由于这四个 C 类网络已经存在于 IP 路由表中，因此工程师单独通告了每个 C 类网络。工程师使用命令 **network 192.168.24.0 mask 255.255.252.0** 汇总了这些网络，但这条命令中指定的路由 192.168.24.0/22 默认并不会被通告出去，因为 IP 路由表中没有这样的一条路由。如果 IGP 支持手动汇总（比如 EIGRP 或 OSPF），并且工程师使用 IGP 命令汇总了这些网络的话，BGP 会通告这条汇总路由。如果 IGP 没有执行路由汇总，而 BGP 又需要通告这条路由，工程师应该创建一条静态路由，把汇总网络放到 IP 路由表中。

这条静态路由应该指向空接口，使用命令 **ip route 192.168.24.0 255.255.255.0 null0**。记住，192.168.24.0/24、192.168.25.0/24、192.168.26.0/24 和 192.168.27.0/24 这些地址已经被放入 IP 路由表中。这条命令会为 192.168.24.0/22 创建一条指向空接口的额外条目。现在假设网络 192.168.25.0/24 变得不可达，那么目的地址为 192.168.25.1 的数据包要与 IP 路由表中现有的条目进行最长匹配。由于 192.168.25.0/24 已从 IP 路由表中移除，因此现在的最优路由是 192.168.24.0/22，这条路由指向空接口。数据包也会被发到空接口，同时路由器会生成 ICMP 不可达消息，并将其发送给数据包的源设备。丢弃这种数据包，可以防止这些数据包通过默认路由消耗带宽资源，这里说的默认路由可能是深入到本地 AS 内部的路由，或者（更糟糕的是）发往 ISP 的路由（这时 ISP 会根据 AS 通告过来的汇总路由把数据包重新发给 AS，从而形成路由环路）。

在这个案例中，工程师使用 **network** 命令一共通告了 5 个网络：4 个 C 类网络和 1 条汇总路由。汇总路由的目的是减少通告的数量，以及减少 Internet 路由表的内容。因此把更精确的网络信息和汇总路由一起通告出去，实际上增加了路由表的大小。

例 C-2 给出了一种效率更高的配置。工程师使用一个条目表示所有四个网络，使用去往空接口的静态路由将汇总路由放入 IP 路由表中，使 BGP 能够找到匹配条目。通过使用 **network** 命令，AS 65100 路由器为分配给这个 AS 的 4 个 C 类网络地址（192.168.24.0/24、192.168.25.0/24、192.168.26.0/24 和 192.168.27.0/24）通告了一条汇总路由。这个新的 **network** 命令（192.168.24.0/22）要想被通告出去，首先要进入本地 IP 路由表中。由于 IP 路由表中存在更为精确的网络，因此工程师创建一条指向空接口的静态路由，使路由器能够在 AS 65000 中通告这个网络（192.168.24.0/22）。

例 C-2 在图 C-2 中的路由器 C 上实施效率更高的 BGP 配置

```
router bgp 65100
  network 192.168.24.0 mask 255.255.252.0
  neighbor 172.16.2.1 remote-as 65000
ip route 192.168.24.0 255.255.252.0 null 0
```

虽然这种配置也能够正常工作，但 **network** 命令本身并不是用来执行汇总的。接下来介绍的 **aggregate-address** 命令是用于这个目的的。

C.1.4 使用 aggregate-address 命令在 BGP 表中创建汇总地址

工程师可以使用路由器配置命令 **aggregate-address** *ip-address mask* **[summary-onl y]**

[as-set]，在 BGP 表中创建聚合或汇总条目。表 C-1 中介绍了这条命令的参数。

表 C-1 **aggregate-address** 命令描述

参数	描述
ip-address	定义要创建的聚合地址
mask	定义要创建的聚合地址掩码
summary-only	（可选）让路由器只通告聚合路由。默认是既通告聚合路由，又通告明细路由
as-set	（可选）生成聚合路由的 AS-Path 信息，其中包含所有明细路由通过的所有 AS 号。聚合路由中默认只列出生成聚合路由的 AS 号

注意命令 **aggregate-address** 和命令 **network** 之间的区别：

- 命令 **aggregate-address** 只聚合已经在 BGP 表中的网络；
- BGP 中的命令 **network** 只有当 IP 路由表中存在相关路由时，BGP 才通告汇总网络。

如果工程师在配置命令 **aggregate-address** 时没有使用关键字 **as-set**，BGP 会把聚合路由通告为本地 AS 产生的路由，同时设置路由聚合（Atomic Aggregate）属性，表示这里缺失了有一些信息。BGP 会设置路由聚合属性，除非工程师使用了关键字 **as-set**。

如果工程师没有使用关键字 **summary-only** 的话，路由器仍会通告每个网络。在有冗余 ISP 链路的环境中，这样做很有用。举例来说，如果一个 ISP 只通告汇总路由，而另一个 ISP 同时通告汇总路由和明细路由，BGP 会使用明细路由。但如果通告了明细路由的 ISP 变得不可访问了，BGP 会使用另一个只通告了汇总路由的 ISP。

当工程师配置了命令 **aggregate-address** 时，路由器会自动为汇总路由在 IP 路由表中添加一条去往空接口（Null0）的路由。

如果 BGP 表中的已有路由属于命令 **aggregate-address** 指定的范围，这条汇总路由就会被放入 BGP 表中，并且会被通告给其他路由器。这个过程在 BGP 表中创建了更多信息。要想获得聚合属性提供的优势，工程师是应该使用 **summary-only** 选项来抑制汇总路由所涵盖的明细路由。当明细路由为抑制状态时，这些明细路由仍存在于执行聚合的路由器的 BGP 表中。但由于它们被标记为抑制状态，因此路由器不会把它们通告给其他路由器。

要想通过命令 **aggregate-address** 使 BGP 通告一条汇总路由，BGP 表中必须至少有一条或几条这个汇总路由涵盖的明细路由。想要让 BGP 表中有路由，通常要使用 **network** 命令来通告这些路由。

如果工程师在 **aggregate-address** 命令中只使用了关键字 **summary-only**，那么就只有汇总路由会被通告出去，并且路径信息中只会显示执行汇总的 AS 号（所有其他路径信息都缺失了）。如果工程师在 **aggregate-address** 命令中只使用了关键字 **as-set**，那么路径信息中会包含一组 AS 号（但如果之前配置了关键字 **summary-only**，将会删除这个配置）。但工程师可能需要在一条命令中同时使用这两个关键字，这样一来，只有汇总路由会被通告出去，并且路径信息中会列出所有相关的 AS 号。

图 C-3 展示了一个案例网络（与图 C-2 所示的网络相同，为了方便查看再次展示）。例 C-3 展示了路由器 C 上的相关配置，使用了命令 **aggregate-address**。

例 C-3 图 C-3 中路由器 C 的配置，使用命令 aggregate-address

```
router bgp 65100
  network 192.168.24.0
  network 192.168.25.0
  network 192.168.26.0
  network 192.168.27.0
  neighbor 172.16.2.1 remote-as 65000
  aggregate-address 192.168.24.0 255.255.252.0 summary-only
```

