第 7 章　BGP 和因特网路由

7.1　BGP 的操作与特点

边界网关协议（BGP）是用于在自主系统（AS）之间进行路由的 IP 路由协议。在与内部网关路由协议有关的内容中，我们已遇到过 AS 的概念。从此定义可以看出，运行两个内部网关路由协议（如 OSPF 和 RIP）的网络实际上由两个 AS 组成。AS 的这个定义在企业网络界是一个有用的概念。但是，AS 概念的范围现在已拓宽，比该术语的定义更全面。AS 还被定义为受共同的控制和管理的一组网络。根据此定义，使用多个 IP 路由协议的单个企业本身就是一个 AS。

BGP 采用了 AS 的这一定义。在 AS 之间路由的协议（如 BGP）称为外部网关协议（EGP）。用于在 AS 内部路由的协议称为内部网关协议（IGP）。因特网被分成多个 AS，每个因特网服务提供商（ISP）连接到多个 AS。BGP 提供 ISP 之间的路由能力，或者更具体地说，提供 AS 之间的路由能力。BGP 有时还用于连接到业务合作伙伴的网络，这又是在 AS 之间进行路由的一个示例。互连两个合并公司的网络正日益成为除因特网之外的另一种广泛的 BGP 应用。

请记住：

- AS 是受共同的控制和管理的一组网络。
- IGP 在 AS 内部路由。
- EGP 在不同的 AS 之间路由。

BGP 概述

在讨论网络设计环境中的 BGP 之前，有必要回顾一些操作特点的基本协议：

- BGP 将 TCP 作为其传送机制。每个 BGP 数据包都封装在 TCP 中，并且由端口号 179 标识。TCP 固有的可靠性免除了在 BGP 内显式确认数据包的需要。
- BGP 依赖于路由器之间的邻居构成。这些路由器称为 BGP 的对等体或邻居。适当的 TCP 会话在两个路由器之间形成后，就形成了 BGP 的邻居。
- BGP 对等体关系有两种类型：

8　外部 BGP 或 EBGP 对等体关系在驻留于不同 AS 中的路由器之间形成。

8　内部 BGP 或 IBGP 对等体关系在驻留于同一 AS 中的 BGP 路由器之间形成。

- 定义了下列 BGP 消息类型：

8　打开

这是建立了 TCP 会话之后，每个路由器发送的第一条消息。一条打开消息包含将协商确定的保持时间和 BGP 的路由器 ID。保持时间是路由器声明 BGP 邻居已停机并将其从邻居表中去除之前，不用接收保活消息的时间长度。BGP 路由器 ID 的计算方式与 OSPF 路由器 ID 相同，即路由器上最高的活动 IP 地址（其环回接口覆盖物理接口）。

8　保活

每个路由器接收的打开消息通过保活数据包来确认。当每个路由器确认了打开消息后，就说已建立了 BGP 会话，从而可以交换 BGP 路由信息。

保活数据包不停地交换，频次高得足以避免由打开消息协商的保持时间到期。

8　更新

最初，BGP 对等体交换它们的完整路由表，从这时起，更新将在路由表发生更改后进行增量。每个 BGP 更新数据包只包含一个 AS 路径的信息。然而，所包含的信息（将在以后讨论）可以很广泛。更新中的路径信息包括 BGP 属性和通过此 AS 路径可以到达的全部网络。

这里值得强调一下，尽管 BGP 无疑是一个非常复杂的协议，但它却是一个距离矢量协议。在 BGP 更新中公布的是路由，而不是链路状态，因此在技术上使其成为了一个距离矢量协议。

8　通知

当检测到错误后会发送通知数据包。在发送这样的消息后，BGP 连接立即关闭。

BGP 维护自己的路由表，该路由表与主 IP 路由表是分开的。可以配置路由器来重新分配两个表之间的路由。在本章的后面，将对围绕主 IP 路由表和 BGP 路由表之间的路由的重新分配问题展开讨论。

基础因特网结构

使用因特网的用户都与 ISP 相连接。每个 ISP 都有自己的网络，其不同的业务提供点（POP）分布在国内或者全球。每个 POP 用于用户连接，它还可能方便与其他 ISP 的通信。ISP 之间的通信方式可能略有不同，具体取决于 ISP 的全球位置。例如在美国，存在一些注册的网络接入点（NAP），它们通常由提供 ISP 互联服务的电信公司运营。大型 ISP 有时互相之间直接连接而不经过 NAP。图 7-1 显示了一个简要的网络示意图。图上 ISP1 和 ISP2 直接连接，而 ISP3 通过它的本地 NAP 连接到因特网的其余部分。

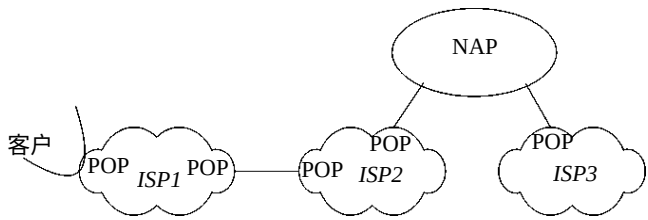

图 7-1　基础因特网结构

AS 号与 IP 地址一样是唯一的，由因特网号码分配机构（IANA）分配。小型 ISP 可能只有一个 AS 号，而较大的 ISP 有可能有多个 AS 号。

AS 指示符是一个 16 位的号码，其值范围从 1 到 65,535。它类似于专用 IP 地址的范围，从 64,512 到 65,530 的 AS 号码留作专用。很明显，如果使用 BGP 的目的是与因特网连接，那么必须使用 IANA 分配的 AS 号码。

