

第8章 IPv6选路

本章首先讨论寻址和地址分配对 IP 网络的影响，然后讨论 IPv6 选路与 IPv4 选路的区别，重点介绍 IPv6 选路协议，也将讨论与选路相关的不同传输类型——单播、任意点播和组播。

8.1 地址对 IP 网络的影响

追溯至 70 年代后期、80 年代初期，IP 刚诞生的时候，几乎无人想到 IP 和 Internet 会发展为上万个不同网络、数千万个主机的规模。在描述早期 IP 实现的文档 RFC814(名字、地址、端口和路由)中，只使用了 32 位地址中的 8 位来标识网络。即，这些互联网络最多只支持 256 个网络。即使较复杂的实现也使用比较简单的寻址机制，即单个网络使用一个选路表项来指定，每个网络内部的单个主机使用一个主机表项来指定。

主机名和网络域名与主机地址和网络地址是通过简单的表链接到一起的。如果一个主机的网络地址改变了，例如由于网络重构而导致地址变化，就必须更新相关的表。如果一个网络域的地址改变，也必须更新选路表。主机地址的变化只需要在主机所在的域内进行更新，而网络地址的变化还需要对外部路由器的表进行更新。通过使用域名系统 (DNS) 服务器可以简化这种情形，而 DNS 还有待充分地规范和实现。在 DNS 的支持下，节点可访问 DNS 服务器以查询与主机名字相对应的网络地址。因此应用程序无需考虑 IP 地址，除非主机名所对应的 IP 地址可能改变。

然而，使用 IP 地址作为主机或节点的全球唯一标识已经有很长的历史，而且暂时还很稳定。即，不仅每个 IP 主机和网络都是通过唯一地址标识的，而且在一段时间内，该地址将保持不变。直到 90 年代中期，这种方案的效果一直很好。当 Internet 作为一种通信媒体，大规模地提供给各机构和个人访问，如同使用电信业务一样，此时 IP 地址的使用和分发也随之发生了变化。此前，大多数使用 IP 和 Internet 的公司直接向负责编址的授权机构申请网络地址和网络域，直接负责自己的 Internet(或 Internet 的前身，如 NSFNet 或 ARPANet) 连接，或与某些专业网络厂商(如 Bolt, Beranek 和 Newman，即 BBN) 合作负责。

但是，当 Internet 进入商用之后，情况就发生了变化。尤其是随着负责编址的授权机构对地址进行严格管理，单独的机构不再直接控制其 IP 地址。这些授权机构把编址任务交给 ISP 来代理，并且与 CIDR 共同使用，这样就可以对路由进行集聚。由于选路表的膨胀，集聚路由成为一个重要特性。

这种趋势导致 IP 寻址发生了巨大变化。首先，若一个机构改变了其 ISP，可能必须要随之改变其网络地址。其次，由于对 IP 地址的控制更加严格，一个有 500 个节点的机构可能只能得到 255 个节点的地址空间。本章将介绍一些与 IP 寻址机制相关的 IP 寻址分支，以及这些分支与 IPv6 选路的关系。

8.1.1 标识符和定位符

RFC 2101(目前 IPv4 地址行为) 发布于 1997 年 2 月，该文档描述了 IPv4 地址的使用如何随时

间的推移而变化。它的要点在于对标识符和定位符的使用进行了区分。文档中将标识符定义为“两台主机的通信会话的整个生存期内使用的位串，用于对其中一台主机相对于另一台进行标识”。即，在用于Internet通信时，标识符看起来类似源主机的IP地址。而定位符被定义为“用于对某个特定包必须交付的位置进行标识的位串，例如它可用于在Internet拓扑中对目的主机所连接的位置进行定位”。即，定位符看起来类似目的主机的IP地址。

因此，标识符用于标识源端，而定位符用于标识目的端。这样做很直观，也很合理，主机IP地址既可以用作标识符，也可用作定位符。但是给予定位功能（即发现目的地）的优先权高于标识功能（即了解数据的源头）。即，与能够准确了解包的源头却不能交付该包相比，能够首先交付包然后再找出其源头更重要。

