

的配置信息，如果将这些信息配置在计算机上，就相当于计算机拥有了公有地址，这种情况下不需要进行地址转换也可以访问互联网。其实 TCP/IP 原本的设计就是这样的。然而，如果使用路由器来上网，BAS 下发的参数就会被配置在路由器上，而且公有地址也是分配给路由器的。这样一来，计算机就没有公有地址了。

这时，计算机会被分配一个私有地址，计算机发送的包需要通过路由器进行地址转换然后再转发到互联网中。Web 和电子邮件等应用程序不会受到地址转换的影响，但有些应用程序会因为地址转换无法正常工作，这一点需要大家注意。这是因为有些应用程序需要将自己的 IP 地址告知通信对象或者告知控制服务器，但在有地址转换的情况下这些操作无法完成^①。

遇到应用程序因地址转换无法正常工作的情况时，我们可以不使用路由器，而是直接让计算机接收来自 BAS 的 PPPoE 消息^②，也就是采用最原始的上网方法。这样一来，计算机就具有了公有地址，不需要地址转换也可以上网了^③。

不过，不用路由器上网也有一点需要注意，因为上网的计算机拥有公有地址，这意味着来自互联网的包可以直接到达计算机，这可能导致计算机被攻击。因此，对于直接上网的客户端计算机，我们应该采取安装防火墙软件等防御手段。

-
- ① 网络电话、聊天、对战游戏等需要客户端之间直接收发网络包的应用程序都需要将自己的 IP 地址告知对方。这些应用程序会受到地址转换的影响，但现在已经有很多解决方案，因此不能说这些应用程序全都不能正常工作。对于某个应用程序来说，如果不知道它是否采用了相应的解决方案，就无法判断它是否会受到地址转换的影响。
 - ② 只要将计算机直接连接到 ADSL Modem、光纤收发器、ONU 等设备，或者通过集线器连接到这些设备，计算机就可以直接接收 PPPoE 消息了。
 - ③ 有一些面向公司提供的服务，如 IP8、IP16 等，可以分配多个公有地址，但这种服务非常昂贵。

4.3.7 除 PPPoE 之外的其他方式

刚才我们讲的内容都是基于 PPPoE 方式的，实际的接入网还有其他方式。下面我们先跑个题，简单介绍一下这些其他方式。

首先，我们先看看使用 PPPoA^① 方式的 ADSL 接入网^②。ADSL 使用 PPPoE 方式时，是先将 PPP 消息装入以太网包中，然后再将以太网包拆分并装入信元，而 PPPoA 方式是直接将 PPP 消息装入信元（图 4.21）。由于只是开头加不加 MAC 头部和 PPPoE 头部的区别，PPP 消息本身是没有区别的，因此密码校验、下发 TCP/IP 配置参数、收发数据包等过程都是和 PPPoE 基本相同的。不过，虽然开头加不加 MAC 头部和 PPPoE 头部看上去只是很小的区别，但却会对用户体验产生一定的影响。

PPPoA 方式不添加 MAC 头部和 PPPoE 头部，而是直接将包装入信元中。

由于 PPPoA 没有 MAC 头部，所以 PPP 消息是无法通过以太网来传输的，这就意味着需要和 BAS 收发 PPP 消息的设备，也就是计算机和路由器，必须和 ADSL Modem 是一体的，否则 PPP 机制就无法工作了。这个一体化的方式主要有以下两种。

第一种是将计算机和 ADSL Modem 用 USB 接口连接起来，这样 ADSL Modem 就和计算机成为一体了。不过，这种方式最终并没有普及。另一种方式是像图 4.21 所示的这样，将 ADSL Modem 和路由器整合成一台设备。这种方式和 PPPoE 中使用路由器上网的方式基本没什么区别，因此得到了广泛的普及。不过，正如我们刚才提到的，当由于地址转换产生问题时，这种方式就不容易处理了，因为我们无法抛开路由器用计算机直接上网。

