第 19 卷第 4 期 广西科学院学报 Vol. 19, No. 4
2003 年 11 月 Journal of Guangxi Academy of Sciences November 2003

选播通信服务及其实现*

Anycast Communication Service and Its Implementation

陈 燕 宋 玲 李陶深
Chen Yan Song Ling Li Taoshen

(广西大学计算机与信息工程学院 南宁 530004)
(College of Computer & Information Engineering, Guangxi University, Nanning, 530004)

摘要 阐述选播通信服务的定义、功能及种类,分析应用层选播通信服务的实现方法和不足,针对应用层选播的不足,提出了网络层选播模式,并指出网络层选播通信服务中选播地址分配和对选播数据包转发的实现方法。

关键词 选播 应用层选播 网络层选播 选播地址 选播路由 地址映射
中图法分类号 TP393.01

Abstract The definition, function and type of anycast service are introduced. The implementations and limitations of application-layer anycast are analyzed. Then the network-layer anycast is given according to the limitations of application-layer anycast. The implementations of distributed anycast address and transmitted anycast data by network-layer anycast are also presented.

Key words anycast, application-layer anycast, network-layer anycast, anycast address, anycast routing, address mapping

随着越来越多的用户通过 WWW 来实现信息共享和查询,某个流行的站点可能因为访问用户过多而发生阻塞。为了增强服务的可用性和改善网络的流量分布,通常在网络中复制服务器,将多个具有相同内容的服务器通过网络连接在一起,共同向用户提供好的服务。例如 Internet 中的 WWW 镜像服务器、视频点播、股票服务器等都属于此类结构。复制服务器技术分本地复制和分布式复制 2 类:本地复制主要采用服务器组群技术,这些服务器都在同一个子网内;分布式复制是将服务器放置在不同的地理位置,通过 Internet 连接提供服务[1]。这些复制的服务器能够向多个客户同时发送信息,从而可有效地实现网络和服务器的负载平衡。采用一种新的服务模型——选播(Anycast)可以支持分布式的复制服务器,改善网络负载分布和简化某些网络应用,以满足人们对网络服务质量(Quality of Service,简称 QoS)的要求。

2003-06-10 收稿。
* 广西教育厅科技项目(桂教科研[2001]401 号),广西自然科学基金项目(桂科自 0229008),广西"新世纪十百千人才工程"专项资金(桂人字 2001213 号)联合资助。

1 选播服务

1.1 选播服务定义

在 RFCl546 中，选播服务首次被定义为：采用无状态的“尽力而为”方式，将选播数据包至少传输到一个具有选播地址的主机，最好仅仅传输到一个主机[2]。随着越来越多的应用需要选播服务，在 RFC2373 中又将选播服务定义为 IPv6 的一种标准服务模型[3]：一个选播地址可分配给一组接口（这些接口通常属于不同的节点）；发送到一个选播地址的数据包将根据选路协议对距离的量度转发至拥有此地址的“最近”的接口。这一定义说明：一组主机或服务器共享一个选播地址；用户数据包通过路由可以到达“最近”的一个。例如，多个 WWW 镜像服务器可共享一个选播地址，为了得到所需信息（如天气情况、股票数据等），用户可通过选播服务访问若干个服务器中“最近”的一个 WWW 服务器。

1.2 选播服务功能

这种选播机制实现以下功能：

（1）最近的服务器选择：客户机能与拥有指定选播地址的“最近”的服务器通信；

（2）提供抽象服务：选播地址可以作为一项服务的标识。如果网络上的每项服务都使用一个唯一的选播地址来标识，用户就可以通过该选播地址在任何地方获得最好的服务。若再能得到 DNS 对选播地址的支持，只需通过对服务名称的访问，即可获得服务。

（3）提高可靠性：选播地址的机制能提高服务的可靠性和冗余。选播地址可分配给网络上的分布式服务器，若其中一个服务器失效，其它服务器仍能保持对客户的服务。

（4）路由策略：通过选播地址结构可以实现基于策略的路由机制。

目前对选播通信的研究有 2 种不同的模式：一种是基于服务或应用度量（如负载量、响应时间等）的应用层选播；另一种是基于网络拓扑（如最少跳数或最小代价等）的网络层选播。

