

IPv6 中 IP 层实现 Anycast 均衡负载的一种设计方案

王晓喃^{1,2} 钱焕延¹

(南京理工大学 南京 210094)¹ (常熟理工学院 江苏常熟 215500)²

摘要 本文在充分分析和研究了 Anycast 服务的基础上,提出了一个确保 Anycast 服务质量并且可以在 Anycast 成员之间实现均衡负载的全新方案,同时对该方案的实施过程做了深入的分析和讨论,充分证明了该方案的可行性及其有效性。

关键词 IPv6, Anycast, ICMPv6, 路由器

A Scheme on Implementing Load-Balance Anycast Service on IP Layer in IPv6

WANG Xiao-Nan^{1,2} QIAN Huan-Yan¹

(Nanjing University of Science & Technology, Nanjing 210094)¹ (Changshu University of Science & Technology, Changshu 215500)²

Abstract On the basis of fully analyzing and studying Anycast service, a new scheme which secures Anycast service quality and achieves loading balance between Anycast members is created in this paper. Meanwhile, the fulfillment and implementation of this scheme are deeply analyzed and discussed, fully proving the feasibility and validity of that scheme.

Keywords IPv6, Anycast, ICMPv6, Router

1 前言

随着计算机网络技术的迅速发展,新的应用需求不断出现。因此,需要研究新的协议和模型以满足不断增长和变化的需求。Anycast 服务就是在此基础上提出来的。

Anycast 服务是 IPv6 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度量来计算)。图 1 说明了 Anycast 的这种功能。

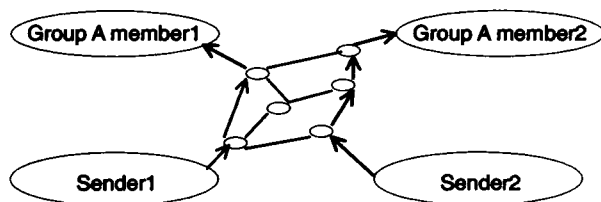


图 1 Anycast 服务

目前, Anycast 技术可以在 IP 层实现,也可以在应用层实现,但是在应用层实现存在着一些问题。因为在应用层实现 Anycast 技术, Anycast 服务不得不需要收集服务器主机是否在线和该主机提供何种服务类型,以及每个潜在客户和各个不同服务器间的以用户指定的标准度量的距离信息。为了收集这两种信息, Anycast 应用程序不得不频繁地探测服务器的状态,或者服务器频繁地向 Anycast 应用程序报告本机状态。当网络中的业务十分繁忙并且服务器数量很多时,网络可能不堪重负而导致整个网络瘫痪。由此可见,在应用层实现 Anycast 服务存在着复杂性和规模性问题。此外,现存的 Anycast 实现方案都没有考虑每个 Anycast 成员自身所能

承受的负载能力,所以单纯凭借某种度量单位计算的最短距离成员可能会同时处理超出其负载能力的多个数据包,从而导致其崩溃。

由于存在以上的问题,本文提出一种在 IP 层实现 Anycast 均衡负载的模型。此模型可以确保 Anycast 服务质量并且可以在 Anycast 成员之间实现均衡负载。下面对此方案加以详细的论述和分析。

2 Anycast 通信模型

Anycast 是一种通信模型,一个 Anycast 地址被分配给提供同一种服务的一组节点,发送到这个 Anycast 地址的数据包可以被路由到具有最短距离的 Anycast 组成员中去。在本模型中,每个与 Anycast 组成员直接相连的路由器都需要维护一个数据库,用来记录本地的 Anycast 组成员的参数,包括每个 Anycast 组成员的 Unicast 地址以及当前的会话数等等,如图 2 所示。

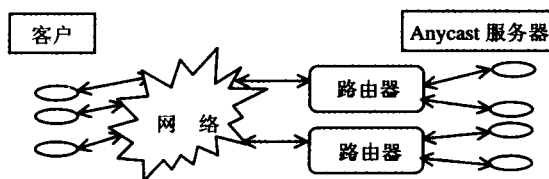


图 2 Anycast 模型体系结构

在本模型中,客户请求 Anycast 服务的过程可以描述如下:(1)客户端通过 DNS 服务器请求解析域名地址, DNS 返回域名所对应的 Anycast 地址;(2)客户端向该 Anycast 地址发送数据包,该数据包按照与 Unicast 相同的方式进行路由,最后到达与 Anycast 服务器相连的路由器,该路由器查询自