RFC 2101的作者指出，对于标识符和定位符的要求有两个重要区别：一是唯一性，二是持久性。

首先讨论唯一性。对于通信节点双方来说，标识符必须是唯一的，即各节点之间进行通信时，其标识符都必须唯一。有唯一合法IP地址的主机能够通过识别有唯一合法IP地址的任何其他主机，且连接到同一个互联网的所有此类主机都是唯一的。而另一方面，对于相互通信的路由器而言，定位符仅在某些情况下要求是唯一的。即，在同一选路域内，定位符必须唯一，但在不同的选路域内，定位符可以重叠。例如，一个路由器可以将10号网络连接到其他网络，但不能将两个或多个10号网络互相连接，否则即使规范没有禁止转发10号网络的包，路由器也不知道该向哪条链路上发送目的地址为10号网络的包。

现在考虑持久性。标识符的生存期要比定位符长。标识符至少要保持到两个节点间的通信结束。如果在通信过程中，一个节点的标识符有所改变，另一个节点则无法对后续包正确寻址。而另一方面，定位符只在相关的选路机制需要时才起作用。即，对于在节点通信过程中定位符改变的情况，路由器有能力进行处理。

目前，尽管定位符和标识符大多来源于节点的IPv4地址，但两者还是有不同的属性，且其理想化特性不同而且不一致。例如，理想的标识符只在节点初次安装到网络上时分配一次，其后永远不变。理想的标识符与一个节点相捆绑，并且只捆绑到一个节点，不能再重新使用或重新分配，这样就可以一直将该标识符与该节点相链接，而不会链接到其他节点。总之标识符的功能是将节点作为数据源进行标识。

而另一方面，定位符用于确定包必须向何处发送，它不需要持续很长时间，但是它应该描述在网络拓扑中节点实际所处的位置。这样，如果主机在网络拓扑中的位置由于某种原因而改变，定位符也随之改变。例如，如果主机从一个网络中迁移到另一个网络中，在理想情况下，其定位符应该改变。同样，如果主机所连接的网络重新编号，主机的定位符也随之变化。

RFC 2101的作者已注意到，不论作为定位符或标识符，IP地址都不理想。由于IP地址不再是全球唯一的，例如网络号10代表了共享同一网络和主机地址的相当大一部分IP节点，因此它不是理想的标识符。同时IP地址缺乏持久性的情况越来越多，因此它更不适合作为标识符。对于依靠DHCP来分配临时IP地址的网络，今天这个IP地址由一个节点使用，明天可能由另一个节点使用。

同样，IP地址作为定位符也有其不足。其一，网络号10无法说明此节点在互联网中的位置。其二，由于历史原因，网络地址无法说明该网络与其他网络的位置关系。当然随着越来越

越多的网络路由使用 CIDR 进行集聚, 这种情况发生了一些变化。某一 CIDR 块内的网络地址通常由负责该块的机构来处理。但是, 对于 B 类网络, 或在 CIDR 广泛应用之前已分配地址的网络, 其地址无法说明此网络在 Internet 中的位置。其三, 如果改变 ISP, 网络拓扑随之发生变化, 但是除非该机构对网络重新编号, 否则网络地址无法反映这种拓扑的变化, 而重新编号又使 IP 地址作为标识符的稳定性受到影响。

8.1.2 地址分配、无缝互操作和网络拓扑

RFC 2008 (Internet 选路的不同地址分配策略的含义) 发布于 1996 年 10 月, 该文档提出了有关 IP 选路的一些问题, 并描述了地址分配的“当前最好惯例”。此文档的基本前提是对“地址借出”方法的研究, 相对于传统的“地址所有权”方法, 该方法极大地改善了性能和扩展性。

换言之, 该 RFC 鼓励能够采用地址借出方法的机构使用由其 ISP 分配的 IP 地址, 一旦机构改变其 ISP, 地址也要随之变化。这意味着该机构是向 ISP 暂借其 IP 地址, 而 ISP 负责为客户集聚业务。通过集聚, ISP 只需维护更少的路由, 且 Internet 的整个扩展性得以改善。但是, 集聚也意味着如果机构决定改变 ISP, 就必须改变其 IP 地址, 以便新的 ISP 可以对其路由进行集聚。