2 应用层选播

应用层选播是通过一些引导机制为客户提供访问选播服务提供者的方法，而这些机制并不需要在 Internet 中增加新的构成或服务[4]，例如不需要改变底层协议、不涉及路由器的修改、不需要路由系统提供支持等。应用层选播主要研究在当前使用的路由算法和协议处理策略不变的情况下，在 Internet 中如何实现选播路由服务。

2.1 应用层选播实现方法

目前应用层选播是通过修改 DNS（域名系统）服务器来实现的。修改 DNS 则是由一个将选播域名（Anycast Domain Name，简称 ADN）映射到多个 IP 地址的方案来实现。DNS 查询是访问 Internet 的最基本的方法，因此只需修改和增加 DNS 服务功能，就可以利用现有的资源实现选播服务，避免了由于增加额外的构成或服务所带来的复杂性。通常情况下 DNS 中一个域名对应着一个不同的 IP 地址，当用户进行域名查询时，DNS 解析器在关系表里查找匹配的记录，并返回相应的、唯一的 IP 地址，这样用户就能访问到该 IP 地址所代表的服务器了。为了支持复制的服务器，要对 DNS 作一些修改，以实现一个 ADN 到多个不同 IP 地址的映射。当客户需要选播服务并发出一个 ADN 查询时，选播解析器处理该查询后向客户端返回多个不同的 IP 地址。在客户端通过基于某种策略或度量标准（如响应时间）设计的过度器对返回

的多个 IP 地址过虑，最终向用户提供一个最合适（如响应时间最短）的 IP 地址[4]。

2.2　应用层选播的不足

应用层选播服务的不足有：(1) 可扩展性差，这是因为应用层选播是通过修改 DNS 实现的，因此每增加一个相同服务的服务器就要浪费掉一个 IP 地址，并且要修改 DNS 中存放的内容；(2) 在通信中用户需要知道 DNS 服务器的位置；(3) 在网络连接之前 DNS 的映射信息就要存在；(4) 每次访问都必须解析域名；(5) 无法判断指定服务器的可用性。当一个服务器出错时客户无法及时了解，访问就会继续，即使是网络管理者发现错误并及时修改 DNS 的记录也无法保证客户的实时访问性。主机最近服务（Host Proximity Service，HOPS）和分布式控制器（Distributed Director）的引入在一定程度上能解决这个问题，但是这种方法也都要求每个连接开始之前客户必须与 HOPS 服务或分布式控制器通信，这就造成了额外的开销和资源的浪费。

总的来说，应用层选播不了解网络的拓扑变化和负载情况，也就无法根据距离或拥塞来决定访问的服务器。这对于要求网络服务质量的应用来说是致命的。事实上应用层选播并不是真正意义上的选播，因为应用层选播的一组复制主机并不共享一个选播地址。因此，为区别于网络层的选播，人们也称应用层的选播为“伪选播”。

3　网络层选播

选播的初衷是支持分布式复制服务器，在网络上实现负载平衡[6]。但是处在应用层的选播无法及时了解网络的动态变化，难以实施进一步的应用。网络层选播只要求为提供相同服务的一组服务器分配同一个选播地址，由路由控制系统完成对服务器的查找、定位。由于对服务器的选择完全在网络层完成，不需要用户参预，也不需要增加新的功能服务器（如 HOPS 服务器，分布式控制器）。因此，人们开始致力于网络层选播的研究，主要解决如何为提供服务的主机分配选播地址和对选播数据包进行正确的转发。

3.1　选播地址

在下一代的互联网协议 IPv6 中定义了一种全新的选播地址，可以同时分配给多个不同的接口。预定义的 IPv6 选播地址格式如表 1 所示。

选播地址中的子网前缀标识一条特定的链路，这种选播地址与接口标识填充为零的同一链路上的单播地址有着相同的格式。IPv6 特别说明，所有路由器都必须支持其所在子网的子网路由器选播地址，发送到子网路由器选播地址的数据报可以转发到子网中任一路由器。同时规定，由于主机无法宣告接收选播包的意图，不允许将选播地址分配给 IPv6 主机，只能分配给 IPv6 的路由器；由于数据报文出错时，必须确定其起始节点，所以选播地址也不允许作为 IPv6 源端地址。

