第 33 卷 第 8 期 计 算 机 工 程 2007 年 4 月
Vol.33 No.8 Computer Engineering April 2007
·网络与通信· 文章编号：1000—3428(2007)08—0109—03 文献标识码：A 中图分类号：TP393

IPv6 中通过应用层实现 Anycast 服务的通信模型

王晓楠[1,2]，钱焕延[1]
(1. 南京理工大学计算机科学与技术学院，南京 210094；2. 常熟理工学院软件工程系，常熟 215500)

摘 要：提出了一种在应用层实现 Anycast 服务的通信模型，此通信模型不仅解决了 Anycast 服务的扩展性问题，同时也解决了 IP 层实现 Anycast 服务所存在的一些其它问题，如路由表膨胀、TCP 通信失败等，分析和讨论了该通信模型的可行性及其有效性，并且根据实验数据对本通信模型的服务性能进行了分析和总结。

关键词：IPv6；Anycast；ICMPv6；路由器

Communication Model of Anycast Service on Application Layer in IPv6

WANG Xiaonan[1,2], QIAN Huanyan[1]
(1. School of Computer Science and Technology, Nanjing University of Science & Technology, Nanjing 210094;
2. Software Department, Changshu Institution of Technology, Changshu 215500)

【Abstract】 A new communication model on implementing Anycast service is proposed in this paper and it solves not only Anycast scalability but also other existing problems, such as router table explosion, TCP communication errors, and so on, which are generally caused by performing Anycast services on IP layer. At last, the feasibility and validity of this new model are analyzed and discussed. Meanwhile, according to experimental data, the performance of this model is analyzed.

【Key words】 IPv6; Anycast; ICMPv6; Router

1 概述

Anycast 是 IPv6 所提供的一种特殊网络服务，它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。

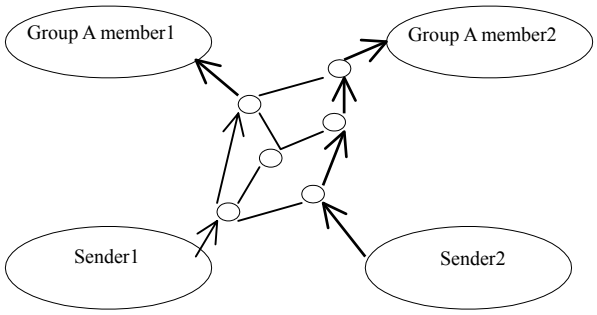

图 1 Anycast 服务

图中 Sender1 和 Sender2 都向同一个 Anycast 地址发出了服务请求数据包，但是该数据包被网络转发到距离发送者最近的一个组成员(接口)。

Anycast 有着广泛的应用，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，他不需要重新配置本地的 DNS，主机可以使用全局的 Anycast 地址访问任何地方的 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，对它的需求也不断地增长。

2 在 IP 层实现 Anycast 服务的性能分析

目前，很多 Anycast 服务在 IP 层实现，但是这种实现方式或多或少地存在着一些问题：

(1) IPv6 中的 Anycast 地址是从 Unicast 地址空间分配而来的，在语法上 Anycast 地址与 Unicast 地址没有任何区别，这就给 Anycast 的路由带来问题。可以这样设想，如果用 Unicast 路由协议路由一个 Anycast 数据包，而这个数据包的 Anycast 地址表示的可能是分散在互联网的各个地方的共享某个特性的结点组，那么每个全球 Anycast 地址必须作为一个独立的路由表项来处理。这就使得路由表会随全球 Anycast 组数成比例增长，从而导致路由表迅速地膨胀。因为这个原因，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，但是这种把 Anycast 组限制在一个预定义的区域内的做法，大大限制了 Anycast 组成员在整个网络中的广泛分布，进而影响了 Anycast 的服务质量；

(2) 在 IP 层实现 Anycast 技术会导致当前网络中的路由表变化频繁，从而造成客户和服务器之间的通信障碍；

(3) 在 IP 层实现 Anycast 服务的时候，Anycast 成员的距离度量只能通过 Hop 的次数来衡量，但是有些时候，这些

作者简介：王晓楠(1973－)，女，博士生，主研方向：网络安全与应用；钱焕延，教授、博导

收稿日期：2006-05-15 **E-mail：**wxn_2001@163.com

距离需要用其它的度量方式，例如 CPU 负载或者服务器当前的某些参数等形式来衡量。

以上存在的种种问题都阻碍了 Anycast 服务的进一步发展。因此，本文提出一种在应用层实现 Anycast 服务的通信模型，此通信模型很好地解决了 IP 层实现 Anycast 服务所存在的问题。