① PPPoA: Point-to-Point Protocol over ATM。

② PPPoA 不能用于 FTTH，因为 FTTH 不使用 ATM 信元。

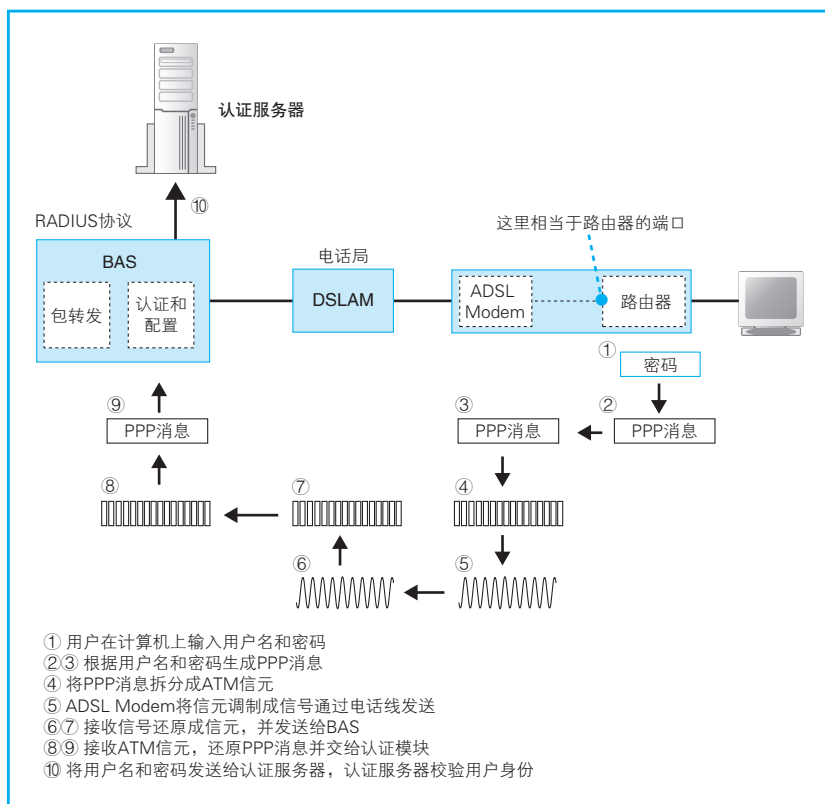


图 4.21 用 ATM 信元装载 PPP 消息的 PPPoA

当然，PPPoA 和 PPPoE 相比也有一些优势。PPPoE 方式中，如图 4.18 所示，需要添加 PPPoE 头部和 PPP 头部，这意味着 MTU 就相应变小了^①，这可能会降低网络的效率。而 PPPoA 不使用以太网包来传输 PPP 消息，因此不会发生 MTU 变小的问题。

PPPoE 会降低网络效率，PPPoA 也有 ADSL Modem 和路由器无法分离的限制，这两个问题其实都是由 PPP 引起的。因此，有一些运营商不使

① PPPoE 一般还会和隧道技术一起使用，这时还需要加上隧道头部，MTU 就更短了。

用 PPP，他们使用 DHCP^① 协议从 BAS 向用户端下发 TCP/IP 配置信息。

DHCP 经常用于通过公司网络向客户端计算机下发 TCP/IP 配置信息，其原理如图 4.22 所示，首先客户端请求配置信息（图 4.22 ①），然后 DHCP 服务器下发配置信息（图 4.22 ②），非常简单，不需要像 PPP（图 4.17）那样需要多个步骤，也不需要验证用户名和密码。没有用户名和密码，就意味着无法通过用户名来切换运营商网络，但这种方式也有优势，它可以单纯地直接传输以太网包，不需要添加额外的 PPP 头部，因此不会占用 MTU。

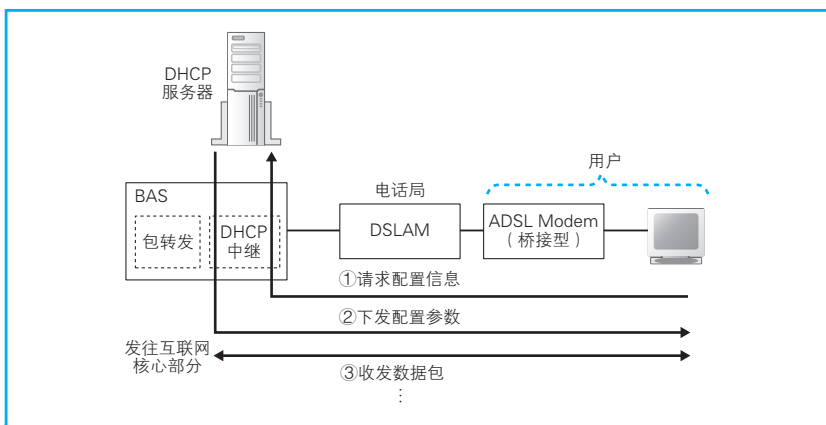


图 4.22 DHCP 的原理

此外，采用 DHCP 的运营商使用的 ADSL Modem 也和 PPPoE、PPPoA 方式不同，这种 ADSL Modem 不使用信元，而是直接将以太网包调制成为 ADSL 信号，因此没有 ADSL Modem 和路由器无法分离的问题^②。