表 1　IPv6 选播地址格式

项目	格式
n 比特	128-n 比特
子网前缀	00000000000000

为了扩展选播的应用，引入主机路由的概念允许选播地址分配给主机。主机路由可以由主机本身具有路由宣告功能，也可以由与之相连的路由器实现它的主机路由宣告。主机只需向路由系统宣告它的选播地址，而不必参预整个路由的信息交换或者实现路由的其它功能。路由器可以通过 Ipv6 的邻居发现路由协议（ND，Neighbor Discovery）或多播侦听路由发现协议（Multicast Listener Discovery）发现同一链路上的主机选播地址。

在 RFC 2373 中，IPv6 已经保证了对选播地址的支持，因此，对网络层选播的研究应该主要集中在对选播数据包进行正确的选径转发问题上，以保证客户访问到的主机是选播组成员中“最近”的一个。

3.2 选播数据包的转发

为了在网络层实现选播服务，路由器必须能识别选播地址并能正确转发数据。选播路由机制将根据协议对距离的量度转发选播数据包到拥有同一个选播地址的“最近”的节点。但是节点的路由会随着网络拓扑结构的改变而改变，到达选播组节点的最短路径就有可能改变。因此，同一个源节点的后续选播数据就有可能发送不到原来进行通信的同一个节点。特别是应用无连接的路由协议时尤为如此。如图 1 所示，图中 $S1$ 和 $S2$ 具有相同的选播地址 M，主机 $H1$ 发送信息探测到与服务器 $S1$ 最近，因此 $H1$ 在与 $S1$ 发生通信。当数据包 $P3$ 到达 $S1$ 以后，网络的拓扑发生了变化，此时 $H1$ 探测到的是与服务器 $S2$ 最近，因此 $H1$ 开始与 $S2$ 通信。这就造成了后续的数据包 $P4$ 等无法到达同个服务器 $S1$。

在 IPv6 中，为了解决这些问题，规定选播地址只能在通信开始时使用，以避免传输路径改变时带来的负面影响[7]。我们知道在网络通信开始前 TCP 必须首先为通信的双方建立一条连接，确认双方的身份和状态，成功建立连接后通信才开始。因此可规定在实施选播通信时，选播地址只在建立连接时使用，也就是利用选播地址找到合适的“最近”服务器，并且在建立连接过程中把该服务器的单播地址返回给发送方，连接建立后的通信就变成发送方与该服务器的点对点的单播通信。可通过源标识符选择器或选播地址映象器的实施从选播地址到对应单播地址的映射，以保证选播服务的连接通信。

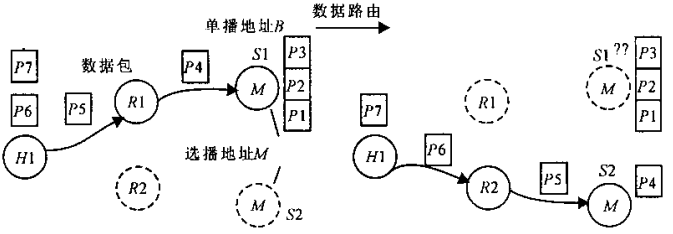

图 1 数据包的转发

3.2.1 源标识符选择器

在 1996 年出版的 Internet 草案中，将使用新 IP 选择器的简单机制称为源标识符选择器。在这种选播机制中，任何带有选播地址的节点都能使用选择器通知其相对应的单播地址。由于目前的传输层协议不支持选播地址服务，可以通过建立 TCP 三方握手连接的机制，使用综合 IP 层与传输层的源标识符选择器来实现地址的映射。图 2 给出了建立 TCP 连接过程示意图，其中在图 2 (a) 中，客户机向选播地址 M 发送一个 SYN 信息包；带选播地址 M 的服务器收到 SYN 信息包后，向客户机发送 SYN＋ACK 信息包应答，同时把它自己的单播地址 B 作为源地址发送，代替选播地址 M 作为目的地址、并附加源标识符选择器作为目的选择器；客户机收到 SYN＋ACK 信息包后，去掉信息包头部的“源标识符选择器”，确认该信息包是否为 SYN 信息包的应答，然后将服务器地址由 M 改为 B；客户机通过发送 ACK 信息包到单播地址 B 以建立连接。