由于本通信模型的工作方式与 DNS 相关，因此本文首先介绍一下 DNS 的一些基本原理，然后再对该通信模型的原理和实现加以详细的分析和讨论。

3 DNS 基本原理

DNS 是一个分布式主机信息数据库，其数据结构类似于 Unix 文件系统的结构。整个数据库将根放在顶端，其结构像一棵倒立的树。树中每个节点都是 DNS 中的一个域，每个域可由不同的组织进行管理。各组织又可将其域再分成一定数目的子域，并委托其它组织进行管理。每个域中的数据通过 Client/Server 模式在整个网络上均可存取，一种称为域名解析器的程序担任 Client/Server 模式中的服务器部分。域名解析器包含了数据库中部分域的信息，它不但要提供有关它所负责的域的数据，而且还要能够搜索域名空间来找到不是它所负责的域的数据，供解析器的客户来访问。

目前，域名空间的最高域为根域；在根域下，按组织形式划分为 14 个一级域，按地理位置划分，为每个国家及地区设立一个一级域，并根据国际标准 ISO3166 建立了官方的、两字母的缩写表示其一级域名，如 cn 代表中国，fr 代表法国等。每一级的下面又相应地建立了多级域。DNS 的域名解析过程可以描述如下：

当主机向 DNS 系统提出域名服务请求时，该请求首先被主机的本地域的域名解析器获得，域名解析器查询本地共享数据库，如果本地区共享数据库中没有指定的域名资源记录，则域名解析器将向上一级域名服务器提出域名解析服务请求。

如果还得不到主机名与 IP 地址对应的解析关系，则继续向上一级域名服务器请求域名解析服务，直到根域名服务器为止。然后根域名服务器将该域名服务请求转向其所在的二级子域的域名服务器。这样，一级级传下去，最后，由此域名所在域的域名服务器查询其共享数据库，如果该共享数据库中没有域名对应的 IP 地址资源记录，则向请求方返回错误信息；如果该共享数据库中有此域名对应的 IP 地址资源记录，则将该记录传给请求方，并将该资源记录保存在请求方的数据库中以便以后的快速查询。这里需要说明的就是域名服务器与域名解析器的区别：本域的域名服务器响应其它域名解析器的域名服务请求，本域的域名服务器定期向其它域的域名服务器提出刷新本地共享数据库的请求，本域的域名解析器负责将本地无法解析的域名请求转发给上一级域名服务器，一个结点上的域名服务器与域名解析器是集成在一起的。

4 Anycast 在应用层的实现

在本通信模型中，Anycast 服务是在应用层实现的，一个 Anycast 组由一个域名来标识，而不是由一个 IP 地址来标识。由于 DNS 中的域名地址是分层结构的，因此为了区分 Anycast 域名与其它域名，本通信模型将 Anycast 域名的一级域名设置为 Any，例如：www.njust.edu.any。当本地域名解析器接收到域名请求时，它首先会判断该域名请求的一级域名的类型，如果一级域名是 Any，那么就将该请求转发给 Anycast DNS 服务器来处理，否则就转发到其它 DNS 服务器进行处理。如图 2 所示。

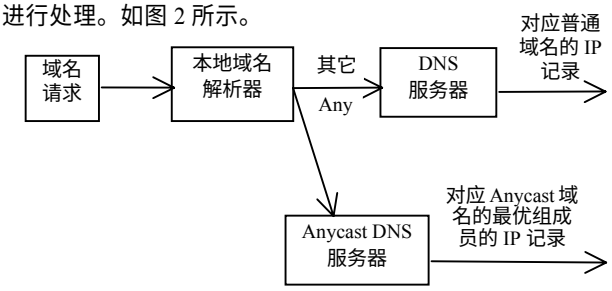

图 2 Anycast 域名转换处理流程

下面再来讨论一下 Anycast 域名寻址的实现细节。

Anycast 是一种通信模型，发送到这个 Anycast 地址的数据包会被路由到具有最短距离的 Anycast 组成员中去，这里所说的最短距离通常由所使用的路由协议来确定，一般可以是 Hop 数、服务器负载、到服务器的往返时间(RTT)以及当前可用的带宽等。