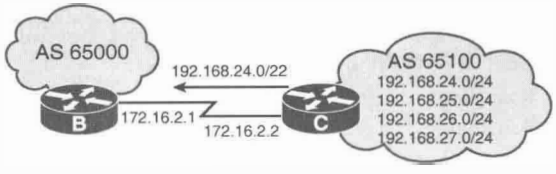


图 C-3 BGP 网络汇总案例

下面来详细说说路由器 C 上的配置。

- router bgp 65100：配置 BGP 进程，AS 号为 65100。
- network 命令：配置 BGP 在 AS 65100 中通告 4 个 C 类网络。这部分配置说明了通告什么。
- neighbor 172.16.2.1 remote-as 65000：指定使用这个地址的路由器（路由器 B）是 AS 65000 中的邻居。这部分配置说命令向哪里发送通告。
- aggregate-address 192.168.24.0 255.255.255.0 summary-only：定义需要创建的汇总路由，并且不向任何邻居通告明细路由。这部分配置说明如何通告。如果工程师没有配置 **summary-only** 选项，这条汇总路由会与明细路由一起通告出去。但是在这个案例中，路由器 B 只从路由器 C 那里收到了一条路由（192.168.24.0/22）。命令 **aggregate-address** 告诉 BGP 进程执行路由汇总，并自动在 IP 路由表中添加一条代表这个汇总路由的去往空接口的路由。

这几条 BGP 命令的主要区别在于：

- 命令 **network** 告诉 BGP 通告什么；
- 命令 **neighbor** 告诉 BGP 向哪里通告；
- 命令 **aggregate-address** 告诉 BGP 如何通告网络。

aggregate-address 命令不能代替 **network** 命令。因为汇总路由中必须至少有一条或者多条明细路由在 BGP 表中。在有些情况中，明细路由是由其他路由器注入到 BGP 表中的，聚合是由其他路由器甚至其他 AS 中的路由器执行的。这种行为称为代理聚合。在这种情况中，工程师只需要在聚合路由器配置正确的 **aggregate-address** 命令，无须 **network** 命令

就可以通告明细路由。

命令 **show ip bgp** 能够查看路由的汇总信息，它会显示出本地路由器 ID、BGP 进程获知的网络、远端网络的可达性以及 AS 路径信息。在例 C-4 中，注意看命令的输出内容，下面四个网络的第一列显示为 *s*，这表示这些网络是被抑制的；它们是通过这台路由器上的 **network** 命令学来的；下一跳地址是 0.0.0.0，表示是这台路由器在 BGP 中创建的这些条目。注意这台路由器还在 BGP 中创建了汇总路由 192.168.24. 0/22(这条路由的下一跳也是 0.0.0.0，表示是这台路由器创建了它)。明细路由都被抑制了，只有汇总路由会被通告出去。

例 C-4 show ip bgp 命令输出显示被抑制的路由

```
RouterC# show ip bgp
BGP table version is 28, local router ID is 172.16.2.1
Status codes: s = suppressed, * = valid, > = best, and i = internal
Origin codes : i = IGP, e = EGP, and ? = incomplete
Network            Next Hop      Metric LocPrf      Weight    Path
*>192.168.24.0/22  0.0.0.0             0               32768   i
s>192.168.24.0     0.0.0.0             0               32768   i
s>192.168.25.0     0.0.0.0             0               32768   i
s>192.168.26.0     0.0.0.0             0               32768   i
s>192.168.27.0     0.0.0.0             0               32768   i
```

C.2 与 IGP 之间的重分布

第 4 章中介绍了路由重分布及其配置。本节主要讨论何时适合在 BGP 和 IGP 之间执行重分布。第 7 章提到过，从图 C-4 也能看出来，运行 BGP 的路由器会维护一张包含有 BGP 信息的表，这张表与 IP 路由表相互独立。路由器会把 BGP 表中的最优路由提供给 IP 路由表，工程师也可以配置路由器共享两个表中的信息（重分布）。

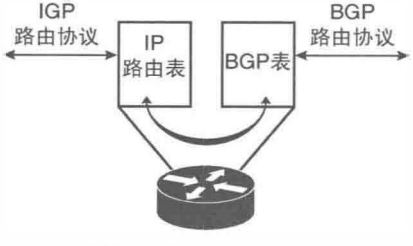


图 C-4 运行 BGP 的路由器维护 BGP 表和 IP 路由表

C.2.1 向 BGP 中通告网络

一个 AS 中的路由信息可以通过下列方式进入到 BGP 中。

- 使用 network 命令：前文提到过，**network** 命令可以让 BGP 通告已经存在于 IP 表中的网络。工程师必须使用 **network** 命令通告这个 AS 中所有希望通告的网络。

- **把去往空接口的静态路由重分布到 BGP 中**：路由器上运行不同的路由协议，当它把从一个协议接收到的路由通告到另一个协议中时，就需要使用路由重分布。静态路由这时也被当作一种协议，工程师可以通过重分布把静态信息通告到 BGP 中（“使用 network 命令进行汇总的注意事项”一节中已经介绍了空接口的用法）。
- **把动态 IGP 路由重分布到 BGP 中**：通常不建议使用这种解决方案，因为这样可能会带来不稳定性。

通常不推荐把 IGP 重分布到 BGP 中，因为 IGP 路由的任何变化（比如一条链路失效）都可能导致 BGP 更新。这种方法会使 BGP 表变得不稳定。

如果工程师使用了重分布，一定要注意只重分布了本地路由。举例来说，从其他 AS 学来的路由（从 BGP 重分布到 IGP 中的路由）一定不能从 IGP 中再次发送出去，否则将会形成路由环路；配置这种路由过滤是很复杂的。

使用 **redistribute** 命令把路由重分布到 BGP 中，会导致这个路由的源属性不完整，工程师可以从 **show ip bgp** 命令的输出内容中看到**?**。

C.2.2 从 BGP 向 IGP 中通告路由

工程师可以通过把 BGP 路由重分布到 IGP 中，使 BGP 的路由信息发送到一个 AS 内。

由于 BGP 是外部路由协议，因此在与内部协议交换路由信息时工程师需要格外谨慎，因为 BGP 表的容量非常庞大。

对于 ISP AS 来说，通常不需要从 BGP 重分布路由信息。其他 AS 可能需要使用重分布，但考虑到路由数量，工程师通常需要对路由进行过滤。接下来的小节中将逐个介绍每种情况。

ISP：不需要从 BGP 重分布到 IGP

ISP 的设备中通常有运行 BGP 的 AS 中的所有路由（或者至少有 AS 中传输路径上的所有路由）。当然了，这需要内部 BGP（iBGP）是全互连环境，并且 iBGP 要用来承载穿越 AS 的外部 BGP（eBGP）路由。在这种环境中，BGP 信息无需重分布到 IGP 中。IGP 只需要把本地信息路由到 AS 以及 BGP 路由的下一跳。

这种环境的好处之一是 IGP 协议不需要考虑所有 BGP 路由，BGP 路由交给 BGP 来处理。在这种环境中，BGP 的收敛速度也更快，因为它不需要等待 IGP 通告来的路由。

非 ISP：可能需要从 BGP 重分布到 IGP

非 ISP AS 通常不知道运行 BGP 的 AS 中的所有路由，网络中可能也没有建立全互连的 iBGP 环境。如果是这样的话，并且如果 AS 内部需要知道外部路由的话，工程师就需要把 BGP 重分布到 IGP 中。但由于 BGP 表中的路由数量过于庞大，因此工程师通常需要执行过滤。

第 6 章中“多宿主 Internet 连接”小节中介绍了另一种方法，也就是企业不用从 BGP 那里接收完整的路由，而是让 ISP 只向 AS 发送一条默认路由，或者几条默认路由以及一些外部路由。

注释 当一个 AS 只在边界路由器上运行了 BGP，AS 内的其他路由器不运行 BGP，但需要知道外部路由时，工程师就需要把 BGP 重分布到 IGP 中。

C.3 团体

第 7 章中已经讨论过，BGP 团体是一种过滤入站或出站 BGP 路由的方法。工程师很难在大型网络中，基于复杂的路由策略来使用分发列表和前缀列表过滤路由。举例来说，工程师可能需要为每台路由器上的每个邻居，都单独配置一条 **neighbor** 命令和访问列表或前缀列表。