另一种方法——地址所有权方法导致了地址表的急剧膨胀。但数据流是由其他节点导向一个特定的 IP 地址时, 即在使用 IP 地址作为标识符时, IP 地址还是有很好的实用价值。

然而, 使用 IP 地址作为标识符将导致很多问题。首先, 在处理网络业务、升级或改变节点功能方面, 用户因此损失了相当多的灵活性。使用 DNS, 用户可以将一个逻辑名字 (如 www.ioshinc.com) 捆绑到一个地址, 该地址可能随时间变化。例如, 用户很容易将其 web 站点的捆绑从已过时的 80486 微机上移到第三方 web 呈现供应商所运行的高端 SMP 服务器上, 用户只需将其逻辑名的 DNS 映射从一个 IP 地址改变为另一个 IP 地址。

灵活性的用途很大。在上例中, Internet 呈现供应商需要将用户的域名映射为自己服务器的 IP 地址, 需要为用户个人系统分配新的主机名。更重要的是, 必须要求 IP 应用程序只使用逻辑节点名, 而不能使用 IP 地址, 这样这些应用才能在 IPv4 和 IPv6 链路上无缝互操作。如果应用程序只涉及逻辑节点名, 就可以采用其他方法来实现节点名和节点地址的映射。有关取决于 IPv6 的协议这方面及其他方面的论题将在第 10 章讨论。

使用地址借出方法的最大优点是 IP 地址可以反映出网络拓扑。在图 8-1 中, ISP B 为 Acme 公司分配了一个网络地址, 这样该公司就可以通过 ISP B 连接到 ISP Q, 再通过 ISP Z 连接到 Internet。如果所有的 IP 地址都借给用户, 网络的性能和可扩展性就可以获得显著提高。对于要发送给 Acme 公司的包, 图中的源节点知道首先要选路到 ISP Z, 对于链接到 ISP Z 左边的所有网络, 都使用该路由。如图所示, 在 ISP Z 只有三条路由, 分别连接到其客户 ISP P、ISP Q 和 ISP R。同样, ISP Q 也只需要有三条路由, 分别连接到其客户 ISP A、ISP B 和 ISP C。

Acme 公司从 ISP B 借用 IP 地址, 如果改变 ISP, 就必须重新编号。例如, 该公司认为 ISP A 能提供更好的服务且价格更低。此时, ISP Q 就必须改变其选路表以呈现 Acme 公司的新地址, 但是对于已经向 Acme 公司发送的包, 更高层的 ISP 和 Internet 内另一侧的路由器仍然通过 ISP Z 进行选路。Acme 公司可以决定自己需要只来自上一层 ISP 的高层服务, 此时需要改变更多的路由。但是, 该公司应该注意到地址借出能够显著改善扩展能力和性能, 代价是一旦改变 ISP, 或其 ISP 改变了上级 ISP, 就必须对该公司的网络重新进行编号。RFC 2008 的作者指出如果使用某种 NAT 或应用网关, 可以无需对内部主机重新编号。他们还指出, 集聚和地址借出的目