但是，实际上用户并不关心这些参数，他们真正感兴趣的是从发送服务请求到接收到服务应答之间的这段时间间隔，即总体应答时间(Total Response Time，TRT)，这段时间越短，客户认为服务质量越好。所以，本模型所采用的距离度量策略是 TRT，因为这个参数不仅反映了服务器负载的繁忙状态，而且反映了网络本身以及所建立的连接的某些属性(例如，带宽以及用户到服务器的 Hop 数)，所以 TRT 是一个综合性参数。

根据这个参数，本模型中的 Anycast 域名寻址实现过程可以描述如下：

(1)客户端向 DNS 系统提出 Anycast 域名服务请求，该请求中包含 Anycast 域名以及自身的 Unicast 地址，该请求首先被主机本地解析器截获；

(2)本地解析器根据请求中的域名类型判断出其为 Anycast 域名解析请求，所以，将该请求转发给权威 Anycast 域名服务器进行解析；

(3)权威域名服务器接收到此 Anycast 域名请求之后，查询其共享数据库，如果共享数据库中包含该 Anycast 域名资源记录，那么域名服务器则根据资源记录中的 TRT 值来获取最佳 Anycast 组成员；

(4)Anycast 域名服务器将获得的最佳 Anycast 组成员的 Unicast 地址直接返回给客户端。

整个过程如图 3 所示。

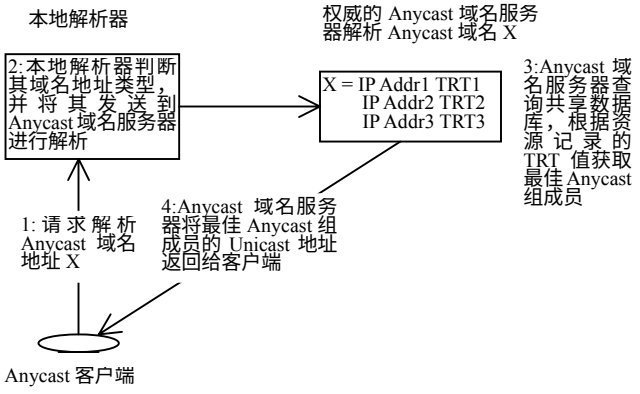

图 3 Anycast 域名寻址实现过程

在本模型中，Anycast 域名服务器的共享数据库中记录着每个 Anycast 地址所对应的组成员的相关参数，这些参数包括每个组成员的 Unicast 地址以及当前的 TRT 值。这里需要注意的是，每个 Anycast 成员的 TRT 值是个动态的参数，它会随着时间的变化而变化，所以为了确保每个 Anycast 成员 TRT 参数的有效性，Anycast 域名服务器必须定期发送查询消息给每个 Anycast 成员，以便及时地更新数据。在本通信模型中，Anycast 域名服务器采用如下的参数和算法来确定更新数据的频率：

参数包括时间间隔 L，最大阈值 T，步长 R，当前的阈值 M，其初始化值为 T。Anycast 域名服务器每隔 L 检测一次本服务器的当前客户请求流量，如果当前的流量值与上次检测到的流量值之差的绝对值大于 M，那么就发送查询消息给 Anycast 组中的每个成员，Anycast 成员接收到这个查询消息之后，计算当前的 TRT 并将其返回给 Anycast 域名服务器；如果当前的流量值与上次检测到的流量值之差的绝对值小于 M，那么就将 M 的值减去步长 R。这样，在客户流量比较平稳的情况下，Anycast 域名服务器至少能每隔 T/R 的时间单位发送一次查询消息，反之，它至多每隔 L 时间单位发送一次查询消息。

在上述过程中，Anycast 组成员可以利用如下公式计算出本次服务的 TRT 值：

$$TRT = S/BW + 3.5 \times RTT$$

其中，TRT 指本次服务的总体时间；S 指传输的数据总量；BW 指当前的带宽；RTT 指数据往返时间。

同样，Anycast 域名服务器在接收到 Anycast 组成员的反馈的 TRT 参数之后，将按照如下的过程来更新自身的数据库：

$$TRT=\alpha TRT_{old}+(1-\alpha)TRT_{new}$$

其中，TRT_{old}指当前数据库中的TRT值，TRT_{new}指Anycast组成员返回的当前服务的参数值，α是一个常量，可以根据网络性能的稳定性来决定α的取值，本模型取值为 0.25。