BGP 团体功能使路由器能够使用标识符（团体）来标记路由，其他路由器能够根据这个标记做出（过滤）决策。BGP 团体用来识别共享某些相同属性的目的地（路由），它们因此而共享相同的策略。因此作为团体出现的是路由器，而不是单条路由。团体属性并不局限于一个网络或一个 AS，它并没有物理边界。

C.3.1 团体属性

团体属性是可选传递的属性。如果路由器不理解团体的概念，它也能把团体属性传递给下一台路由器。但如果路由器理解团体的概念，工程师必须明确配置它传播团体属性；否则它默认会丢弃团体属性。

每个网络都可以是多个团体的成员。

团体属性是长度为 32 比特的号码，它的取值范围是 0～4 294 967 200。前 16 比特指明了定义这个团体属性的 AS 号。后 16 比特表示团体号，只具有本地意义。工程师可以以十进制数值输入团体值，也可以使用格式 *AS:nn*，其中 *AS* 是 AS 号，*nn* 是后 16 比特本地号码。路由器默认把团体值表示为十进制数值。

C.3.2 设置并发送团体配置

工程师可以使用 route-map 来设置团体属性。

要想在 route-map 中设置 BGP 团体属性，工程师需要使用 route-map 配置命令 **set community** {[*community-number*] [*well-known-community*] [**additive**]} | **none**。表 C-2 中列出了这条命令的参数。

表 C-2 **set community** 命令描述

参数	描述
community-number	团体号，取值范围是 0～4 294 967 200
well-known-community	下面这些是预定义的公认团体 ■ **internet**：把这条路由通告到 Internet 团体，或任何属于 Internet 团体的路由器 ■ **no-export**：不向 eBGP 对等体通告这条路由 ■ **no-advertise**：不向任何对等体通告这条路由 ■ **local-AS**：不把这条路由发送到本地 AS 之外
additive	（可选）把这个团体属性添加到现有的团体中
none	为 route-map 匹配的前缀移除团体属性

工程师需要把 **set community** 命令与 **neighbor route-map** 命令一起使用，把 route-map 应用到路由更新中。

工程师需要使用路由器配置命令 **neighbor** {*ip-addres* | *peer-group-name*} **send-community**，来设置应该向 BGP 邻居发送的 BGP 团体属性。表 C-3 中列出了这条命令中的参数。

表 C-3 neighbor send-community 命令描述

参数	描述
ip-address	BGP 邻居的 IP 地址，要向它发送团体属性
peer-group-name	BGP 对等体组的名称

默认情况下，路由器并不向任何邻居发送团体属性（路由器会在出站 BGP 更新中剥除团体属性）。

图 C-5 中的路由器 C 正在向路由器 A 发送 BGP 更新，但它不希望路由器 A 将这些路由传播给路由器 B。

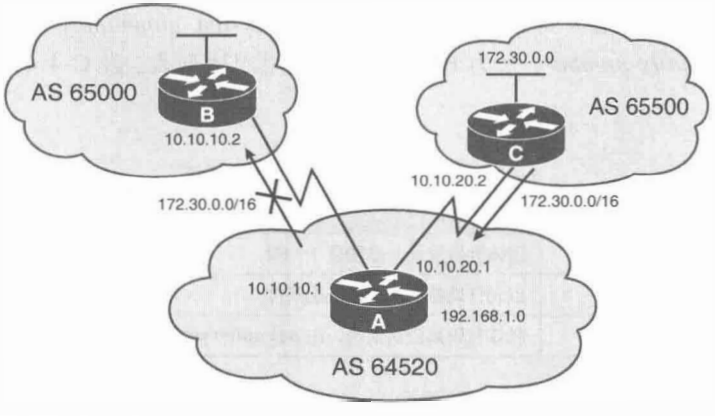


图 C-5 BGP 团体案例使用的网络

例 C-5 展示了本例中路由器 C 上的相关配置。路由器 C 在向路由器 A 通告的 BGP 路由中设置了团体属性。设置 **no-export** 团体属性是为了告诉路由器 A：不应该把这些路由发送给外部 BGP 对等体。

例 C-5 配置图 C-5 中的路由器 C

```
router bgp 65500
  network 172.30.0.0
  neighbor 10.10.20.1 remote-as 64520
  neighbor 10.10.20.1 send-community
  neighbor 10.10.20.1 route-map SETCOMM out
!
```

（待续）

```
route-map SETCOMM permit 10
  match ip address 1
  set community no-export
!
access-list 1 permit 0.0.0.0 255.255.255.255
```

在本例中，路由器 C 只有一个邻居，也就是 10.10.20.1（路由器 A）。当它与路由器 A 进行通信时，根据命令 **neighbor send-community** 的设置，它会发送团体属性。在向路由器 A 发送路由时，路由器 C 使用 route-map SETCOMM 来设置团体属性。所有匹配 **access-list 1** 的路由都会被设置上 **no-export** 团体属性。access-list 1 匹配所有路由，因此所有路由都会被设置上团体属性 **no-export**。

在本例中，路由器 A 会收到路由器 C 的所有路由，但不会把它们发送给路由器 B。

C.3.3 使用团体配置

工程师需要使用全局配置命令 **ip community-list** *community-list-number* {**permit** | **deny**} *community-number*，来为 BGP 创建并使用团体列表。表 C-4 中列出了这条命令中的参数。

表 C-4 **ip community-list 命令描述**

参数	描述
community-list-number	团体列表号码，范围是 1～99
permit | **deny**	以允许或拒绝来设置匹配条件
community-number	团体号码或公认属性，由 **set community** 命令配置

工程师可以使用 route-map 配置命令 **match community** *community-list-number* **[exact]**，把 BGP 团体属性与团体列表中的值进行匹配。表 C-5 中列出了这条命令中的参数。

表 C-5 **match community 命令描述**

参数	描述
community-list-number	团体列表号码，范围是 1～99；用来对比团体属性
exact	（可选）表明这里需要精确匹配。团体列表中的所有团体，并且只有这些团体才会出现在团体属性中

图 C-6 中的路由器 C 正在向路由器 A 发送 BGP 更新。路由器 A 根据路由器 C 设置的团体值，为这些路由设置权重。

例 C-6 展示了图 C-6 中路由器 C 的相关配置。路由器 C 只有一个邻居，就是 10.10.20.1（路由器 A）。

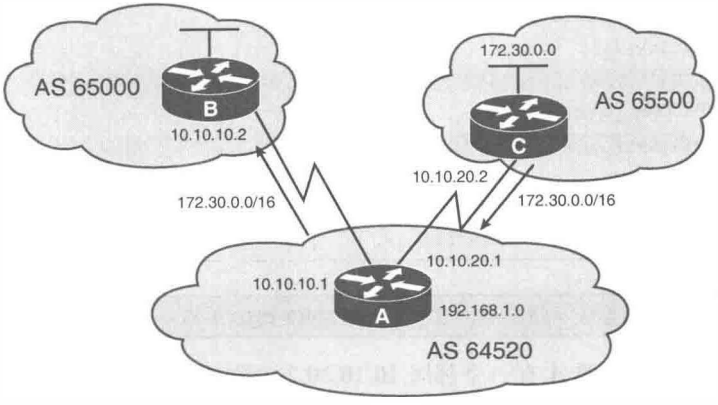


图 C-6 使用权重的 BGP 团体属性案例

例 C-6 图 C-6 中路由器 C 上的相关配置

```
router bgp 65500
  network 172.30.0.0
  neighbor 10.10.20.1 remote-as 64520
  neighbor 10.10.20.1 send-community
  neighbor 10.10.20.1 route-map SETCOMM out
!
route-map SETCOMM permit 10
  match ip address 1
  set community 100 additive
!
access-list 1 permit 0.0.0.0 255.255.255.255
```