什么时候应该使用 BGP？

当连接到因特网或另一个机构的网络时，并不总是需要使用 BGP。事实上，在实现 BGP 协议前，必须仔细地考虑和评估使用它的必要性和后果。

在下面的情况中，可能没必要或者不值得使用 BGP：

- 存在到 ISP 或业务合作伙伴的单个连接。

对于此情况，到 ISP 或业务合作伙伴的默认路由已能满足使用要求。事实上，使用 BGP 所取得的结果只不过是对网关路由器施加了巨大的处理负载和内存需求。

- 没有实现关于应该从 AS 中采取什么通信路径的策略。

即使多条路径通向 ISP 或邻近的 AS，如果次优路由是可以接受的，则也可能不需要 BGP。相反，两条默认路由可以注入本地 AS 中，并且 IGP 量度确定了要选择的路径。在图 7-2 中，由 AS65000 表示的管理网络连接到两个 ISP。使用了两条默认路由分别连接这两个 ISP。每条默认路由被重新分配为 IGP。AS65000 中的每个路由器将使用具有低 IGP 量度的重新分配的默认路由到因特网。例如，如果 OSPF 是 IGP，则 AS65000 中的每个路由器将使用“更接近”的默认路由，也就是具有低 OSPF 开销的路由。

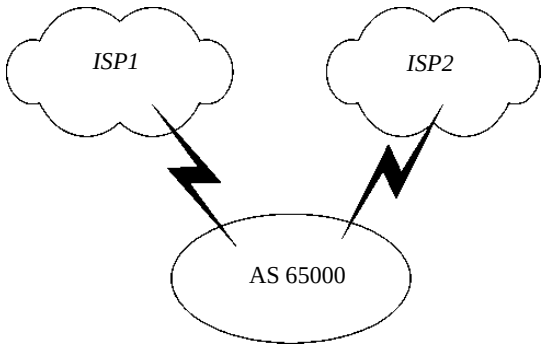

图 7-2　多个 ISP 连接

如果到同一个 AS 或 ISP 有多个链路，也可以使用这种方法。这种方法的优点是减少了因特网网关路由器上的开销、处理和内存需求。这种非 BGP 方法的缺点是有可能产生次优路由，因为路径选择仅由 IGP 执行。由于只使用默认路由，所以无法运用对特定网络的路径选择的控制。

- 如果 AS 之间的链路使用很繁重。

每个 BGP 邻居或对等体关系都要求维护 TCP 会话。增量更新也将在整个链路内传播。如果 ISP 公布所有因特网路由（或者即使是其中的大部分），则会产生巨大的路由通信量，并可能因此阻塞 WAN 链路。繁忙链路另一个潜在的不利因素是可能会丢失 BGP 通信。

这会使两个对等体之间的路由信息产生不一致，更严重，可能使 TCP 会话和 BGP 对等体被重置。

- 因特网网关路由器上的内存不足，无法存储因特网 BGP 路由表。

运行 BGP 会使用许多资源，下载因特网路由表时更是如此。即使是部分因特网路由表也会给路由器施加巨大的内存需求。维护 BGP 会话将大量占用 CPU，因此路由器必须能够满足这些需求。

这一切都提出了这个问题：何时需要 BGP？在下列情况时应该采用 BGP：

- 存在到 ISP 的多个连接。

如果存在到 ISP 的多条路径，并且要求对经过每条路径的通信量进行一定的控制。

- 使用多个 ISP，并且必须避免次优路由。

如果路由到特定的目标需要照顾某个 ISP 的利益，则必须使用 BGP。

- 要求操作路由参数以便影响路径选择。

当要为 ISP 路径选择实现任何路由策略时，通常使用 BGP。

- 路由信息在两个 AS 之间来回传送。

此方案存在于 ISP 网络中，或者存在于本地 AS 连接到不止一个 AS 的大企业的 ISP 网络中。如果本地 AS 在它所连接的其他 AS 之间分配路由信息，则必须使用 BGP。例如，参

考图 7-2，如果 ISP1 依赖于 AS 6500 来获悉本地到 ISP2 的路由和 ISP2 到本地的路由，则将使用 BGP。

EBGP 和 IBGP

外部 BGP　*外部 BGP*（EBGP）在位于不同 AS 中的两个路由器之间运行。图 7-3 显示的是路由器 R1 和 R2 之间的 EBGP 会话，这两个路由器分别位于 AS100 和 AS200 中。注意，一个路由器只能属于一个 AS，但串行链路的两端可以位于不同的 AS 中。EBGP 对等体通常必须直接连接。

内部 BGP　*内部 BGP*（IBGP）在位于同一个 AS 中的路由器之间运行。在图 7-3 中，R2 和 R3 都位于 AS200 中，因此 R2 和 R3 之间存在 IBGP 对等体关系。必须进行 IBGP 对等化，在整个 AS 中分配 BGP 路由信息，这样才能向第三个 AS 传送。