的不应是构造尽可能小的选路表，而应是减缓现有选路表的增长速度，以确保额外的增长不会影响到选路的性能。

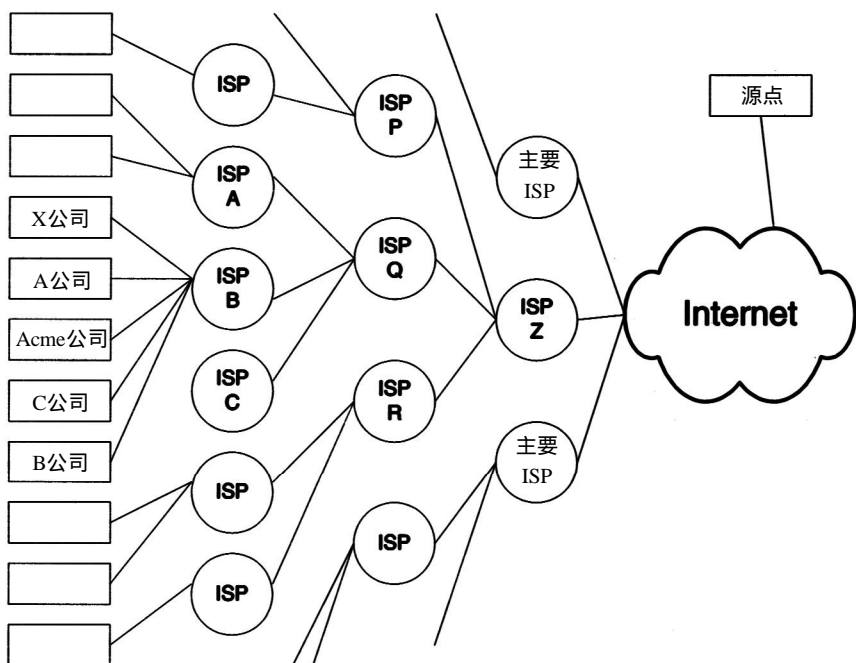


图8-1 IP路由集聚和地址借出使IP地址能反映Internet拓扑

本节提出的问题和IPv4寻址机制密切相关，但是这些RFC的写作都是以向IPv6升级为前提的，指出IPv4的可以改进之处有助于说明IPv6寻址从何处及如何改进。

8.2 选路问题

目前好像几乎人人都知道 IPv4网络地址即将耗尽。另一个问题却不是如此显而易见，即非默认路由器，或者是列出Internet上所有路由的路由器，即在Internet骨干网上或骨干网附近、因而必须知道全部路由的路由器，它们如何处理日益庞大的路由表。路由表中必须列出到达所有独立网络的路由，因此 CIDR广受欢迎。使用 CIDR，一个上述骨干路由器可以用一个涵盖8位地址空间的CIDR路由代替256个C类网络的256条路由。所有的256条路由可以经由一个Internet访问供应商来选路，因此CIDR可以显著减少映射到Internet所需要的路由数目。

IPv6没有IPv4中的地址类别的概念。不论 A、B、C类地址的存在对于IPv4如何有用，长期以来这种分类都是对地址的浪费，对于网络地址体系结构，子网或超网能力好像用处更多。而且出于选路目的，IPv6地址可以积累起来，理论上有很大潜力可以显著地减少非默认选路表的大小。

当然，这种高度集聚的体系结构也有缺点，即一旦一个机构改变其供应商，就必须对网络重新编号。同样，多宿主网络可能引起更多的问题。实际上，基于供应商的 CIDR模式集聚方法的反对者把这个问题称为“专制”，他们已经提出了替代方案。很显然，这些替代方案在IPv6中没有采纳，但是这些方案有助于使自动配置和供应商移动性成为 IPv6过渡策略的关键部分。自动配置和供应商移动性的相关机制将在第11章中介绍。

看起来IPv6选路协议和IPv4选路协议好像没有显著的不同,这一点也许会令人吃惊。毕竟IPv6寻址体系结构自身将显著改进选路效率,并减少非默认选路表的大小,因此选路算法和协议只需要进行极少的修改便可取得更好的执行效果。为支持 IPv6,对这些协议所做的修改大部分都与如何处理较长的IPv6地址有关。

IP选路协议

IP选路协议实质上可以分为链路状态协议和矢量距离(或路径矢量)协议两类,也可以按照内部选路和外部选路来分类。这两种分类方法看起来很简单,但足以满足本书的要求。

1. 内部选路和外部选路

内部选路和外部选路的概念对 Internet的结构非常重要。这两个概念与 Internet以及相连的互联网络之间的交互方式密切相关。例如,某个公司的内联网通过一条链路与 Internet相连,该内联网与全球 Internet之间的全部业务流都经由该链路来传送。如果这条链路中断,内联网就不再具有外部连接能力。这种类型的网络称为自治系统(AS),因为网络内部的一切都由单一的管理机构来管辖。这种系统的自治性体现在如果想要访问系统内的任何节点,全球 Internet路由器只需要了解一条路由。同样,AS内的任何节点可以使用默认路由来向AS之外的任何节点发送包。默认路由用于标识链接该AS与全球Internet的路由器。

内部选路参与AS内部包的选路。换言之,在相对小型的互联网络内的选路,所谓小型是相对于全球 Internet而言。AS内部的路由器保持的路由表相对较小,这些路由表包含到AS内部的子网和网络的路由,如果包要寻址的网络的地址没有明确列在路由表中,那么表中也包含一条或多条此时要选的默认路由。