还有一种 DHCP 方式，它不使用 PPP，而是将以太网包直接转换成 ADSL 信号发送给 DSLAM。

① DHCP: Dynamic Host Configuration Protocol, 动态主机配置协议。

② 使用信元的 PPPoE 和 PPPoA 方式中，BAS 需要配备比较昂贵的 ATM 接口，因此不使用信元还可以控制成本。

4.4 网络运营商的内部

4.4.1 POP 和 NOC

下面回到正题，现在网络包已经通过接入网，到达了网络运营商的路由器。这里是互联网的入口，网络包会从这里进入互联网内部^①。

互联网的实体并不是由一个组织运营管理的单一网络，而是由多个运营商网络相互连接组成的（图 4.23）。ADSL、FTTH 等接入网是与用户签约的运营商设备相连的，这些设备称为 POP^②，互联网的入口就位于这里。

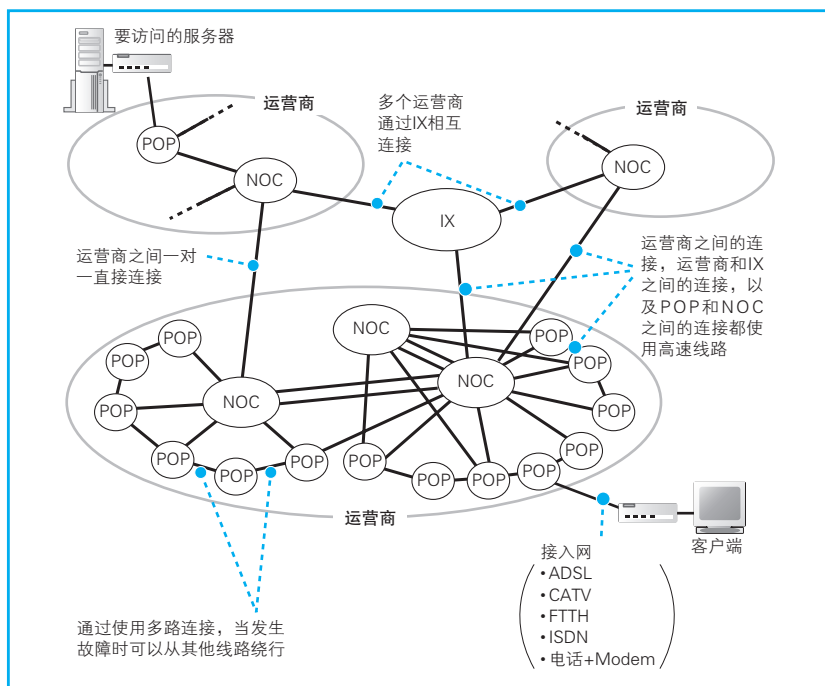



图 4.23 互联网内部概览

① 简单来说，此后网络包的传输轨迹就是通过路由器的不断转发向目的地前进，基本过程和我们之前介绍的内容大同小异。

② POP: Point of Presence, 中文一般叫作“接入点”。



网络包通过接入网之后，到达运营商 POP 的路由器。

那么，POP 里面是什么样的呢？POP 的结构根据接入网类型以及运营商的业务类型不同而不同，大体上是图 4.24 中的这个样子。POP 中包括各种类型的路由器，路由器的基本工作方式是相同的，但根据其角色分成了不同的类型。图 4.24 中，中间部分列出了连接各种接入网的路由器，这里的意思就是根据接入网的类型需要分别使用不同类型的路由器。