在本例中，依照 **neighbor send-communiy** 命令的设置，路由器 A 收到了团体属性。route-map SETCOMM 用来设置向路由器 A 发送的团体属性。所有匹配 access-list 1 的路由，都会在路由现有的团体属性中添加团体 100。在本例中，access-list 1 匹配任意路由，因此所有路由的团体列表中都会添加上 100。如果工程师在 **set community** 命令中没有设置关键字 **additive**，那么这个团体属性就会代替现有的团体属性。由于本例中使用了 **additive**，因此 100 会被添加到路由所属的团体列表中。

例 C-7 展示了图 C-6 中路由器 A 上的相关配置。

例 C-7 图 C-6 中路由器 A 上的相关配置

```
router bgp 64520
  neighbor 10.10.20.2 remote-as 65500
  neighbor 10.10.20.2 route-map CHKCOMM in
!
```

（待续）

```
route-map CHKCOMM permit 10
  match community 1
  set weight 20
route-map CHKCOMM permit 20
  match community 2
!
ip community-list 1 permit 100
ip community-list 2 permit internet
```

注释 例 C-7 中没有展示路由器 A 上的其他 **router bgp** 配置。

在本例中，路由器 A 有一个邻居 10.10.20.2（路由器 C）。route-map CHKCOMM 用来在从路由器 C 收到路由时，检查团体属性。任何团体属性匹配了团体列表 1 的路由，其权重团体都会被设置为 20。团体列表 1 允许团体属性为 100 的路由。因此从路由器 C 发来的所有路由（团体列表中都有 100）都会被设置权重 20。

在本例中，所有不匹配团体列表 1 的路由都会与团体列表 2 进行对比。匹配了团体列表 2 的路由可以被放行，但其团体属性不会有任何变化。团体列表 2 中指定了关键字 **internet**，这表示所有路由。

例 C-8 中展示的案例输出内容来自于图 C-6 中的路由器 A。命令输出中展示了路由器 C 发来的路由 172. 30.0.0 的详细信息，其中包括它的团体属性 100 和添加后的权重属性 20。

*例 C-8 图 C-6 中路由器 A 上的 **show ip bgp** 命令输出*

```
RtrA # show ip bgp 172.30.0.0/16
BGP routing Table entry for 172.30.0.0/16, version 2
Paths: (1 available, best #1)
  Advertised to non peer-group peers:
    10.10.10.2
  65500
    10.10.20.2 from 10.10.20.2 (172.30.0.1)
      Origin IGP, metric 0, localpref 100, weight 20, valid, external, best, ref 2 Community:
100
```

C.4 路由反射器

BGP 规定从 iBGP 学来的路由永远不能传播给其他 iBGP 对等体（有时这种规则称为 BGP 水平分割）。这个规则导致一个 AS 中需要建立全互连的 iBGP 对等体。但如图 C-7 所示，全互连的 iBGP 环境扩展性很差。在只有 13 台路由器的 AS 中，就需要维护 78 条 iBGP 会话。随着路由器数量的增加，所需的会话数量也会增加，计算公式如下所示，其中 n 表示路由的数量：

iBGP 会话的数量 $= n(n - 1) / 2$

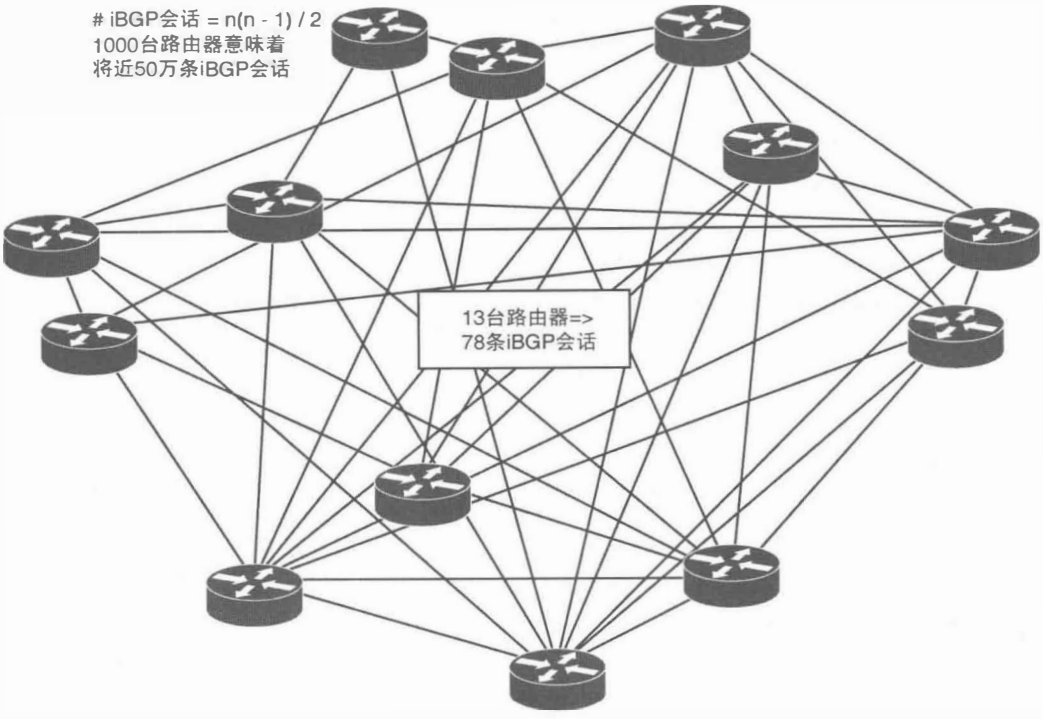


图 C-7 全互连 iBGP 需要多条会话，因此扩展性不好

除了必须要创建并维护的 BGP TCP 会话数量外，路由流量的总量也可能是个问题。根据 AS 的拓扑，当流量穿越每个 iBGP 对等体时，可能会在一些链路上复制多次。比如，如果一个大型 AS 的物理拓扑中包含一些 WAN 链路，那么运行在这些链路上的 iBGP 会话就会消耗大量的带宽。

对于上述问题的解决方案就是使用路由反射器（RR）。本节将介绍 RR 的工作原理和配置方式。

RR 修改了 BGP 的规则：允许配置为 RR 的路由器将从 iBGP 学到的路由传播给其他 iBGP 对等体，详见图 C-8。

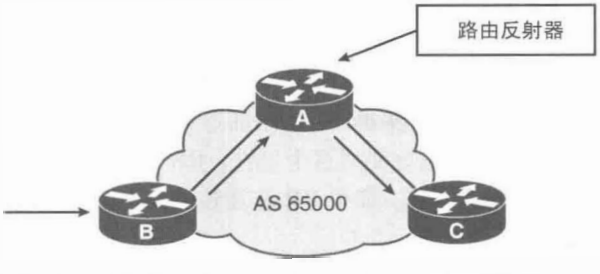


图 C-8 路由器 A 为 RR，它可以把 iBGP 路由从路由器 B 传播给 C

这种做法减少了必须维护的 BGP TCP 会话数量，也减少了 BGP 路由流量。

C.4.1 路由反射器的优势

当配置了 BGP RR 后，就不再需要使用全互连的 iBGP 对等体拓扑了。RR 能够把从 iBGP 学到的路由传播给其他 iBGP 对等体。RR 主要是 ISP 在用，尤其是当 ISP 内部 **neighbor** 命令过多时使用。路由反射器通过使用一个中心路由器来为 RR 客户端复制路由更新，而减少了一个 AS 内 BGP 邻居关系的数量（因此也节省了 TCP 连接）。