3.2.2 选播地址映象器

选播地址映象器是一种新的选播地址映射机制，它利用 ICMP ECHO 请求/应答找到选播

地址所对应的单播地址。客户机在与选播服务器连接之前，先使用选播地址映象器找到选播服务器对应的单播地址；然后客户机再使用该单播地址作为目的地址与服务器建立连接。如图 2（b）所示，客户机向选播地址 M 发送 ICMP ECHO 请求信息包；服务器收到该请求后，把其单播地址 B 作为源地址，向客户机发送 ICMP ECHO 应答；客户机收到此 ICMP ECHO 应答信息包后，则将服务器地址 M 替换为 B；客户机向单播地址 B 的服务器建立连接。

通过上面 2 种地址映射方法建立的连接，可以避免必须连续的多个数据到达多个服务器。例如，图 1 的客户机 $H1$ 发出地址 M 的选播服务请求；“最近”的服务器 $S1$ 响应；$S1$ 向 $H1$ 返回自己的单播地址 B；$H1$ 把 B 作为目的地址开始通信。实现了 $H1$ 所有的地址 M 的选播请求服务都由 $S1$ 提供，避免了后续的包 $P4$、$P5$ 等到了其它的服务器 $S2$。

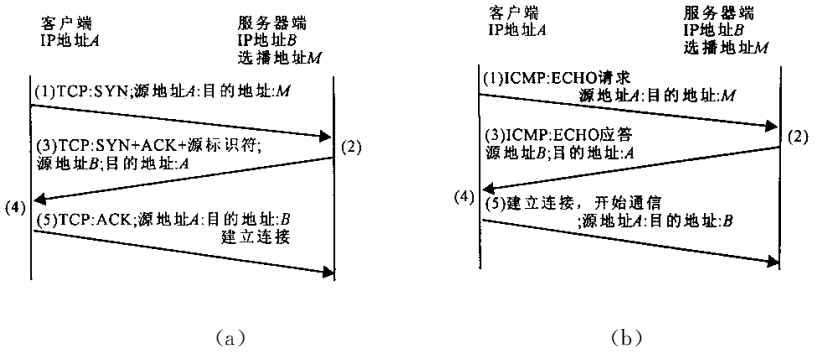

图 2　建立 TCP 连接过程示意图

（a）源标识符选择器方法；（b）选播地址映象器方法

可此可知，对选播服务器的定位由路由控制系统利用选播地址实现；找到合适的“最近”服务器后要通过地址映射保证双方的一致通信。这样，网络层的选播就可以实现。

4　结束语

从本文的分析可知，选播服务是一种新的服务体系，它对于平衡网络和服务器负载、优化网络资源都十分重要。选播服务可以在应用层实现，也可以在网络层实现。尽管应用层的选播服务不需要修改路由系统，在现有的网络基础上实施比较简单、易行，但是它并不适应网络的快速变化和发展；而网络层的选播则充分利用了 IPv6 提供的选播地址，由路由控制系统支持实现，能实时响应网络的变化要求。因此，对网络层选播的研究势必是今后人们对网络服务模型研究的方向。我们将开展选播组的管理、选播数据包进行路由的适用协议开发、以及相关算法的设计。

参考文献

1　Zegura E et al.. Application-layer anycasting：a server selection architecture and use in a replicated Web service. IEEE/ACM Trans Net，2000.
2　Partridge C，Mendez T，Milliken W. Host Anycasting Service. RFC 1546，1993.
3　Hinden R，Deering S，IP version 6 addressing architecture. RFC 2373，1998.
4　Bhattacharjee S et al.. Application-layer anycasting. Proc of IEEE Infocom'97，1997，1390～1398.
5　Basturk E，Engel R，Heas R et al.. Using network layer anycast for load distribution in the internet. IBM Research Report，1997，20938.
6　Hinden R，Deering S. IP version 6 addressing architecture. RFC 1884 IFTF，1995.
7　Bound J，Roque P. IPv6 Anycasting Service：Minimum requirements for end nodes. Draft-bound-anycast－00. txt(EXPIRED)，1996.

（责任编辑：黎贞崇）