我们从上面开始看，首先是专线，这里用的路由器就是具有通信线路端口的一般路由器。专线不需要用户认证、配置下发等功能^①，因此用一般的路由器就可以了。接下来是电话、ISDN 等拨号方式的接入网，这里使用的路由器称为 RAS。拨号接入需要对用户拨电话的动作进行应答，而 RAS 就具备这样的功能。此外，之前我们讲过通过 PPP 协议进行身份认证和配置下发的过程，RAS 也具备这些功能。再往下是 PPPoE 方式的 ADSL 和 FTTH。PPPoE 方式中，ADSL、FTTH 接入服务商会使用 BAS，运营商的路由器则与 BAS 相连。PPPoE 中的身份认证和配置下发操作由接入服务商的 BAS 来负责，运营商的路由器只负责对包进行转发，因此这里也是使用一般的路由器就可以了。如果 ADSL 采用 PPPoA 方式接入，那么工作过程会有所不同，DSLAM 通过 ATM 交换机^②与 ADSL 的运营商的 BAS 相连，然后再连接到运营商的路由器。用户端传输的信号先经过 ADSL Modem 拆分成 ATM 信元并进行调制，然后 DSLAM 将信号还原成信元，通过 ATM 交换机转发到 BAS，最后 BAS 将信元还原成网络包，再通过运营商的路由器转发到互联网内部。

① 专线是固定连接线路，不需要进行身份认证，参数是根据传真、书面等方式下发后进行手动配置的，因此也不需要 PPP、DHCP 等机制。其实，这就是最古老的互联网接入方式。

② ATM 交换机是转发 ATM 信元的设备，负责将 DSLAM 输出的信元转发给 BAS。

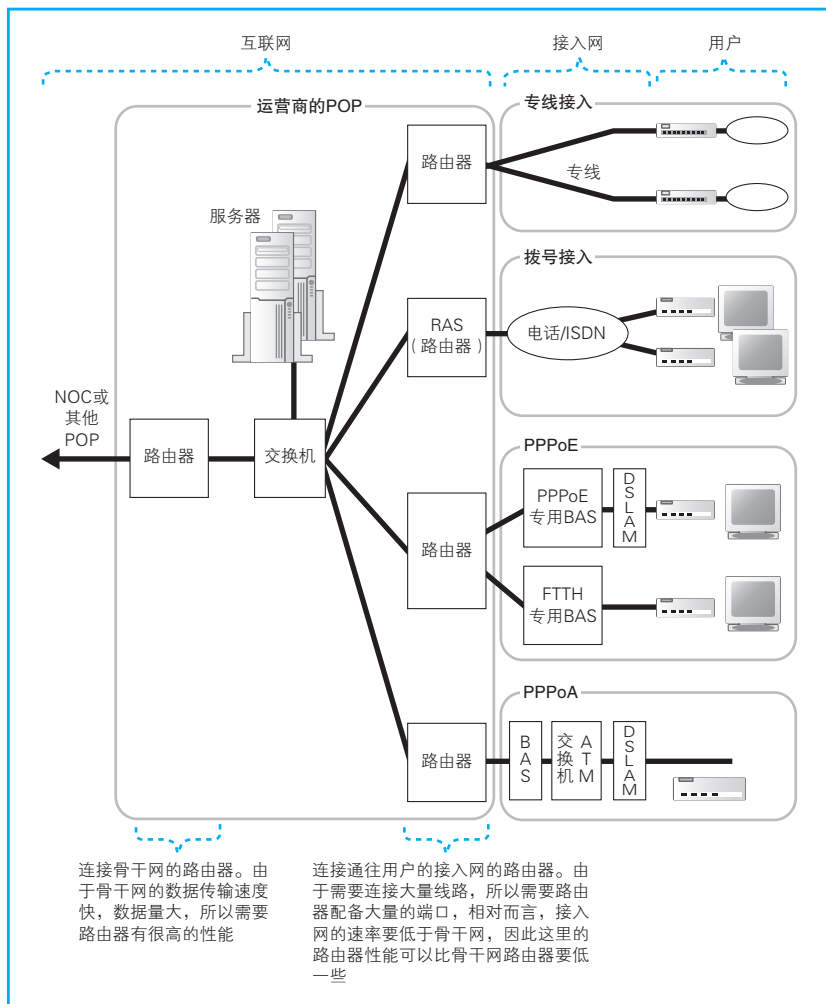


图 4.24 POP 概览

对于连接接入网的部分来说，由于要连接的线路数量很多，所以路由器需要配备大量的端口，但能传输的网络包数量相对比较少，这是因为接入网的速率比互联网核心网络要低。因此，端口多且价格便宜的路由器适用于这些场景。相对地，图中左侧的路由器用于连接运营商和核心 NOC