路由反射器并不会影响 IP 数据包传输的路径，它只会影响分布路由信息所使用的路径。不过如果 RR 的配置不正确，网路中有可能形成环路，后文“路由反射器迁移提示”小节中给出了案例。

一个 AS 内可以有多个 RR，既可以提供冗余性，又可以以分组的形式减少未来所需的 iBGP 会话数量。

迁移到 RR 只需要很少的配置，并且无需一次性完成配置，因为一个 AS 内可以同时有非 RR 路由器和 RR 路由器。

C.4.2 路由反射器的术语

路由反射器（Route Reflector）是一台拥有特殊配置的路由器，它能够把从 iBGP 学到的路由通告给（或者说发射给）其他 iBGP 对等体。RR 和其他路由器之间建立部分 iBGP 对等体关系，这些路由器称为客户端（Client）。客户端之间无需建立对等体关系，因为路由反射器会在客户之间通告路由信息。

RR 及其客户端的组合称为一个集群（Cluster）。

其他不是 RR 客户的 iBGP 对等体称为非客户端（Nonclient）。

起源 ID（Originator ID）是可选项，是由 RR 创建的非传递 BGP 属性。这个属性中记录了本地 AS 中这条路由的初始路由器 ID。如果由于配置问题产生了环路，这个路由更新再次回到了源路由器，源路由器会忽略它。

通常一个集群中只有一个 RR，这时使用 RR 的路由器 ID 来标识这个集群。为了提高冗余性，并且避免单点故障，一个集群中也可以有多个 RR。当使用多个 RR 时，工程师需要为集群中的所有 RR 都配置集群 ID（ClusterID）。路由反射器可以通过集群 ID，来识别同一集群中其他 RR 发来的更新。

集群列表（ClusterList）是路由途经的集群 ID 序列。当 RR 将路由从它的客户端反射给集群外的非客户端时，它会在集群 ID 上添加本地集群 ID。如果这个更新的集群列表是空的，RR 会自己创建一个。通过使用这个属性，RR 可以在配置有缺陷的环境中，发现由于环路回到相同集群的路由信息。如果 RR 在通告的集群列表中看到了本地集群 ID，它会忽略这个通告。

起源 ID、集群 ID 和集群列表都有助于在 RR 配置中预防路由环路。

C.4.3 路由反射器的设计

在一个 AS 中使用 RR 时，工程师可以把这个 AS 分割为多个集群，每个集群中至少有一个 RR 和少量客户端。出于冗余性的考量，工程师也可以在一个集群中设置多个 RR。

RR 之间必须建立全互连的 iBGP 关系，确保所有学到的路由能够在 AS 内正确传播。

工程师仍需使用 IGP，IGP 需要承载本地路由和下一跳地址。

RR 及其客户端之间也遵守普通的水平分割原则。因此 RR 从一个客户端收到的路由，不会再通告给这个客户端。

> **注释** 对于一个 RR 可以有多少个客户端并没有官方限制，这完全取决于路由器的内存容量。

C.4.4 路由反射器的设计案例

图 C-9 提供了一个 BGP RR 设计案例。

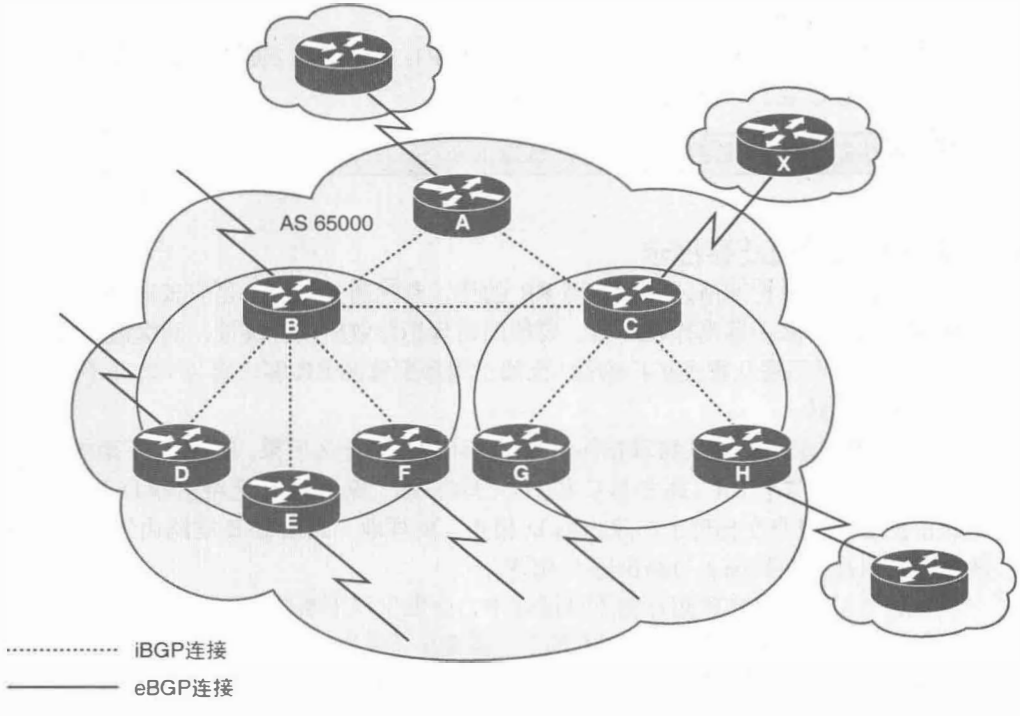


图 C-9 路由反射器设计案例

> **注释** 图 C-9 中并没有展示 AS 65000 中的物理连接。

在图 C-9 中，路由器 B、D、E 和 F 组成了一个集群。路由器 C、G 和 H 组成了另一

个集群。路由器 B 和 C 是 RR。路由器 A、B 和 C 之间建立了全互连的 iBGP 关系。

C.4.5 路由反射器的工作原理

当 RR 收到一个更新时，它会根据发送这个更新的对等体类型，采取以下对策。

- 如果是从客户端对等体收到的更新，它会把更新发送给所有非客户端对等体以及所有客户端对等体（除了路由的始发设备）。
- 如果是从非客户端对等体收到的更新，它会把更新发送给集群内的所有客户端。
- 如果是从 eBGP 对等体收到的更新，它会把更新发送给所有非客户端对等体以及所有客户端对等体。

以图 C-9 为例，在这个环境中会发生以下事件。

- 如果路由器 C 从路由器 H（客户端）那里收到了更新，它会把更新发送给路由器 G，以及路由器 A 和 B。
- 如果路由器 C 从路由器 A（非客户端）那里收到了更新，它会把更新发送给路由器 G 和 H。
- 如果路由器 C 从路由器 X（通过 eBGP）那里收到了更新，它会把更新发送给路由器 G 和 H，以及路由器 A 和 B。

> 注释 路由器也会向相应的 eBGP 邻居发送更新。

C.4.6 路由反射器的迁移提示

当工程师想要把网络迁移为使用 RR 的话，首先需要考虑的是应该把哪些路由器当作反射器，把哪些路由器当作客户端。要按照物理拓扑做出设计决策，确保数据包转发路径不受影响。如果不遵从物理拓扑的话（比如工程师配置的 RR 客户端与 RR 并不物理相连），可能会造成环路。

图 C-10 展示了不遵从物理拓扑的 RR 设计会带来什么后果。在图中，下面的路由器（路由器 E）同时是两个 RR（路由器 C 和 D）的客户端。路由器 A 是路由器 D 的 RR 客户端，但路由器 A 并没有在物理上与路由器 D 相连。同样地，路由器 B 是路由器 C 的 RR 客户端，但它也没有在物理上与路由器 C 相连。