己的数据库,获取会话次数最少的 Anycast 成员的 Unicast 地址,并将该数据包转发到该 Anycast 组成员;(3)此后,Anycast 组成员与客户端之间开始利用各自的 Unicast 地址进行直接通信。

在本模型中,客户端的功能不需要修改,但是需要增加 DNS、主机以及路由器的某些功能。下面具体介绍本模型的实现过程。

3 Anycast 通信模型实现

本模型不仅解决了应用层实现 Anycast 服务的规模性和复杂性问题,而且确保了 Anycast 服务质量以及实现了在 Anycast 成员之间均衡负载的功能。

3.1 Anycast 成员加入和注销

在本方案中,一台主机如果想加入一个 Anycast 组,那么它必须向本地的路由器提出请求。此外,每个本地路由器还需要维护一个 Anycast 数据库,用以记录本地每个 Anycast 成员的 Unicast 地址以及每个 Anycast 成员的会话次数。一台主机加入 Anycast 组的过程可以描述如下:

(1)一台主机向本地路由器发送一个要求加入 Anycast 组的请求;

(2)路由器接到该请求后,对其身份进行验证,如果验证通过则允许加入,并发送一个允许加入的消息返回给主机,同时更新 Anycast 数据库,将新成员的 Unicast 地址加入到数据库中,会话次数设置成初始值 0,否则发送拒绝加入的消息;

(3)主机接收到路由器的消息后,如果是允许加入消息,那么该主机就发消息给本地的所有成员,通知其新的身份,否则放弃该消息。

下面再讨论一下 Anycast 成员从 Anycast 组注销的情况。一个 Anycast 成员从 Anycast 组注销的过程可以描述如下:

(1)一个 Anycast 成员向本地路由器发送一个 Anycast 注销请求消息;

(2)路由器接到该消息后,发送一个允许注销的消息返回给该 Anycast 成员,并且更新 Anycast 数据库,将该成员的表项从数据库中删除;

(3)该 Anycast 成员接收到路由器的消息后,向本域所有成员发送消息,通知它们自己已经注销了 Anycast 成员身份。

在上述(2)中,路由器在删除了请求注销的 Anycast 成员表项之后,如果此时的数据库为空,证明没有 Anycast 成员存在。那么,该路由器必须向相邻路由器发送消息,通知它们它们现在已经没有任何 Anycast 组成员。相邻路由器收到这个消息之后,首先从路由表中删除该 Anycast 地址对应的路由表项,同时发送一个查询消息给其他相邻路由器,以获取该 Anycast 地址所对应的其他成员的路由信息。

上述过程可以通过在 BGP 或者 IGMP 中增加新的消息类型来实现。但是,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。

3.2 路由器

在本方案中,Anycast 地址和 Unicast 地址是不加以区分的,所以路由器对于 Anycast 路由以及 Unicast 路由的过程基本是一致的。但是,这里有一点不同,就是当 Anycast 数据包到达与 Anycast 组成员直接相连的路由器的时候,该路由器会查询自己的 Anycast 数据库,找出会话次数最少的 Anycast 成员的 Unicast 地址,并将数据包转发到该组成员。此外,路

由器还需要接收 Anycast 组成员发送的服务开始与结束的消息,并相应地将该 Anycast 组成员所对应的会话次数加 1 或者减 1。另外,路由器还需要接收新的 Anycast 成员加入或者撤销消息,并根据其更新 Anycast 映射表,详见 3.1。本模型中,路由器需要维护一个 Anycast 数据库,用以记录 Anycast 组成员的参数,该表的数据结构见表 1。

表 1 Anycast 映射表的数据结构

Anycast 地址	对应的 IP 地址	会话次数
AnycastAddr _i	IP _{i1}	3
	IP _{i2}	2
	...	...
	IP _{in}	1
...	...	...