不同于 EBGP，只要每个路由器拥有一个到另一个对等体的 IGP 路由，就不需要直接路由 IBGP 对等体。

现在讨论 IBGP 的一些功能和规则。必须清楚地理解了这些功能和规则后，才能正确地设计和配置 BGP 网络。

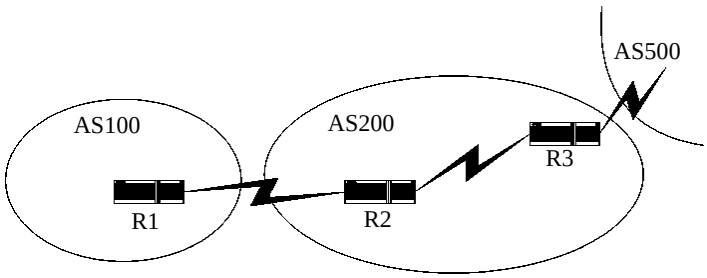

图 7-3　EBGP 和 IBGP 对等体

- IBGP 不会把它从某个 IBGP 邻居获悉的路由传播到任何其他的 IBGP 邻居。IBGP 的这一功能旨在防止路由循环。它只将此路由信息发送到 EBGP 邻居。出于这个原因，如果 IBGP 在 AS 内被广泛用于路由，则维护 IBGP 邻居的完全结网十分重要。例如，在图 7-4 中，如果 R5 是 R7 的 IBGP 邻居，则它不会把从 R7 获悉的路由信息传播到 R3。同样，R5 不会把它从 R3 获悉的任何 BGP 信息传播到 R7。除了充分结网方法外，有若干种方法可以解决由这种 IBGP 属性引起的问题。路由反射器的概念使此功能对于已配置路由反射器的客户端来说不那么严格。在本章的后面将探讨路由反射器的设计应用以及其他解决方案。
- 如果路由器从 EBGP 邻居收到有关某个特定网络的更新，它将把该路由传播到它的具有相同下一个中继段的 IBGP 邻居。换句话说，下一个中继段地址将被传入 IBGP 并得到

维护。例如，在图 7-4 中，如果 R6 将关于 166.60.0.0 的更新发送到 R3，则 R3 将用下一个地址为 167.10.1.6 的中继段接收此路由。然后它用同样的下一个 167.10.1.6 中继段将 166.60.0.0 公布给 R5。如果 R5 没有到 167.10.1.0/24 的路由，则会产生问题。

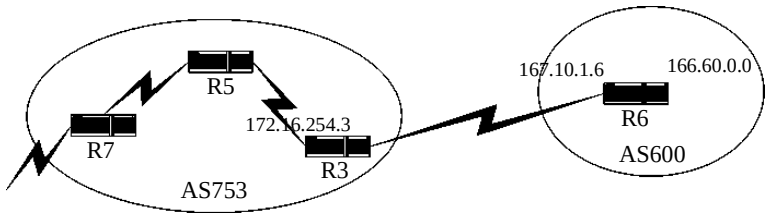

图 7-4　IBGP 路由传播

如果 R5 没有到 167.10.1.0 的路由，另一个解决方案是配置 R3 将“自己”作为下一个中继段，将它的所有路由公布给 R5。大多数路由器厂商使用 next-hop-self（下一个中继段是自己）参数来达到此效果。如果 R3 被配置为以“下一个中继段是自己”的方式将路由公布给 R5，它将把下一个中继段为 172.16.254.3 的 BGP 路由公布给 R5。

- IBGP 不会把从某个 IBGP 邻居获悉的路由公布给第三个 IBGP 邻居。
- 与通过 EBGP 获悉的路由关联的下一个中继段地址不经改变地通过 IBGP 域传播。

同步

同步　同步原则涉及 BGP 与 IGP 的同步。如果采用这种同步，则会影响 BGP 路由器对通过 IBGP 获悉的路由的处理方式。使用同步时，将强制执行下面的行为：

- BGP 路由器不会把从 IBGP 对等体获悉的路由放在它的 IP 路由表中，直到它通过 IGP 获悉了那个同样的路由。换句话说，它直到通过 IGP 获悉了 BGP 路由时才会使用它。
- BGP 路由器不会把通过 IBGP 获悉的路由公布给它的 EBGP 邻居，直到它通过 IGP 获悉有关该路由的信息。

有两种方案可能要求 BGP 和 IGP 之间同步：

- 当某个 AS 将一个自主系统的信息传送给第三个 AS 时。如果中间的 AS 将一个 BGP 路由公布给第三个 AS，而中间的 AS 还没有到达该 BGP 路由的 IGP 路由，并且第三个 AS 中的 BGP 路由器试图访问该路由，则数据包有可能丢失。如果依赖 IGP 提供到 IBGP 更新中的下一个中继段地址的路径，则可能发生这种情况。如果尚未到达该 IGP 更新，则将丢失到该目标的数据包。

例如，在图 7-4 中，如果启用了同步，则 R5 不会立即将它从 R3 收到的关于 166.60.0.0 的更新放在它的 IP 路由表中。同样，R7 路由器的 BGP 表中也有到 166.60.0.0 的路由，但它不会将其放在它的 IP 路由表中，直到它通过 IGP 获悉相同的路由。R7 也不会将 166.60.0.0 路由公布给它的 EBGP 对等体。相反，R3 会将该路由立即放在 IP 路由表中，因为它是通过

EBGP 而不是以 IBGP 获悉该路由的。

如果使用 IBGP 而不是 IGP 在 AS 内部路由，同步将导致问题。在此情况中，因为不会通过 IGP 获悉任何路由，所以这两种协议将永远不会同步，并且通过 IBGP 获悉的路由将永远不会被放在 IP 路由表中或公布给 EBGP 对等体。