在这个没有遵从物理拓扑的不良设计中，会发生以下事件。

- 路由器 B 知道去往 10.0.0.0 的下一跳是 x（它从 RR 路由器 C 学到这个下一跳）。
- 路由器 A 知道去往 10.0.0.0 的下一跳是 y（它从 RR 路由器 D 学到这个下一跳）。
- 路由器 B 去往 x 的最优路由可能要穿越路由器 A，因此路由器 B 向路由器 A 发送目的地为 10.0.0.0 的数据包。
- 路由器 A 去往 y 的最优路由可能要穿越路由器 B，因此路由器 A 向路由器 B 发送目的地为 10.0.0.0 的数据包。
- 环路形成了。

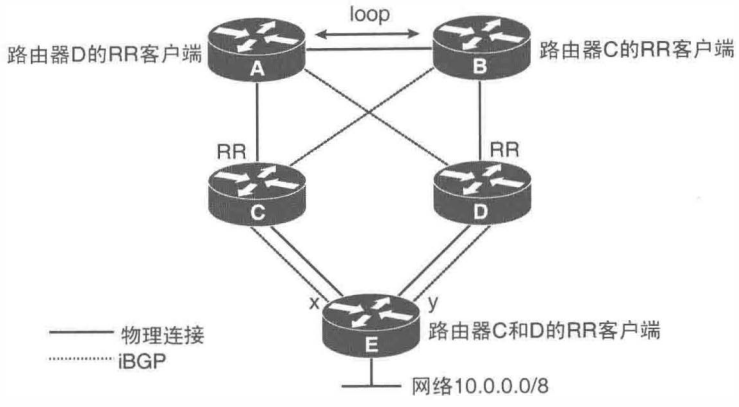


图 C-10 没有遵从物理拓扑的不良 RR 设计

图 C-11 展示了一个相对较好的设计（相对较好是因为它遵从了物理拓扑）。在这张图中，下面的路由器（路由器 E）仍是两台 RR 的客户端。

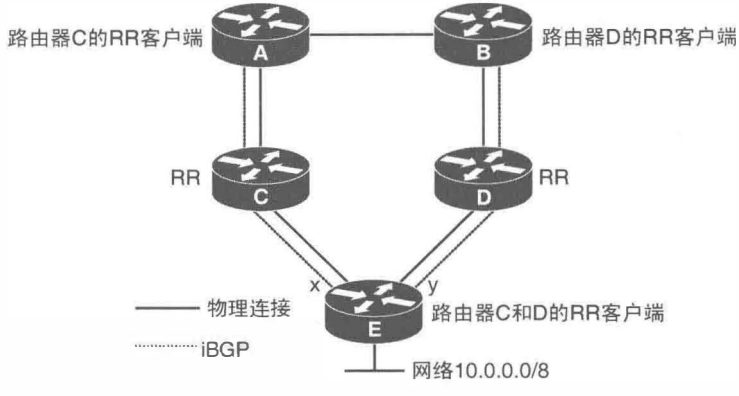


图 C-11 遵从物理拓扑的较好 RR 设计

在这个遵从了物理拓扑的较好设计中，会发生以下事件。

- 路由器 B 知道去往 10.0.0.0的下一跳是 *y*（它从 RR 路由器 D 学到这个下一跳）。
- 路由器 A 知道去往 10.0.0.0的下一跳是 *x*（它从 RR 路由器 C 学到这个下一跳）。
- 路由器 A 去往 *x* 的最优路由要穿越路由器 C，因此路由器 A 向路由器 C 发送目的地为 10.0.0.0 的数据包，路由器 C 将数据包转发给路由器 E。
- 路由器 B 去往 *y* 的最优路由要穿越路由器 D，因此路由器 B 向路由器 D 发送目的地为 10.0.0.0 的数据包，路由器 D 将数据包转发给路由器 E。
- 没有环路。

当工程师把网络迁移为使用 RR 时，要一次配置一个 RR，然后删除客户端之间重复的

iBGP 会话。建议工程师为每个集群就配置一个 RR。

C.4.7 路由反射器的配置

工程师可以使用路由器配置命令 **neigh bor** *ip-address* **rout e-reflector-clinte** 把一台路由器配置为 BGP RR，并把指定邻居配置为它的客户端。*ip-address* 参数设置的是要成为客户端的 BGP 邻居地址。

配置集群 ID

如果工程师配置了多个 BGP 集群，并且需要配置集群 ID 时，可以在集群中的所有 RR 上使用路由器配置命令 **bgp cluster-id** *cluster-id*。注意，在配置了 RR 客户端后就无法更改集群 ID 了。

使用 RR 后，会对一些命令带来限制，其中包括以下命令。

- 当工程师在 RR 上使用命令 **neigh bor next-hop-self** 时，只会影响到从 eBGP 学来的路由的下一跳，因为反射 iBGP 路由的下一跳不应该发生变化。
- RR 客户端的配置无法与对等体组共存。这是因为对等体组中的路由器必须向这个对等体组中的所有成员发送更新。如果 RR 的所有客户端都在一个对等体组中，并且其中一个客户端发送了更新，那么 RR 要负责为所有其他客户端共享这个更新。由于水平分割原则，RR 绝不会把更新发送给源客户端。

C.4.8 路由反射器的案例

图 C-12 展示了一个案例网络，其中路由器 A 是 AS 65000 中的 RR。例 C-9 展示了路由器 A（RR）上的相关配置。

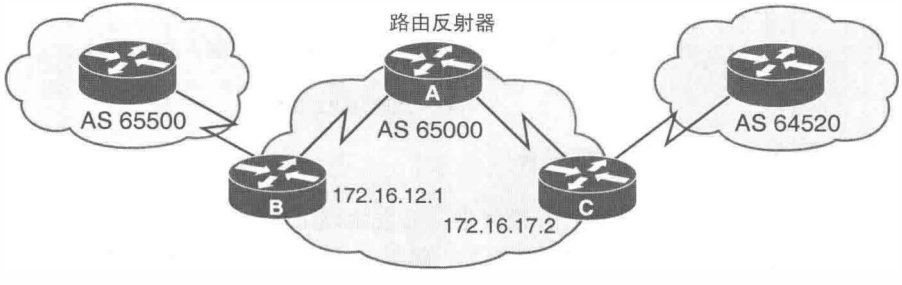


图 C-12 路由器 A 是 RR

例 C-9 配置图 C-12 中的路由器 A

```
RTRA(config)# router bgp 65000
RTRA(config-router)# neighbor 172.16.12.1 remote-as 65000
RTRA(config-router)# neighbor 172.16.12.1 route-reflector-client
RTRA(config-router)# neighbor 172.16.17.2 remote-as 65000
RTRA(config-router)# neighbor 172.16.17.2 route-reflector-client
```

工程师使用 **neighbor route-reflector-client** 命令指定了哪个邻居是 RR 客户端。在本例中，路由器 B 和 C 都是路由器 A 的 RR 客户端。

C.4.9 检查路由反射器

工程师可以使用命令 **show ip bgp neighbors** 来查看哪个邻居是 RR 客户端。例 C-10 展示了这条命令的部分输出内容，以图 C-12 中的路由器 A 为例，显示出邻居 172.16.12.1（路由器 B）是路由器 A 的 RR 客户端。