以及其他 POP，所有连接接入网的路由器发出的包都会集中到这里，使用的线路速率也比较高，因此这里需要配备转发性能和数据吞吐量高的路由器。

NOC^① 是运营商的核心设备，从 POP 传来的网络包都会集中到这里，并从这里被转发到离目的地更近的 POP，或者是转发到其他的运营商。这里也需要配备高性能的路由器。

话说回来，到底需要多高的性能才行呢？我们来看实际产品的参数。面向运营商的高性能路由器中有些产品的数据吞吐量超过 1 Tbit/s^②，而一般面向个人的路由器的数据吞吐量也就 100 Mbit/s 左右，两者相差 1 万多倍。当然，路由器的性能不完全是由吞吐量决定的，但从这里可以看出规模和性能的差异。

其实，NOC 和 POP 并没有非常严格的界定。NOC 里面也可以配备连接接入网的路由器，很多情况下是和 POP 共用的。从 IP 协议的传输过程来看，也没有对两者进行区分的必然性，因为无论是哪个路由器，其转发网络包的基本工作原理都是相同的。因此，大家可以简单地认为，NOC 就是规模扩大后的 POP^③。



4.4.2 室外通信线路的连接

POP 和 NOC 遍布全国各地，它们各自的规模有大有小，但看起来跟公司里的机房没什么太大区别，都是位于一幢建筑物中的，其中的路由器或者通过线路直接连接，或者通过交换机进行连接，这些和公司以及家庭网络都是相同的。只不过，公司的机房一般使用双绞线来连接设备，但运营商的网络中需要传输大量的包，已经超过了双绞线能容纳的极限，因此

① NOC: Network Operation Center，网络运行中心。

② T 表示 10^{12} 。

③ 从探索之旅的角度来看，运营商内部似乎只要有路由器就行了，但实际上 POP 和 NOC 中的设备不只有路由器。因为运营商还会提供如网站、邮件等各种服务，所以机房里面还会配备各种服务器。

一般还是更多地使用光纤^①。

大楼室内可以用线路直接连接，对于距离较远的 NOC 和 POP 来说，它们之间的连接方式可以分为几种。

对于自己拥有光纤^②的运营商来说，可以选择最简单的方式，也就是用光纤将 NOC 和 POP 直接连接起来。

这种方式虽然想法简单，但实现起来却并不简单。光纤需要在地下铺设，需要很大的工程费用，而且当线路发生中断时还必须进行维修，这些维护工作也需要费用。因此，只有有限的几家大型运营商才拥有光纤。

那么，其他运营商怎么办呢？其实也不难，只要从其他公司租借光纤就可以了，但所谓租借并不是光纤本身。

拥有光纤的公司一般都会提供光纤租用服务。以电话公司为例，电话公司会在其拥有的光纤中传输语音数据，但一条光纤并不是只能传输一条语音数据，光纤是可以复用的，一条语音数据只占其通信能力的一部分。换句话说，电话公司可以将自己的光纤的一部分通信能力租借给客户。对于客户来说，只要支付一定的费用就可以使用其中的通信能力了。对于电话公司来说，其拥有的光纤不会全部自己使用，通过租借的方式也可以带来一定的收益，无论其业务本质是电话还是互联网，这一点都是共通的^③。这种服务就叫作通信线路服务。

不拥有光纤的运营商则可以使用租借通信线路的方式将相距较远的 NOC 和 POP 连接起来。电话使用的通信线路（电话线）只能传输语音这种单一形式的数据，但运营商使用的通信线路则种类繁多。首先，在速率上就分为很多种，其中比较快的种类，其速率为电话线的 100 万倍左右。除

① 光纤基本上和 FTTH 没有区别，只不过在大楼内部短距离连接时，一般采用多模光纤。

② 比如，电话公司由于自身业务需要，通过电线杆等方式铺设了很多光纤，那么这些公司属于拥有光纤的。电力公司通过继承电线杆上架设的光纤来开展通信业务，也算是自己拥有光纤的。此外，高速公路沿途铺设的光纤也会归一些公司所有，因此拥有光纤的方式是多种多样的。

③ 互联网之外的其他通信线路服务在本质上都是一样的。

了速率之外，数据的传输方式也分为很多种。以前，将多条电话线捆绑在一起的方式比较主流，现在我们有了各种类型的通信线路，其中也有一些公司不对光纤进行细分，而是直接将整条光纤租借出去^①。不同的通信方式和速率对应着不同的价格，对于不拥有光纤的运营商来说，需要根据需要从中进行选择。