- 如果没有设置 IBGP 完全结网。如果已设置完全结网，则可以禁用同步，这是因为不需要等待通过 IGP 获悉路由。笔者使用“ BGP 完全结网 ”这个术语来表示每个路由器都运行 BGP 的 AS。但这并不一定意味着每个路由器与所有其他路由器是对等的。可以使用诸如路由反射器和 BGP 联盟等解决方案来实现完全 BGP 连接，而不必完全对等化。本章稍后将探讨这些技术。

如果 BGP 没有重新分配为 IGP，则不需要同步。这是因为 BGP 和 IGP 在一定程度上相互独立，因而不需要进行同步。如果 AS 中的每个路由器都运行 BGP，则可能不必重新分配，ISP 就是一个典型的例子。

不同的厂商以自己的方式来对待同步问题。Cisco 路由器默认情况下启用该功能，如果不需要 BGP 和 IGP 之间的同步，则可以手动禁用该功能。Juniper Networks 对该功能采取了不同的方法，它在其当前产品系列中从不使用该功能。不使用该功能是因为 Juniper 设备主要针对的是 ISP 市场，并且往往是深入到 BGP 完全结网（不重新分配为 IGP）的网络内部提供服务。

请记住：

- 同步规定 BGP 路由器不会把路由公布给它的 BGP 邻居，直到它从 IGP 获悉该路由。
- 如果存在 GP 完全结网或者 BGP 不重新分配为 IGP，则不需要同步。

BGP 稳定性——问题与解决方案

路由器内存问题　2000 年的最后一个季度，因特网的路由表包含的范围是 85,000 个路由和 7,000 个 AS 号，并且仍在不断增长。仅对路由表来说，这就意味着内存需求超过 32MB。如果 BGP 路由器确实是不经过任何筛选接收整个因特网路由表，那么很明显，由于它的内存限制，有时它会经历不稳定。即使路由器是接收经过筛选的因特网路由表，这仍旧是一个问题。有鉴于此，应当经常监视运行 BGP 的路由器的内存使用。

路由处理和 CPU 使用　BGP 维护与它的每个邻居的 TCP 会话。除了维护这些会话所需要的开销之外，增量路由更新被发送到这些邻居中的每个邻居和从每个邻居接收。若 BGP 邻居的数量中等偏上，这就会大量占用 CPU。一般情况下，如果路由器的平均 CPU 使用经常超过 80-90%，CPU 正确处理数据包的能力将暂时削弱。而这会给网络运行造成严重的负面影响。与 BGP 关联的 CPU 使用随着邻居会话的数量和必须处理的增量更新程度成比例增加。

路由抑制　在任何大型网络中，特别是在使用 BGP 的网络中，速度忽高忽低的链路会

造成严重的路由破坏。持续地传播与速度忽高忽低的链路关联的 UPDATE 和 WITHDRAWN 消息，会使链路带宽并增加每个路由器的 CPU 使用，因为路由器会更新它的邻居。

使用称为路由抑制的功能可以降低由速度忽高忽低的路由引起的网络不稳定。这要求抑制公布在短时间内速度改变过于频繁的路由。配置路由抑制的确切方式因路由器厂商而异，但不论哪种情况，均采用以下原则：

- 每次链路速度的改变都会导致性能损失。
- 性能损失达到一定的阈值后，将不再公布或抑制路由。这称为抑制限制。路由即使被抑制后仍会继续导致性能损失。
- 通过半衰期的定义，路由有机会从其性能损失中“恢复”。与路由关联的性能损失值按指数衰减或减少。半衰期是指性能损失减少一半所用的时间。
- 当被抑制的路由的性能损失值减少到一个称为重用限制的预定义值时，该路由将被再次公布。

图 7-5 显示了一个性能损失与时间对比图例，以此说明了路由抑制的原理。路由抑制确保不稳定的链路不会被公布，因而避免了路由不稳定性。被抑制的路由只有在由半衰期值确定的一段时间内保持稳定后才会被重新公布。此解决方案的唯一不足之处是，在从路由恢复稳定到再次被公布这段时间内有时间滞后发生。不过，这个额外的损耗时间很可能是值得的，因为它换来了路由抑制的稳定效果。半衰期、抑制限制和重用限制这类参数是可配置的，在选择它们的值时应基于网络最近的稳定历史记录。

会话重置　有关 AS 路径的 BGP 更新包括通过该路径可到达的 IP 网络上的信息，以及与这些网络关联的路由量度和属性。如果这些属性中有任何是由管理员操作以便实现特定的路由策略，则必须重置涉及 BGP 对等体关系的 TCP 会话。

重置 TCP 会话将导致在重新建立新会话时刷新路由表和缓存项。因此，在公布和重新获悉具有新属性的路由时将中断路由。路由会消失并最终在下游路由器上重新出现，这个事实有可能导致某段时间内不稳定，因为增量路由更新可能通过几个 AS 传播。

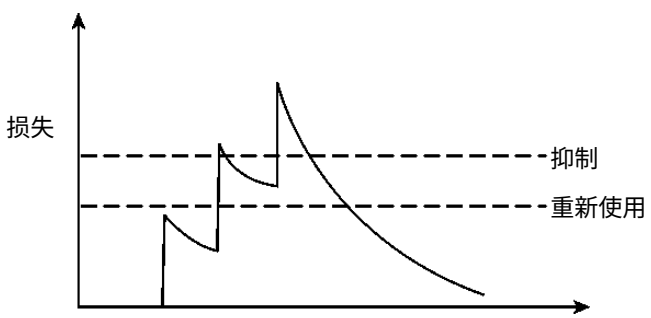

图 7-5　性能损失与时间对比显示的路由抑制

现在，大多数路由器生产商支持不必重置 TCP 会话即可重新配置 BGP 属性的能力。这通常称为 BGP 软重新配置，它消除了配置更改后进行任何不必要的路由更新。可以为出站更新或入站更新在路由器上配置软重新配置。对于出站更新的软重新配置，所有路由更新在配置发生更改后，只是被重新发送给每个邻居，不要求进行会话重置。入站更新若不重置会话则更难更改，并且本地路由器必须存储所有入站更新（与策略无关）。显然，这样会大量占用内存。通过触发会话另一端的出站软重新配置，可以强制新的策略生效。在配置任何相关功能之前，清楚路由器厂商平台上是如何处理配置更改的很重要。