例 C-10 在图 C-12 中路由器 A 上查看 show ip bgp neighbors

```
RTRA# show ip bgp neighbors
BGP neighbor is 172.16.12.1, remote AS 65000, internal link
 Index 1, Offset 0, Mask 0x2
  Route-Reflector Client
  BGP version 4, remote router ID 192.168.101.101
  BGP state = Established, table version = 1, up for 00:05:42
  Last read 00:00:42, hold time is 180, keepalive interval is 60 seconds
  Minimum time between advertisement runs is 5 seconds
  Received 14 messages, 0 notifications, 0 in queue
  Sent 12 messages, 0 notifications, 0 in queue
  Prefix advertised 0, suppressed 0, withdrawn 0
  Connections established 2; dropped 1
  Last reset 00:05:44, because of User reset
  1 accepted prefixes consume 32 bytes
  0 history paths consume 0 bytes
--More--
```

C.5 通告默认路由

工程师可以使用路由器配置命令 **neighbor** {*ip-address* | *peer-group-name*} **default-originate** [**route-map** *map-name*]配置 BGP 路由器，让它向某个邻居发送默认路由 0.0.0.0，邻居会把这条路由当作默认路由使用。表 C-6 介绍了这条命令的参数。

表 C-6 **neighbor default-originate 命令描述**

参数	描述
ip-address	BGP 邻居的 IP 地址
peer-group-name	BGP 对等体组的名称
route-map *map-name*	（可选）指定 route-map，将默认路由按照工程师的规划发送给邻居

C.6 不通告私有 AS 号

第 6 章中提到过，IANA 定义了私有 AS 号范围 64512～65534，用于私有目的，这跟

私有 IPv4 地址的概念类似。

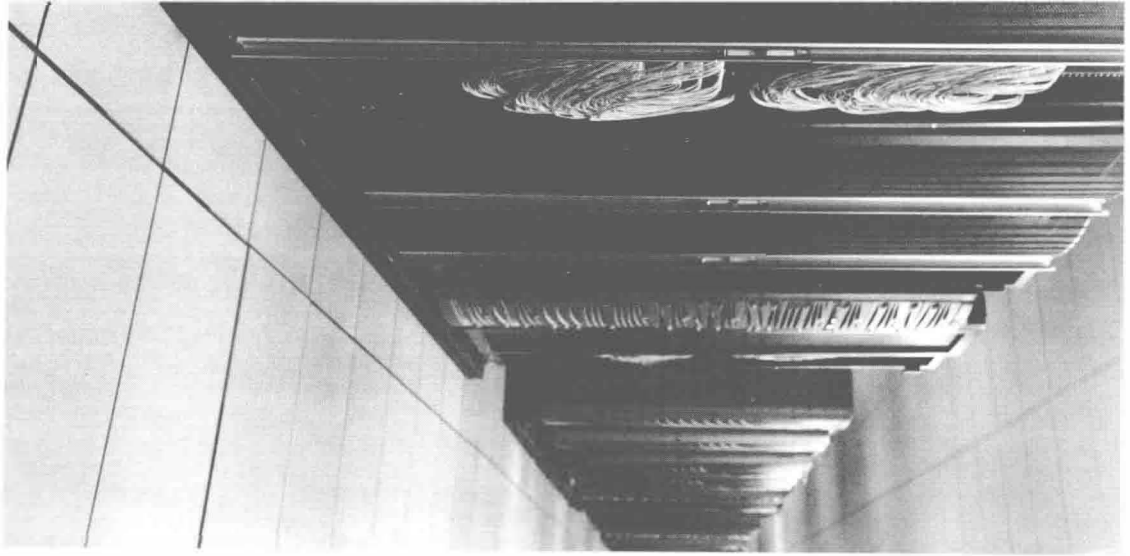
只有公有 AS 号应该通过 eBGP 邻居发送到 Internet 上。工程师可以使用路由器配置命令 **neighbor** {*ip-address* | *peer-group-name*} **remove-private-as** [**all** [**replace-as**]]，从 AS-Path 属性中移除私有 AS 号；工程师只能针对 eBGP 邻居配置这条命令。

表 C-7 介绍了这条命令的参数。

表 C-7 **neighbor remove-private-as** 命令描述

参数	描述
ip-address	BGP 邻居的 IP 地址
peer-group-name	BGP 对等体组的名称
all	（可选）从出站更新的 AS-Path 中移除所有私有 AS 号
replace-as	（可选）只有当配置了关键字 **all** 时才有效，关键字 **replace-as** 可以把 AS-Path 中的所有私有 AS 号都替换成路由器本地 AS 号。这样做可以维持 AS-Path 属性的长度（用于 BGP 路径选择过程中），使之与替换 AS 号之前一样长

注释 从命令的语法中可以看出，关键字 **all** 可以单独使用；通过测试得出，只使用关键字 **all** 和不使用关键字 **all** 得出的结果相同。



缩写与简称

本附录旨在介绍本书和互联网行业涉及的缩写、简称和首字母缩写。

缩写	全称
3DES	三重 DES
6-to-4	IPv6 到 IPv4
AAA	认证、授权、审计
ABR	区域边界路由器
ACK	1．确认 2．确认包 3．TCP 分段中的确认位
ACL	访问控制列表
AD	通告距离
AES	高级加密标准
AfriNIC	非洲网络信息中心
AH	认证头部
APNIC	亚太互联网络信息中心
ARIN	美洲互联网号码注册管理机构
ARP	地址解析系统
AS	自治系统
ASBR	自治系统边界路由器
ASN	AS 号
BDR	备份指定路由器
BGP	边界网关协议
BGPv4 或 BGP-4	BGP 版本 4
bps	每秒位
BSCI	构建可扩展的 Cisco 互联网络
CCDP	Cisco 认证资深网络设计工程师
CCNA	Cisco 认证网络工程师
CCNP	Cisco 认证网络资深工程师
CCSP	Cisco 认证网络安全资深工程师
CDP	Cisco 发现协议
CE	客户边缘
CEF	Cisco 快速转发

续表

缩写	全称
CEFv6	IPv6 Cisco 快速转发
CIDR	无类域间路由
CoS	服务类型
CPE	客户提供商边缘 客户边缘设备
CPU	中央处理单元
DAD	地址冲突检测
DBD	数据库描述数据包
DES	数据加密标准
DESGN	设计 Cisco 互联网络解决方案
DHCP	动态主机配置协议
DHCPv6	IPv6 DHCP
DHCPv6-PD	DHCPv6 前缀代理
DLCI	数据链路连接标识符
DMVPN	动态多点 VPN
DNA	永不老化
DNS	域名服务或域名系统
DR	指定路由器
DUAL	弥散更新算法
E1	外部类型 1
E2	外部类型 2
eBGP	外部 BGP
EGP	外部网关协议
EIGRP	增强型内部网关路由协议
ESP	封装安全负载
EUI-64	64 位扩展唯一标识
FD	可行距离
FHRP	第一跳冗余协议
FIB	转发信息库
FLSM	固定长度子网掩码
FS	可行后继
FTP	文件传输协议
Gbps	每秒吉比特
GE	吉比特以太网
GLBP	网关负载分担协议
GRE	通用路由加密
HMAC	散列消息认证码
HSRP	热备份路由器协议