另一方面,外部选路发生在AS环境之外。骨干路由器不能有默认路由。作为骨干路由器,如果要正确地发挥其作用,它就必须了解每个目的网络的显式路由。这意味着它的选路表将非常庞大,由此可见,使用诸如CIDR之类的集聚机制对于改善骨干选路性能极为重要。

如下所述,所有的选路协议都使用了链路状态选路算法和距离矢量选路算法中的一些要素。目前的IPv4外部路由器依靠RFC 1771中定义的边界网关协议4(BGP-4),这是一种支持CIDR的距离矢量选路协议。尽管最初有一些评论(参见Huitema所著《IPv6 The New Internet Protocol》)认为BGP非常适用于IPv4的32位地址,但完全不适于IPv6,但是BGP-4似乎仍将用于IPv6中的外部选路。有一个Internet草案中描述了为能够正确处理IPv6单播地址(链路本地地址、站点本地地址和全球地址)而进行的某些扩展,RFC2283中指出,经过上述扩展和支持多协议选路所进行的相对少量的修改,BGP-4将能够处理IPv6外部选路。

另一个重要的外部选路协议是域间选路协议(IDRP),该协议源自ISO/OSI的努力。IDRP在描述实质上规模无限的网络和支持网络地址体系结构方面提供了更大的灵活性。有些人认为IDRP是更好的IPv6外部选路协议,但是随着Internet的进一步发展,IDRP是否能继续保持其重要性,这一点还存在质疑。

2. 链路状态和矢量距离协议

通常矢量距离协议较简单。选路信息协议(RIP)就是一个重要的矢量距离协议,该协议很简单,但是,因为它要求互联网络中的每个路由器都要周期性地向网络中所有其他路由器广播自己的选路信息,故该协议有一定的局限性。正如其名称所示,每个路由器所广播的信息包括路由和表示该路由长度的整数的列表。见图8-2,网络中的路由器1到网络A的路径距离是

1跳，到网络B、D、F的路径距离是2跳，到网络C和E的路径距离是3跳。图中路由器2到网络B是单跳路径，到网络A是一条2跳路径。

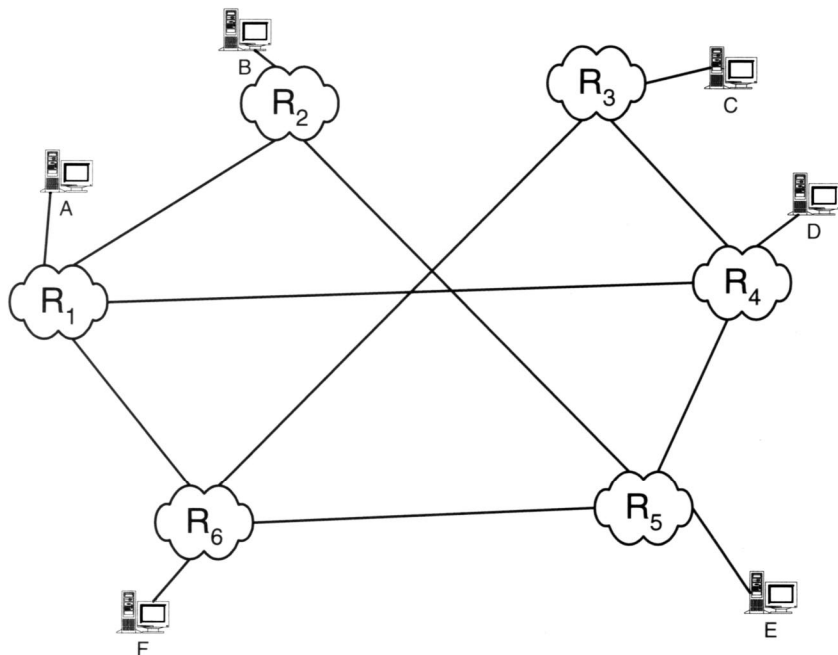


图8-2 路由器使用RIP链接而成的一个简单互联网络

使用这个广播信息，路由器2得知任何要发往网络A的包应该向路由器1选路，因为其他路由器没有通告到网络A的更短路径。同样，路由器2知道应该避免把发往网络A的包转发给路由器5，因为该路由器与网络A的距离为3跳。路由器2也清楚路由器3与网络A的距离为3跳，与自己的距离也为3跳，因此如果经由路由器3来转发给网络A将需要6跳。