IGP 重新分配为 BGP

将 IGP 重新分配为 BGP 可能导致 BGP 域中的不稳定性，原因在于 IGP 域中的链路每次更改状态时，BGP 都会生成增量更新。因此，不建议将动态 IGP 重新分配为 BGP。以下是一些替代方案：

- 从 IP 路由表发出路由。

特定 AS 的一个本地网络或网络范围可以被插入该 AS 内路由器上的 BGP 中。每个网络都必须已经驻留在该路由器的 IP 路由表中。然后，这些网络将通过 BGP 被公布。例如，通过使用网络命令，Cisco 路由器使 BGP 能够公布 IP 路由表中的 IGP 路由。这样做比将 IGP 重新分配为 BGP 更加稳定。因为主要网络或 Supernet 可以插入 BGP，所以只有此网络或网络范围的状态更改才会导致更新在整个 BGP 内传播。例如，如果整个 B 类网络 172.16.0.0 是用此方式插入 BGP 的，则只有 172.16.0.0 的所有子网都丢失才会导致 BGP 域内 172.16.0.0/16 状态的改变。路由也可以以汇总 Supernet 的形式（如 172.0.0.0/8）插入 BGP。

- 将静态路由重新分配为 BGP。

静态路由可以重新分配为 BGP 而不会引起稳定性问题，这是因为它们不改变状态。但如果目标停机，那么即使关联的静态路由仍在公布中，也会出现黑洞（即没有目标的路由）。

对因特网的移动主机访问需求相对较小，但无疑一直在增长，该需求突出了重新分配静态路由的另一个局限性。如果需要对 ISP 的直接移动访问而不是通过企业网挂接到因特网，就会导致问题的产生。只有当移动主机需要保留它们的注册 IP 地址（这可能与移动 Web 服务器的情况一样）时，这才称得上是问题。对于简单的移动拨号访问因特网，给客户机分配的地址只是来自 ISP 的注册 IP 地址池；因此，路由并不算是问题。

BGP 重新分配为 IGP

如果 BGP 不必要地重新分配为 IGP，或者 BGP 没有得到正确的控制，则在 IGP 域内还会跟着发生不稳定性。下面将分析可能需要（也可能不需要）一定程度的（从 BGP 到 IGP）重新分配的实例。

ISP 网络　一般情况下，ISP 会在每个 POP 的所有路由器上运行 BGP。通常需要 IGP 提供到 IBGP 下一个中继段地址的路由。如果 BGP 在 ISP 网络内的所有主要路由器上运行，

应该不会需要 BGP 到 IGP 的重新分配。来自其他 AS 的路由通过使用 IBGP 的 AS 运送。图 7-6 显示了 ISP 的一个完全结网 AS，它避免了将 BGP 重新分配为它自己的 IGP。

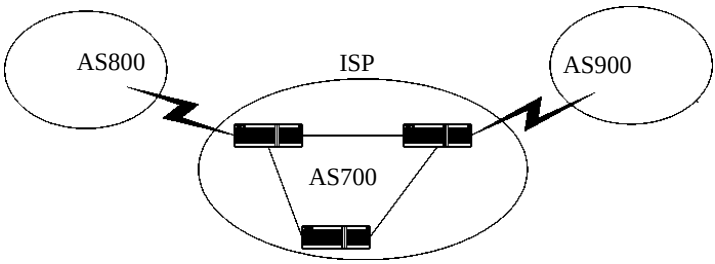

图 7-6　ISP IBGP 充分结网

不将 BGP 重新分配为 IGP 有两个主要好处。第 1 个好处是：网络内 IGP 路由器上的资源需求显著减少，这还减少了 IGP 域内的路由开销。第 2 个好处是：BGP 收敛得更快，因为 BGP 不用等待有关路由的 IGP 更新就可以公布这些路由。换句话说，在 BGP 和 IGP 之间不必采用同步。

请记住：

- 如果所有路由器在 AS 内运行 BGP，则不需要从 BGP 到 IGP 的重新分配。
- 这有两个好处：(a) 减少 IGP 域内的路由开销，(b) 因为不需要同步，BGP 收敛速度更快。

企业网　企业网一般不会在所有路由器上运行 BGP。如果可以忍受到因特网的次优路由，则默认路由可以出于此目的被重新分配为 IGP。

但是，如果必须实现涉及如何访问相邻 AS 的某些路由策略，则必须将 BGP 重新分配为 IGP。对于到业务合作伙伴网络的连接，这可能并不成问题，因为远程 AS 中的 BGP 网络数可能不是特别多。请看图 7-7 中显示的网络。所讨论的企业网是 AS5000，它通过两个链路连接到业务合作伙伴 AS6000。BGP 被用来实现一个策略：最重要的远程网络通过 T-1 链路接入，而其余的网络使用 56K 串行线路接入。出于安全原因，AS6000 只公布 AS5000 需要连接的网络。这样也具有减少 BGP 路由开销的效果，还使将这些路由重新分配为 IGP 域成为可能。因此，根据此策略，可以获得从 AS5000 内的任何路由器到 AS6000 中每个网络的优化路由。