在本方案中,为了保证 Anycast 服务质量,在路由器接收到 Anycast 数据包之后,它会查找自己的数据库以获取最佳 Anycast 组成员。同时发送一个 ICMPV6 消息,测试一下该 Anycast 成员是否可达。如果可达,则转发给数据包;否则,路由器会选取另外一个 Anycast 成员(会话次数最少的 Anycast 成员)进行测试,同时将不可达的 Anycast 组成员的路由信息从 Anycast 映射表中删除或者进行标记(以便过一段时间继续测试其是否可达)。然后依照上述过程,直到找到一个可达的 Anycast 组成员为止。在某些极端情况下,可能 Anycast 组的所有成员都不可达,此时它将发送一个 ICMPV6 差错报文,通知源主机目的服务器不可达。

3.3 DNS 服务器

在本模型中,DNS 需要增加解析 Anycast 域名的能力。由于 DNS 中的域名地址是分层结构的,所以为了区分 Anycast 域名与其他域名,将 Anycast 域名的一级域名设置为 Any,例如 www.njust.edu.any。本模型要求 DNS 能解析一级域名为 Any 的域名,并返回相应的 Anycast 地址。DNS 解析 Anycast 地址的详细过程同其他类型的域名解析过程相同,这里不再赘述。

3.4 主机

如果一台主机要求加入一个 Anycast 组或者从该组注销,它需要发送一个请求消息给本地路由器。同时,Anycast 服务器还需要在服务开始和结束的时候,向本地路由器发送消息,以便其更新 Anycast 数据库。这里需要注意的一个问题就是 Anycast 组成员和路由器之间的通信要采用一定的安全措施,以防假冒和恶意攻击。

3.5 Anycast 地址与 Unicast 地址的转换

客户端发出 Anycast 服务请求数据包的时候,该数据包按照与 Unicast 相同的方式进行路由,最后到达与 Anycast 服务器相连的路由器。该路由器查询自己的数据库,获取会话次数最少的 Anycast 成员的 Unicast 地址,并将该数据包转发到该 Anycast 组成员。然后,Anycast 组成员与客户端之间开始利用各自的 Unicast 地址进行直接通信。这里一个值得注意的问题就是,客户端发出的申请 Anycast 服务数据包中的目的地址是 Anycast 地址,但是 Anycast 组成员返回的应答数据包中的源地址为其本身 Unicast 地址,那么客户端如何知道 Anycast 地址与其组成员的 Unicast 地址之间的对应关系呢?本方案中采用 IPv6 中的 Source Identification Option 字段来解决这个问题。Source Identification Option 字段的值设置为原来的目的 Anycast 地址。这样,当源主机接收到

Anycast 组成员的响应数据包之后,首先查看这个 Source Identification Option 字段,以便确认它是 Anycast 服务请求的响应数据包。如果是 TCP 连接,还需要将该字段提交到传输层,然后传输层将修改 TCP 连接管理表中的相应数据,以便以后正确通信。

但是,这种方法的缺点是 Source Identification Option 字段是在原有的 IP 协议中新增加的选项。如果采用这种方法实现 Anycast 地址和 Unicast 地址的转换,必须修改客户端的程序。所以,本方案提出了一种新的 Anycast 地址与 Unicast 地址的转换方法,即利用已经存在的 IP 选项 Source Route Option 来完成此功能,如图 3 所示。