续表

缩写	全称
HTTP	超文本传输协议
HTTPS	安全 HTTP
Hz	赫兹
IANA	互联网号码分配局
iBGP	内部 BGP
ICANN	互联网名称与数字地址分配机构
ICMP	互联网控制消息协议
ICMPv4	IPv4 ICMP
ICMPv6	IPv6 ICMP
ID	标识符
IDRP	域间路由协议
IEEE	电子电子工程师学会
IETF	互联网工程任务组
IGMP	互联网组管理协议
IGP	内部网关协议
IGRP	内部网关路由协议
IKE	互联网密钥交换
INARP	逆向地址解析协议
IND	逆向邻居发现
IOS	互联网操作系统
IP	互联网协议
IPSec	IP 安全
IPv4	IP 版本 4
IPv6	IP 版本 6
IPX	互联网络数据包交换
IS	1. 信息系统 2. 中间系统
IS-IS	中间系统到中间系统
IS-ISv6	IPv6 IS-IS
ISP	互联网服务提供商
ISR	集成服务路由器
ITU-T	国际电信联盟电信标准化部门
Kbps	每秒千比特
LACNIC	拉丁美洲及加勒比地区互联网地址注册管理机构
LAN	局域网
LS	链路状态
LSA	链路状态通告
LSAck	链路状态确认

续表

缩写	全称
LSDB	链路状态数据库
LSR	链路状态请求
LSU	链路状态更新
M	度量
MAC	1. 媒体访问控制
	2. 消息认证码
MB	兆比特
Mbps	每秒兆比特
MD5	消息摘要算法 5
MED	多出口鉴别器
MIB	管理信息库
MOTD	当日消息
MP-BGP	多协议 BGP-4
MP-BGP4	多协议边界网关协议版本 4
MPLS	多协议标签交换
ms	毫秒
MTU	最大传输单元
NA	邻居通告
NAT	网络地址转换
NAT64	NAT IPv6 到 IPv4
NAT-PT	NAT-协议转换
NBMA	非广播多路访问
ND	邻居发现
NGE	下一代加密
NHRP	下一跳解析协议
NLRI	网络层可达性信息
NMS	网络管理系统
NPTv6	IPv6 到 IPv6 网络前缀转换
NS	邻居请求
NSSA	非完全末节区域
NTP	网络时间协议
NVI	NAT 虚拟接口
OS	操作系统
OSI	开放式系统互联
OSPF	最短路径优先协议
OSPFv2	OSPF 版本 2
OSPFv3	OSPF 版本 3
OUI	机构唯一标识符

续表

缩写	全称
P	传播
PA	可汇聚提供商
PAT	端口地址转换
PBR	策略路由
PDM	协议相关模块
PDU	协议数据单元
PE	提供商边缘
PI	独立于提供商
PPP	点到点协议
pps	每秒数据包
QoS	服务直连
RA	路由器通告
RD	报告距离
RFC	征求修正意见书
RIB	路由信息库
RIP	路由信息协议
RIPE-NCC	欧洲 IP 网络资源协调中心
RIPng	下一代路由信息协议
RIPv1	路由信息协议版本 1
RIPv2	路由信息协议版本 2
RIRs	区域性 Internet 注册机构
RO	只读
RR	路由反射器
RS	路由器请求
RTO	重传超时
RTP	可靠传输协议
RTT	往返延迟
RTTMON	往返延迟监测
RW	读写
SA	安全关联
SHA	安全散列算法
SHA256	256 位 SHA
SIA	停滞在活动状态
SIEM	安全信息与事件管理
SLAAC	无状态地址自动配置
SLA	服务等级协定
SM	源度量
SMTP	简单邮件传输协议

缩写	全称
SNMP	简单网络管理协议
SNMPv1	SNMP 版本 1
SNMPv2	SNMP 版本 2
SNMPv3	SNMP 版本 3
SP	服务提供商
SPF	最短路径优先
SPI	安全参数索引
SRTT	平滑的往返延迟
SSH	安全外壳协议
SSHv1	SSH 版本 1
SSHv2	SSH 版本 2
SSL	安全套接字层
STP	1．屏蔽双绞线
	2．生成树协议
SYN	同步
TCP	传输控制协议
TCP/IP	传输控制协议/互联网协议
TFTP	小型文件传输协议
TLV	类型、长度、值
ToS	服务类型
TTL	生存时间
UDP	用户数据报协议
U/L	全局/本地
UPS	无间断电源
URL	统一资源定位符
uRPF	单播逆向路径转发
UTP	非屏蔽双绞线
VC	虚链路
VLAN	虚拟 LAN
VLSM	可变长子网掩码
VoIP	IP 语音
VPN	虚拟专用网
VRF	VPN 路由与转发
VRRP	虚拟路由器冗余协议
vty	虚拟终端
WAN	广域网
WWW	万维网

CCNP ROUTE 300-101 学习指南

本书是 CCNP ROUTE 300-101 的官方学习指南，涵盖了如下内容：

- 基本路由协议的特性和限制的比较；
- RIPv2 和 RIPng；
- EIGRP 在 IPv4 和 IPv6 环境下的操作和实施；
- OSPFv2 的实施，以及用于 IPv4 和 IPv6 的 OSPFv3；
- 通过路由更新来优化网络性能；
- 在 CEF 交换、策略路由和 SLA 中引入路径控制；
- 使用单个或冗余 ISP 连接来解决企业网络 Internet 连通性问题；
- BGP 术语、概念、操作、配置、验证和排错；
- 使用认证和推荐的其他方法来保护 Cisco 路由器的管理平面。

本书是 Cisco Press 出版的学习指南系列丛书之一，该系列丛书是 Cisco 唯一授权的学习工具，旨在帮助网络从业人员理解网络概念，为参加 CCNP 认证考试做好准备。

本书是 Cisco 唯一授权的 CCNP ROUTE 300-101 学习指南，详细阐述如何规划、配置、维护和扩展现代的路由网络。

本书以用于连接 LAN 和 WAN 的 Cisco 路由器为基础，讲解了如何选择和实施 Cisco IOS 服务来构建可扩展的路由网络。本书详细讲解了基本的网络和路由协议原理，介绍了 IPv4 和 IPv6，全面回顾了 EIGRP、OSPF 和 BGP，还讨论了企业网络的 Internet 连通性问题，探究了路由更新和路径控制，并且给出了当前的路由器安全最佳做法。

本书每章开篇列出了将要讲解的主题以及重点，每章末尾给出了供读者快速学习使用的关键概念的总结，以及供读者评估和加强每章知识掌握情况的复习题。穿插于全书中的配置示例和验证输出示例强调了网络维护和排错中的一些关键问题。

本书是备考 CCNP 认证的最佳读物，所有考生都可以从本书找到 CCNP ROUTE 300-101 认证考试的全部内容。

Diane Teare，拥有 P.Eng、CCNP、CCDP、CCSI、PMP 认证，是网络、培训、项目管理和在线教育领域的专家。她有 25 年的网络设计、实施和排错的经验。Diane 现在是一家 Cisco 培训合作伙伴的课程负责人，以及 CCNA 和 CCNP 路由交换课程的讲师。

Bob Vachon，是加拿大安大略省萨德伯里市凯布莱恩学院（Cambrian College）的教授，讲解网络架构课程。他有超过 30 年的网络和 IT 从业经验，其中有 13 年以团队领导、主要作者和主题专家的身份参与到 Cisco 和 Cisco 网络技术学院的各种 CCNA、CCNA-S 和 CCNP 项目中。

Rick Graziani，在加利福尼亚州阿普托斯的卡布利洛学院讲授计算机科学与网络课程。Rick 拥有近 30 年的计算机网络和 IT 领域的工作和教学经验。在从事教学工作以前，Rick 曾在多家公司的 IT 部门就职，其中包括 SCO 公司（Santa Cruz Operation）、天腾电脑公司、洛克希德公司。

ciscopress.com

异步社区 www.epubit.com.cn
新浪微博 @人邮异步社区
投稿/反馈邮箱 contact@epubit.com.cn

分类建议：计算机／计算机考试／思科认证
人民邮电出版社网址：www.ptpress.com.cn

ISBN 978-7-115-42507-2

ISBN 978-7-115-42507-2
定价：108.00 元