在本模型中，如果 Anycast 域名服务器在指定的时间内没有接收到 Anycast 组成员的信息反馈时，它会将相应的 Anycast 成员进行标记，以便以后进一步检测，如果多次检测此 Anycast 成员均不可达，就将其从共享数据库中删除。

5 性能分析

本通信模型与在 IP 层实现 Anycast 服务相比，具有很多优点：

(1)本通信模型有效地解决了在 IP 层实现 Anycast 服务所存在的扩展性问题；

(2)本通信模型是根据 Anycast 组成员的 TRT 这个综合参数来确定其是否为最佳成员，这样可以有效地保证客户能获得响应速度快的高质量服务；

(3)在 IP 层实现 Anycast 技术会导致当前网络中的路由表变化频繁，从而造成客户和服务器之间的通信障碍，本通信模型很好地解决了这个问题；

(4)在 IP 层实现 Anycast 技术，其距离度量只能通过 Hop 的次数来衡量，本通信模型很好地解决了这种局限性。

本通信模型在 IPv6 的模拟环境下已经成功实施，根据实验数据表明，由于 IP 层实现 Anycast 服务只能通过 Hop 的次数来确定 Anycast 最优组成员，因此它所出现的由于 Anycast 组成员繁忙而导致宕机或者丢包的概率要比本通信模型高出 4%左右，因此可以想象，在 Internet 网络上实施本通信模型要比 IP 层实现 Anycast 服务有着更好的性能。此外，对于用户最关心的服务响应时间，本通信模型并没有考虑 IP 拓扑结构的实际物理距离而是利用了 TRT 这个综合参数，根据实验数据表明，本模型的平均响应速度比在 IP 层实现 Anycast 服务的平均响应速度提高 1.2 倍左右。由此可见，这种处理方法对于真正庞大的 Internet 网络特别适用，因为现在数据传输介质的高效性使得实际的物理距离并不是提高服务质量的真正瓶颈所在，相反，网络的拥塞情况以及服务器的繁忙程度是能否提供高质量服务的关键所在，所以，在大规模的网络中，本通信模型能提供响应速度更快、服务质量更高的 Anycast 服务。

在本通信模型中，为了确保 Anycast 域名服务器中每个 Anycast 成员相关参数的有效性，服务器还需要定期发送查询消息给每个 Anycast 成员，以便及时地更新数据。由于本通信模型采用一定的算法来确定更新数据的频率，而且查询消息与其应答消息的数据传输量都非常小，因此它对于主干网络的性能基本不会造成影响。

最重要的是，本通信模型有效地解决了 Anycast 服务的扩展性问题，它可以使 Anycast 成员的物理分布不受地域限制，从而分布到世界上的各个角落。

6 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 服务的一种新通信模型，并对该通信模型的可行性、有效性以及综合性能加以分析和讨论。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[S]. RFC 1546, 1993.

2 Deering S, Hinden R. Internet Protocol Version 6(IPv6) Specification [S]. RFC 2460, 1998.

3 Hinden R, Deering. IP Version 6 Addressing Architecture[S]. RFC 2373, 1998.

4 Hagino J I, Ettikan K. An Analysis of IPv6 Anycast Internet Draft[Z]. Internet Engineering Task Force, 2001.

5 JohnSon D, Deering S. Reserved IPv6 Subnet Anycast Addresses[S]. RFC2526, 1999.

6 Katabi D, Wroclawski J. A Framework for Scalable Global IP-Anycast(GIA)[C]//Proc. of SIGCOMM. 2000: 3-15.

7 Narten T, Nordmark E, Simpson W. Neighbor discovery for IP Version 6 (IPv6)[S]. RFC 1970, 1996.

8 Huitema C. Routing in the Internet[M]. Prentice Hall, 1996.

9 王晓楠, 钱换延. IPv6 中通过应用层解决 Anycast 的扩展局限性问题[J]. 计算机工程, 2007, 33(4).