4.5

跨越运营商的网络包

4.5.1

运营商之间的连接

让我们重新回到运营商内部，看一看到达 POP 路由器之后，网络包是如何前往下一站的。首先，如果最终目的地 Web 服务器和客户端是连接在同一个运营商中的，那么 POP 路由器的路由表中应该有相应的转发目标。运营商的路由器可以和其他路由器交换路由信息，从而自动更新自己的路由表，通过这一功能，路由信息就实现了自动化管理^②。于是，路由器根据路由表中的信息判断转发目标，这个转发目标可能是 NOC，也可能是相邻的 POP，无论如何，路由器都会把包转发出去，然后下一个路由器也同样根据自己路由表中的信息继续转发。经过几次转发之后，网络包就到达了 Web 服务器所在的 POP 的路由器，然后从这里被继续转发到 Web 服务器。

那么，如果服务器的运营商和客户端的运营商不同又会怎样呢？这种情况下，网络包需要先到服务器所在的运营商，这些信息也可以在路由表中找到，这是因为运营商的路由器和其他运营商的路由器也在交换路由信息。这个信息交换的过程稍后再讲，我们暂且认为路由表中能找到对方运营商的路由信息，这时网络包会被转发到对方运营商的路由器。

总之，对于互联网内部的路由器来说，无论最终目的地是否属于同一家运营商，都可以从路由表中查到，因此只要一次接一次按照路由表中的

^① 这种服务称为 Dark Fibre，中文一般叫作“直驳光纤”。

^② 关于路由信息请参见 3.3.2 节。

目标地址来转发包，最终一定可以到达 Web 服务器所在的 POP。这样一来，我们就可以把包发到任何地方，包括地球的另一面。

4.5.2 运营商之间的路由信息交换

只要路由表中能够查到，我们当然可以把包发到任何地方，包括地球的另一面，但这些路由信息是如何写入路由表的呢？如果路由表中没有相应的路由信息，路由器就无法判断某个网络的位置，也就无法对包进行转发，也就是说，仅仅用线路将路由器连起来，是无法完成包转发的。下面我们来看看运营商之间是如何交换路由信息，并对路由器进行自动更新的。

其实方法并不难。如图 4.25 所示，只要让相连的路由器告知路由信息就可以了。只要获得了对方的路由信息，就可以知道对方路由器连接的所有网络，将这些信息写入自己的路由表中，也就可以向那些网络发送包了。

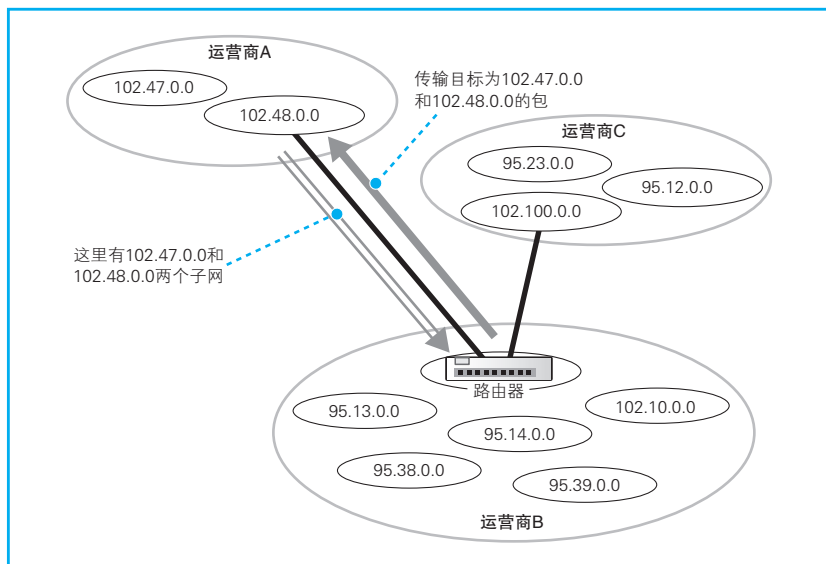


图 4.25 运营商之间的路由信息交换

获得对方的路由信息之后，我们也需要将自身的路由信息告知对方。这样一来，对方也可以将发往我们所在子网的包转发过来。这个路由信息交换的过程是由路由器自动完成的，这里使用的机制称为 BGP^①。