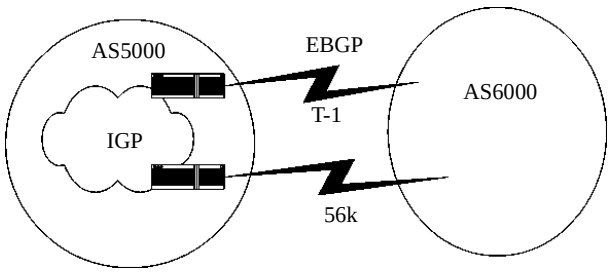

图 7-7　BGP 重新分配为 IGP

如果 AS5000 连接到 ISP，则情况就不是这么简单了。通过 EBGP 从 ISP 获悉整个因特网路由表将会使 EBGP 路由器的资源更紧张。由于路由的数量问题，将所有这些路由重新分配为 IGP，以便企业网内的每个路由器可以拥有到因特网的优化路由，这种方法肯定是不可取的。

要达到对通向本地 ISP 的路径有一定的控制，一个更现实的方法是从 ISP 接受一些大量使用的网络，而对因特网的其余部分使用默认路由。从 ISP 接受的路由可以在链路的企业端加以筛选；但是，最好安排由 ISP 自己来进行筛选，以免不必要的更新经过 ISP 的链路。

如果将 BGP 重新分配为 IGP，则由于 BGP 路由表非常大，可能需要进行筛选以减少路由开销。

如果执行 IGP 和 BGP 之间的重新分配或相互重新分配，则必须小心避免路由循环的可能。每个协议应该只重新分配从它自己的路由域内始发的路由。BGP 不应将它从 IGP 获悉的路由重新分配回 IGP。从 IGP 重新分配为 BGP 的路由也一样。这可以通过适当的路由筛选来实现。

请看图 7-8 所示的网络。路由器 R1 和 R2 与 IGP 和 BGP 域毗连。每个路由器上都配置了 BGP 和 OSPF 之间的相互重新分配。如果 OSPF 将它最初从 BGP 获悉的 150.10.0.0 路由重新分配回该协议，或者同样，如果 BGP 将 171.40.0.0 路由重新分配回 OSPF，则会发生路由循环。每个协议中都有防止这种情况发生的措施。例如，OSPF 在默认情况下不会把外部路由重新分配回另一个协议。但是，已知路由循环有潜在的负面影响，因此不应只依赖这些安全措施。应对路由筛选进行配置，使每个协议只将本地路由重新分配为其他协议。例如，不应将 172.40.0.0 BGP 路由重新分配为 OSPF，因为它们最初是 OSPF 的本地路由。

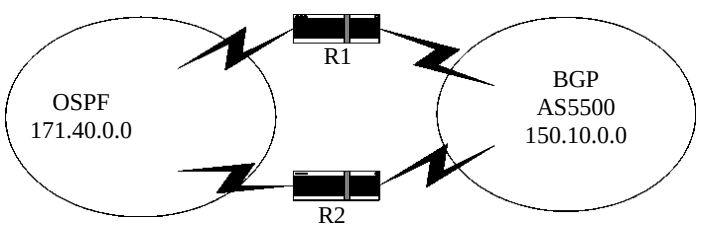

图 7-8　IGP/BGP 相互重新分配

在 BGP 和 IGP 之间进行重新分配或相互重新分配时应执行路由筛选。这样可以避免将路由公布返回获悉这些路由的协议，从而减少了路由循环的可能性。

另一个方法是在网关路由器的位置接受 BGP 路由，但仅将默认路由重新分配为 IGP 域。这个方法对路由到特定网络的 ISP 时选择的路径有一定的控制。当访问因特网网关时，IGP 域内可能会发生次优路由。在图 7-9 中，R1 和 R2 分别接入 ISP1 和 ISP2。如果这些网关路由器中的每一个都接收 BGP 路由，则这些网关路由器能够就选择哪一个链路路由到本地 ISP（即对于各种目标是选择链路 1 还是链路 2）作出基于策略的路径选择决定。这是可以执行的优化路由的范围。因为 R1 和 R2 仅将默认路由公布到 IGP 域中，因此将由 IGP 决定哪个默认路由具有较低的量度以及选择哪个 ISP。而对于只是为了使 R1 和 R2 可以控制所选的 ISP 本地链路，是否值得在 R1 和 R2 上运行 BGP 的问题，则应在设计时决定。使用到因特网中相关网络的静态路由可以取得同样的效果。

