包转发到Anycast所在的主域,并且发送一个ICMPv6差错报文,通知发送该Anycast数据包的域边界路由器此路径不可达。

③发送查询消息的域边界路由器如果在时钟指定时间内没有接收到任何回应消息,那么它会发送一个ICMPv6差错报文给请求Anycast服务的主机,通知其该目标不可达。

3.5 性能分析

这里所说的性能分析是指Anycast服务在IP层实现与其在应用层实现的比较分析。

对于域内路由,由于本方案采用了传统的Unicast域内路由方法,所以排出路由算法的影响,两者获得的Anycast最短路径应该是一致的。

对于域间路由,考虑到3.4节中所提到的意外情况,域边界路由器发起的查询消息里的TTL值一般设置为2或者3是比较高效的。因为一般情况下,Internet的直径只有10个Domain Hop的距离,而域边界路由器一般经过2或者3 Domain Hop也会到达主干路由器。但是,由于域边界路由器发出的查询消息只发送给它的相邻对等体,所以这些消息一般很少到达主干网,因此也就不会影响主干网的运行性能。当TTL设置为3时,本方案所查询的Anycast最短路径的平均值一般会控制在其在IP层查询的Anycast最短路径平均值的1.2倍左右。此外,上述已经提到的影响Anycast服务扩展的主要原因是路由表会随全球Anycast组数呈比例增长,在本方案中,Anycast的路由表是可控的,因为主干网上的路由器只需要维护Unicast路由信息,而无需保存任何Anycast路由信息。从前边的分析可以看出,主干网路由器只要记录域边界路由器之间的路由即可,而由域边界路由器来维护本域内的Anycast路由信息(对于本域以外的Anycast路由信息并不关心)。由此可见,此处的路由信息是可控的,而且比在IP层实现Anycast服务大大节省了路由表资源。另外,由于Anycast和Unicast地址区分开来,也就是说Anycast的路由表和Unicast的路由表区分开来,所以Anycast服务的存在并没有降低Unicast的服务性能,并且由于Anycast无需处理网络前缀匹配的问题,所以可以采用更简洁的数据结构,对于路由表项的删除和修改也比Unicast更容易。最重要的是,本方案不会因为Internet网络的膨胀而影响其性能。在当前的Internet中本方案可以很容易地支持几百万个Anycast组。

结束语 Anycast是IPv6的一个新特性,它可以支持许多服务。本文在IPv6的模拟环境下,提出了在应用层实现Anycast服务的方案,并对该方案的可行性加以分析和讨论。Anycast作为一种新型的通信模式,具有广泛的前景,但是它还存在许多问题,有待进一步探讨和研究。

参考文献

1 Partridge C, Mendez T, Milliken W. Host anycasting service. RFC 1546, 1993
2 Deering S, Hinden R. Internet Protocol Version 6 (Ipv6) specification, RFC2460, 1998
3 Hinden R, Deering S. IP version 6 addressing architecture. RFC 2373, 1998
4 Hagino J, Ettikan K. An analysis of Ipv6 anycast Internet Draft. Internet Engineering Task Force, 2001
5 JohnSon D, Deering S. Reserved Ipv6 Subnet anycast addresses. RFC2526, 1999
6 Katabi D, Wroclawski J. A framework for scalable global IP-Anycast (GIA). In: Proc. of SIGCOMM, New York: ACM Press, 2000. 3~15
7 Narten T, Nordmark E, Simpson W. Neighbor discovery for IP version 6 (Ipv6), RFC 1970, 1996
8 Huitema C. Routing in the internet. Prentice Hall, 1996