根据所告知的路由信息的内容，这种路由交换可分为两类。一类是将互联网中的路由全部告知对方。例如图 4.26 中，如果运营商 D 将互联网上所有路由都告知运营商 E，则运营商 E 不但可以访问运营商 D，还可以访问运营商 D 后面的运营商 B、A 和 C。然后，通过运营商 D 就可以向所有的运营商发送包。像这样，通过运营商 D 来发送网络包的方式称为转接。

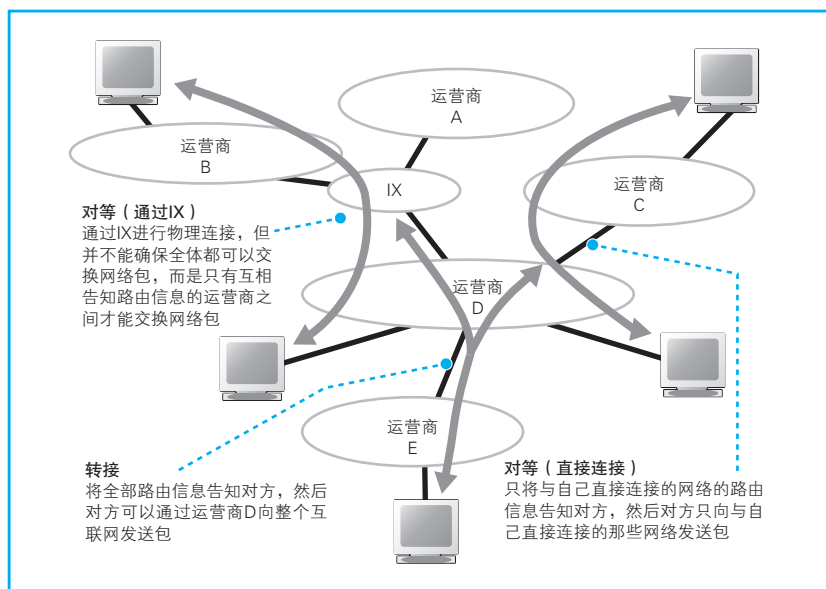


图 4.26 路由信息交换的类型

另一种类型是两个运营商之间仅将与各自网络相关的路由信息告知对方。这样，只有双方之间的网络可以互相收发网络包，这种方式称为非转

① BGP: Border Gateway Protocol, 边界网关协议。

接，也叫对等^①。



互联网内部使用 BGP 机制在运营商之间交换路由信息。



4.5.3 与公司网络中自动更新路由表机制的区别

路由器之间相互交换信息自动更新路由表的方式在公司网络中也会用到，不过公司内部和运营商之间在路由交换方式上是有区别的。

公司中使用的方式是寻找与目的地之间的最短路由，并按照最短路由来转发包，因此，周围的所有路由器都是平等对待的。

公司内部采用这样的方式没问题，但运营商之间就不行了。假设某个运营商拥有一条连接日本和美国的高速线路，那么要访问美国的地址时，可能这条线路是最短路由。如果单纯采用最短路由的方式，那么其他运营商的包就都会走这条线路，这时，该运营商需要向其他运营商收取相应的费用，否则就成义务劳动了。在这种情况下，如果使用最短路由的方式，就无法区分哪个运营商交了费，哪个运营商没交费，也就是说无法阻止那些没交费的运营商使用这条线路，这样就很难和对方进行交涉了。

正是出于这样的原因，互联网中不能单纯采用最短路由，而是需要一种能够阻止某些来源的网络包的机制，互联网的路由交换机制就具有这样的功能。

首先，互联网中可以指定路由交换的对象。公司中，路由信息是在所有路由器间平等交换的，但运营商之间的路由交换是在特定路由器间一对一进行的。这样一来，运营商就可以只将路由信息提供给那些交了费的运营商，那些没交费的运营商也就无法将网络包发送过来了。

其次，在判断路由时，该机制不仅可以判断是否是最短路由，还可以

^① 对等的英文是 peer，BGP 规格中将互相交换路由信息的节点都称为 peer，但 BGP 的 peer 实际上包含了转接和非转接两种节点，但“对等”的 peer 仅包括非转接的节点，它们的意思不同，请不要混淆。

设置其他一些判断因素。例如当某个目的地有多条路由时，可以对每条路由设置优先级。

运营商之间需要对交换路由信息的对象进行判断和筛选，但这样一来，对于没有交换路由信息的运营商网络，我们就无法将网络包发送过去了，如果要访问的 Web 服务器就在那个运营商网络中，我们不就访问不了了吗？其实不用担心，运营商在进行路由交换时会避免出现这样的情况。互联网中有很多运营商，每个运营商都和其他多个运营商相互连接。因此，如果一个运营商走不过去，可以走另一个运营商，无论网络包要发送到什么地方，都会确保能够获取相应的路由信息。如果某个运营商做不到这一点，那它也就该倒闭了。