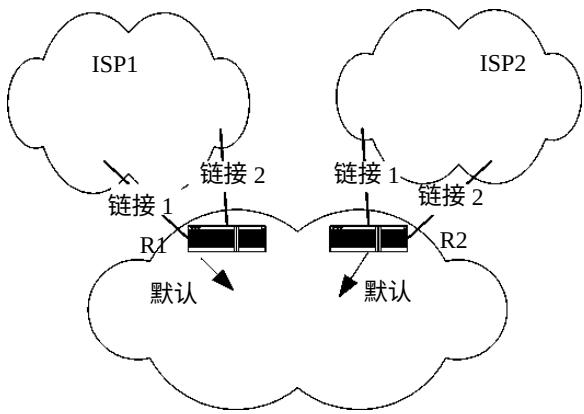

图 7-9　到 EPGP 网关的默认路由

7.2　BGP 路径选择和操作

BGP 属性

到现在为止，已多次谈到 BGP 实现“ 路由策略 ”的能力。但是，还未讨论 BGP 执行此路由功能的方式。为阐述这个问题，必须先介绍一下 BGP 属性的概念。

大多数 IGP 使用特定类型的路由量度来执行路径选择。BGP 并不使用单一的量度参数；相反，它具有多个参数，这些参数的重要程度不同，并且它们在 BGP 内的分配也有差异。这些量度参数称为路径属性。操作这些属性的能力显著提高了网络管理员影响 BGP 路径选择和实现路由策略的能力。

BGP 属性描述了 BGP 协议的主要功能，特别是在它如何执行路径选择方面。在讨论主要 BGP 属性之前，必须先说明一下这些属性可以划分成的类别。

- 公认或可选

公认属性是指在 RFC 标准中描述的并且必须被所有 BGP 识别的属性。公认属性在 BGP 邻居中间传播。

可选属性是指可以属于非标准或专有 BGP 实现的属性。所以，并不是所有 BGP 实现都有必要支持或了解它。根据属性的含义，某些受支持的可选属性被传播到 BGP 邻居。

- 强制或任意

公认属性被进一步划分为强制属性和任意属性。

强制属性必须作为 BGP 路由更新的一部分。

任意属性并不非得作为路由说明的一部分出现在更新数据包中。

- 可传递或不可传递

可选属性被划分为可传递属性和不可传递属性。

可选可传递属性是指可以由不识别该属性的 BGP 路由器透明传递的属性。

不支持或不识别可选可传递属性的路由器必须删除可选不可传递属性。

现在将就操作和功能上的重要性来说明主要的 BGP 属性。还将阐明描述每个属性时所依据的类型或类别。然后，将解释这些属性在影响路径选择方面的相对重要性。

AS 路径　当 BGP 更新通过 AS 时，该 AS 号就追加到更新中。请考虑这么一种情况：路由器接收从远处 AS 始发的更新，该更新在抵达要讨论的路由器前经过若干 AS。

该路由器将有与路由关联的路径信息，而该路由沿着每个中间 AS 行进。用一个例子可以很好地说明这一点。在图 7-10 中，每个路由器都将所显示的网络公布为 BGP。路由器 B 到达 130.1.0.0 的路径信息是（10 30），到达 180.11.0.0 的路径是（10）。同样，路由器 C 到达 120.0.0.0 的路径是（10 20），到达 180.11.0.0 的路径是（10）。路由器 A 路由到 120.0.0.0 网络和 130.1.0.0 网络的路径信息应是什么呢？

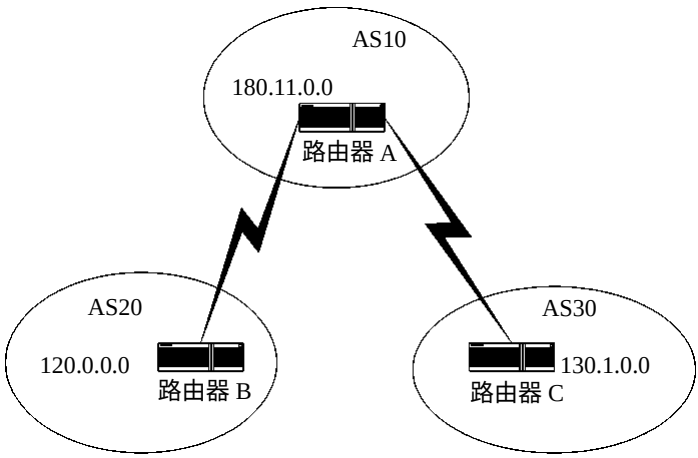

图 7-10　BGP AS 路径示例

AS 路径属性是公认强制属性。因此，该属性必须作为路由说明符的一部分包含在 BGP 更新中，并且将被所有 BGP 实现识别。

该属性对于 BGP 的操作中不产生循环是很重要的。BGP 路由器不接受将自己的 AS 包含在 AS 路径属性中的路由。这个例子的意思是路由已经过该 AS，并可能会因此产生路由循环。

为了确保 EBGP 路由不产生循环，BGP 路由器不会接受 AS 路径属性包含自己的 AS 号码的路由。

原点　原点属性还进一步分为公认属性和强制属性。它包含在所有 BGP 路由更新中，其目的是指出路径信息的原点。它可以具有下面三个值之一：

- IGP：这表明网络是从 IGP 重新分配为 BGP。
- EGP：原点就是 EGP。EGP 是 BGP 的快要被淘汰的前身。

因此，在大多数新式网络中不大可能遇到这种原点。

- INCOMPLETE：原点未知。当静态路由重新分配为 BGP 时会发生这种情况。

下一个中继段　下一个中继段属性是公认强制属性，这在 IBGP 一文中已经讨论过。与 BGP 更新关联的是下一个中继段的 IP 地址。IBGP 公布它从 EBGP 获悉的相同的下一个中继段，这个事实可能会对没有到下一个中继段的路由的下游 IBGP 邻居造成问题。下一个中继段属性的这个问题已经在图 7-4 中所示的网络例子中阐述过。