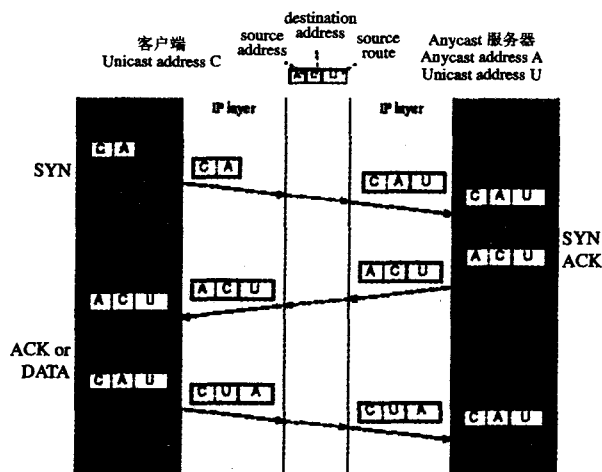


图 3 Anycast 地址与 Unicast 地址的转换

这里假设客户端的 Unicast 地址为 C, Anycast 地址为 A, Anycast 组成员的 Unicast 地址为 U, 那么客户端与 Anycast 组成员的通信过程可以描述如下: (1) 客户端发起 TCP 连接, 其源地址为 C, 目的地址为 A, 这个 SYN 数据包被路由器截获之后, 获取其最优的 Anycast 组成员的 Unicast 地址 U, 同时在数据包中增加一个 Source Route 选项, 并将 U 作为其元素(本方案中, 该字段只包括一个元素), 将其发送出去; (2) Unicast 地址为 U 的 Anycast 组成员接收到该 SYN 数据包之后, 通过目的地址中的 Anycast 地址可以判断这是发往本服务器的请求, 同时在 IP 层记录了 Anycast 地址 A 与组成员 Unicast 地址 U 之间的对应关系。该成员对该请求处理之后, 把 A 作为源地址, C 作为目的地址, 同时颠倒 Source Route 中的元素顺序, 将 SYN ACK 数据包发往客户端; (3) 客户端接收到该 SYN ACK 数据包之后, 其 IP 层通过数据包的目的地址 A 以及 Source Route 选项值可以知道 Anycast 地址与 Anycast 组成员的 Unicast 地址之间的对应关系。由于 Anycast 地址 A 为源地址, 而不是 U 作为源地址, 所以客户端的传输层无需修改 TCP 连接管理表中的数据就可以直接处理这个数据包。客户端在处理 SYN ACK 数据包之后, 为了确保将 ACK 应答数据包发送到正确的 Anycast 组成员, 它会将其源地址设置为 C, 目的地址设置为 Anycast 组成员的 Unicast 地址, Source Route 值设置为 A, 然后发送出去; (4) 服务器接收到 ACK 应答数据包之后, 因为 IP 层已经记录了

目的地址 A 与 Source Route 选项值之间的对应关系, 所以可以对该数据包直接处理。但是, 传输层并不知道它们之间的对应关系, 所以 IP 层在数据包处理完之后, 需要将该数据包的目的地址设置为 A, Source Route 的值设置为 U, 然后交给传输层去处理。这样, 传输层无需做任何修改, 就可以对这个数据包进行处理。此时, TCP 的三次握手完成, 连接正式建立, 通信双方可以利用上述方式进行传输数据。

这种方式的好处是无需修改 IP 协议, 同时不需要修改客户端程序。

4 性能分析

本模型不仅解决了应用层实现 Anycast 服务的规模性和复杂性问题, 而且确保了 Anycast 服务质量以及实现了在 Anycast 成员之间均衡负载的功能。因为本模型对客户端没有任何修改, 只对 DNS 服务器、路由器以及服务器的功能稍加改动, 所以它与当前的网络应用程序以及协议都能很好地兼容。此外, Anycast 成员的加入或者注销, 以及为了保证 Anycast 服务质量都需要发送和接收一些消息。因为这些消息的传输都是在本地完成的, 所以根本不会影响主干网的运行功能。在本方案中, 通过选取当前会话次数最少的 Anycast 成员, 可以达到负载均衡功能, 此功能大大提高了 Anycast 的服务质量。

本方案在模拟的 IPv6 环境下已经实施成功, 它无论是在服务响应速度还是容错性能方面都要比传统的应用层实现的 Anycast 服务提高很多, 从根本上解决了应用层实现 Anycast 服务的规模性以及复杂性的问题。

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, 实现了确保 Anycast 服务质量并且可以在 Anycast 成员之间实现均衡负载的 Anycast 路由方案, 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. RFC 2460, 1998
- 3 Hinden R, Deering S. IP version 6 addressing architecture. RFC 2373, 1998
- 4 Hagino J H, 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