4.5.4 IX 的必要性

图 4.26 中有一个叫作 IX^① 的东西，我们来说说它是干什么用的。对于两个运营商来说，像图 4.26 中运营商 D 和运营商 C 这样一对一的连接是最基本的一种连接方式，现在也会使用这种方式。但这种方式有个不方便的地方，如果运营商之间只能一对一连接，那么就需要像图 4.27 (a) 这样将所有的运营商都用通信线路连接起来。现在光日本国内就有数千家运营商，这样连接非常困难。对于这种情况，我们可以采用图 4.27 (b) 的方式，设置一个中心设备，通过连接到中心设备的方式来减少线路数量，这个中心设备就称为 IX。

现在日本国内有几个这样的设备，其中具有代表性的包括 JPIX^②、NSPIX-2^③、JPNAP^④。经过这 3 个 IX 的数据总量约为 200 Gbit/s^⑤，而且还

① IX: Internet eXchange，中文一般叫作“互联网交换中心”。

② JPIX: JaPan Internet eXchange，日本 Internet Exchange 公司运营的 IX。

③ NSPIX-2: Network Service Provider Internet eXchange Point-2，是由政府、学校、民间三方共同运营的 WIDE 项目的 IX。

④ JPNAP: Japan Network Access Point，日本 Internet Multifeed 公司运营的 IX。

⑤ 2007 年 2 月时的估算值。

在持续增加。

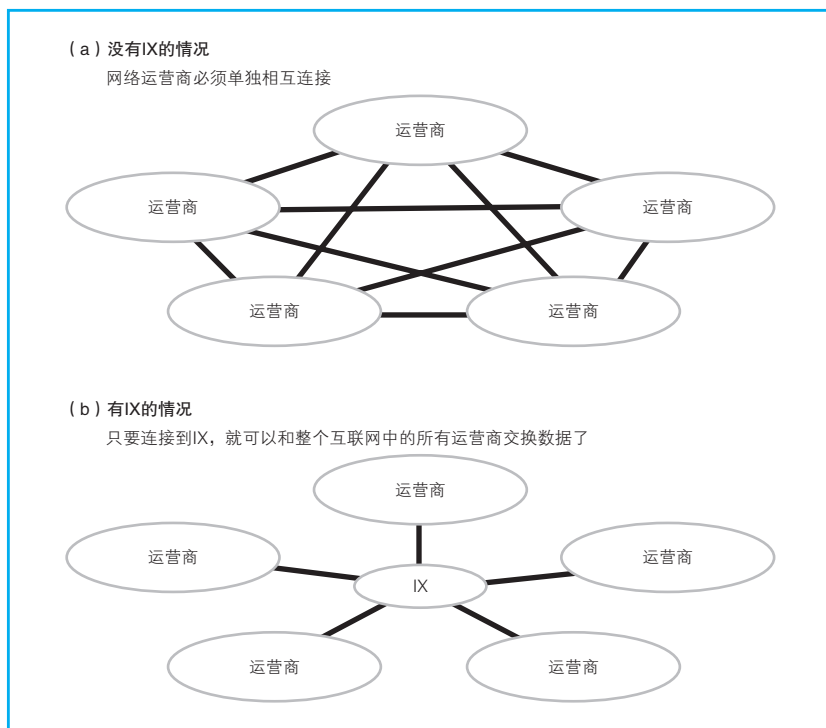


图 4.27 IX 的必要性

4.5.5 运营商如何通过 IX 互相连接

下面我们来探索一下 IX。首先是 IX 的部署场所。为了保证在遇到停电、火灾等事故，以及地震等自然灾害时，路由器等网络设备还能继续工作，IX 所在的大楼都装有自主发电设备，并具有一定的抗震能力。其实这样的要求也不仅限于 IX，运营商的 NOC 也是一样。现在在日本，拥有如此高安全性的大楼其实并不多，因此符合这样要求的大楼里面都可能会有 NOC 和 IX。运营商和 IX 运营机构会租下大楼中的一块地方用于放置 NOC 和 IX 的设备，换句话说，IX 就在这些大楼中某一层的某个角落中。