显而易见，RIP有一些缺陷。首先，该协议“噪音”很大。每个路由器都频繁地发送状态信息，默认情况下每30秒就发送一个报文，这样随着互联网络规模的扩充、路由器数目的增加，整个网络的业务流将急剧增长。其次，由于RIP定义的限制，它只能支持不超过16跳的互联网络，即，如果路由的长度超过16跳，就不能使用RIP来选路。虽然RIP头可以支持的跳长度多达232-1，但由于选路收敛问题，开发者们在早期就确定RIP应该限制在16跳之内。所谓选路收敛问题，是指在不正确路由通过网络传播之前，有关连通性问题的报文无法转播报给所有的路由器。

选路协议有必要允许包的选路独立于网络拓扑。这意味着源节点不必在内存中保留Internet的拓扑结构，也能够向网络中的任何目的地发送包。中间路由器应该了解网络的连通性，以便正确地转发包，但是它们也不必了解整个网络结构，只需要了解本地部分。因此，RIP之类的协议使得路由器能够获得来自其他路由器的有关它们的连接状态的通知。设计IP的目的是使其工作在既不完全可靠也不完全冗余的网络基础设施中。以图8-2为例，这意味着如果路由器6失效，来自网络E的包仍然可以发送到网络A，不是通过路由“路由器5-路由器6-路由器1”，而是通过替代路由“路由器5-路由器2-路由器1”。

问题是，如果路由器1到网络A的链接失效，网络A和互联网络之间就不再有连通性，即它被从互联网络切断了。但是，路由器1直接收到有关其他路由器与网络A之间的链路的信

息。所有这些路由至少是2跳，对于这些信息路由器1予以忽略，因为它自身到网络A只有1跳。为保证协议有效，必须矫正此问题及类似问题，但是相应地增加了复杂度。

上述问题并不说明RIP无用，对于小型或中型互联网（又称为内联网）而言，对网络带宽的限制并不重要，因而RIP还是很重要的选路协议。但是，对于大型内联网，RIP不适用；当有很多不同路由需要考虑时，对于选路骨干业务流，RIP也不适用。

针对矢量距离方法的缺陷，诸如开放的最短路径优先（OSPF）协议之类的链路状态协议得以大量引入。采用这种方法，路由器不是周期性地向所有其他路由器通知自己的所有路由，而是只通告自己的直接链路。见图8-2，路由器1通告自己和网络A、路由器2、路由器4及路由器6有直接连接。其他的路由器也通告它们的直接连接，这样所有的路由器根据这些通告的报文就可以产生合适且理想的路由。如果路由器5得知与自己直接连接的路由器2又连接到路由器1，而路由器1直接连接到网络A，那么它就可以产生一条从网络E到网络A的路由。路由器只在连接改变或其他路由器询问时才发出通告，这样使用诸如OSPF之类的协议会减少与选路相关的噪声。由于没有RIP的跳限制，也不会在路由器间产生大量业务流，因而OSPF能够支持较大型网络。但是，OSPF比RIP要复杂得多。对多层次的支持和对基于服务类型选路的支持也是OSPF的重要特性。

选路协议也可允许更多的选路信息。例如，路由器可以根据可用带宽、延时、甚至价格对某些链路分配不同的值。这样，在图8-2中，对于要发送给网络A的业务流，如果到路由器4的链路在某方面看起来更优（例如更短、更快或价格更低），路由器5可能就倾向于通过路由器4来选路。

3. IPv6对选路协议的更新

如上所述，在为适合IPv6地址和地址范围而进行简单修改之后，BGP-4和IDRP很可能继续用作IP的外部选路协议。同样，在RFC 2080（IPv6的RIPng）中描述的RIPng与现有协议很相似。OSPF版本2最近作为Internet标准在RFC 2328中定义，而用于IPv6的OSPF版本还在定义中。预计用于IPv6的OSPF将保留OSPF的概貌，包括基本特性和功能，并使之适合于处理IPv6选路。