权重　权重是一个可选属性，最初被 Cisco Systems 采纳；不过，其他厂商如 Juniper Networks 的 BGP 实现也识别该属性。如果到目标网络的路由不止一个，则该属性影响来自路由器的路径选择。可以基于每个邻居来配置该属性，但这在该路由器的外部没有意义。权重的数值越高越优先。在图 7-11 中，R1 到 140.1.0.0 有两个可能的路径。它将选择通过 R3 的路径，因为与该路径关联的权重更高。权重属性值可以基于每个邻居来设置，也可以应用

到特定的 AS 路径。例如，在 R1 上，对从 R3 接收的所有更新可以将权重设置为 30，或者可以使该权重值只应用于在 AS670 中始发的路由。如何配置路由器取决于将要实现的策略。

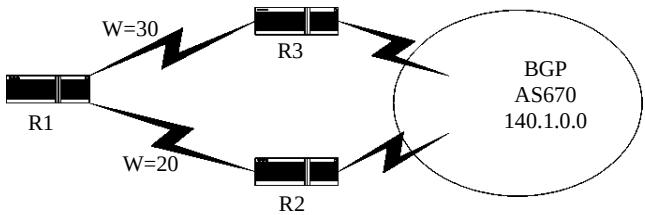

图 7-11　可选权重属性

Cisco 路由器上由本地路由器始发的路由的默认权重是 32,768，而所有其他路由的默认权重是零。权重属性从不在 BGP 邻居之间传播。

本地优先权　本地优先权属性在同一 AS 中的路由器之间分配。它旨在影响 AS 的首选出口点的选择。该属性的值越高越优先，典型默认值为 100。同往常一样，应该仔细检查厂商的文档以验证 BGP 属性的默认值。本地优先权不分配给 EBGP 邻居。本地优先权公认为任意属性。

在图 7-12 中，R2 将作为 AS5000 的首选出口点，因为它将本地优先权 200 应用于它从 AS700 中它的 EBGP 邻居处接收的所有更新。此值比 R1 的到此网络的路由的本地优先权高，所以 R1 将把其去往 AS700 的通信量发送到 R2。因为该属性在 IBGP 邻居之间传播，因此每个路由器都知道其他路由器的本地优先权值。

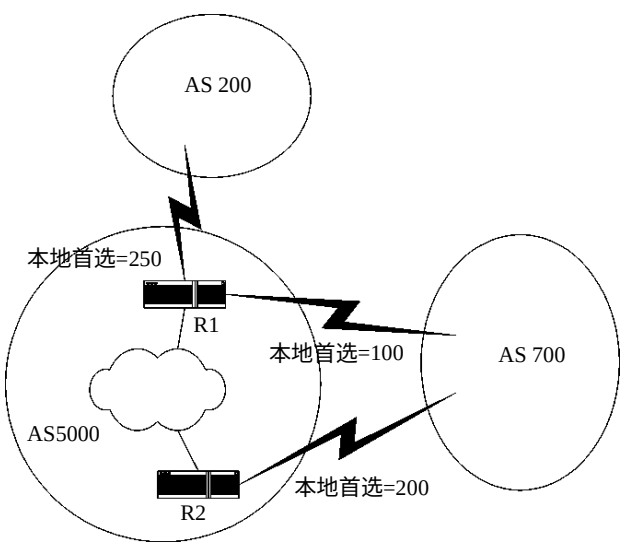

图 7-12　本地优先权属性

可以以一定程度的差别来操作本地优先权。对于在 AS 路径属性中有特定项的某些路由来说，该属性的值可以改变。例如，某些本地优先权值可以应用到从 AS200 中始发或经过 AS300 的所有路由。回到图 7-12，如果通信量去往 AS200，为了使 R1 成为 AS5000 的首选出口点，可以应用这样的策略。这是因为 R1 比 R2 更靠近 AS200。如果 AS200 是目标的话，R1 就成为出口点。所以，网络上的策略是使 R2 成为去往 AS700 的通信量的首选出口点，而 R1 则成为去往 AS200 的通信量的出口点。通过适当的路由器配置可以实现这种操作。在 Cisco 路由器上，采用了称为路由映射的功能。另一个例子是 Juniper 路由器，其相应的功能称为策略语言。

显然可以利用本地优先权属性操作来确保选择到下一个中继段的最佳路径（当然，在到达该下一个中继段后对路由选择决定不必再有任何控制）。除了优化路由选择外，本地优先权操作还可以应用于实现负载平衡。如果 AS 的两个出口点的开销相同，则对于第一条路径，可以给某些目标更高的本地优先权，而对于第二条路径，其他目标可以具有更高的本地优先权。这使网络管理员对从 AS 出去的通信量的负载平衡执行方式有更多的控制。

请记住：

- 操作本地优先权可以从特定的目标中选择不同的 AS 出口点。这可以用于提供到下一个中继段的优化路由选择。还可以用于负载平衡。

量度　量度属性（也称为多出口鉴别器或 MED）为可选不可传递属性。它被公布给 EBGP 邻居，其目的是影响从一个 AS 进入另一个 AS 的路径选择。量度值越小越优先。量度属性在 AS 内分配，以便决定接收它的 AS 中的最佳路径。量度不传播给第三个 AS；相反，它被重置为它的默认值零。这与大多数 IGP 不同，在这些 IGP 中，量度在整个路由域中传播时会随着更新增加。

在图 7-13 所示的网络中，R1 从 AS 20 中的两个源和从 AS 40 中的一个源接收有关 152.16.0.0 网络的更新。因为 R2 公布的路由的量度比 R3 低，R1 将选择 R2 作为到 152.16.0.0 的下一个中继段。