HCRSE107-BGP 双栈原理 Advance&Internet 设计

在较大规模组网或者路由条目较多的情况下，出于简化配置，减少路由条目，提升设备性能等等因素的考虑，会需要用到以下几种工具或技术：

路由聚合（Aggregation）
对等体组（Peer Group）
团体属性（Community ）
路由反射（Route Reflection ）
BGP 联盟（Confederations ）

路由聚合

只向对等体发送聚合后的路由，从而缩小路由表规模 ，明细路由如果发生路由振荡，不会对网络造成影响 ，路由聚合分为自动聚合和手动聚合。

路由聚合是会使用 Aggregator 属性（可选过渡属性），该属性标识发生聚合的节点，携带发生聚合节点的 router-id 和 AS 号。

对于 IPv4 路由，BGP 支持：手工聚合 ，自动聚合
对于 IPv6 路由，BGP 只支持：手工聚合

自动聚合

对 BGP 引入的路由进行有类聚合；配置聚合后，成员明细路由将被抑制；

仅对 import 引入的路由进行聚合。聚合之后的路由将带有 atomic_aggregate 和 aggregator 属性。

bgp 100
summary automatic
当打开时，系统会有提示信息，说明 BGP 自动路由聚合只适

用于通过路由引入方式引入的路由

```
[R1-bgp]summary automatic
Info: Automatic summarization is valid only for the routes imported through the
import-route command.
[R1-bgp]
```

手工聚合

对 BGP 本地路由表中的路由进行聚合。手动聚合可以控制聚合路由的属性，以及决定是否发布明细路由。聚合路由不会携带成员明细路由的 AS_PATH 属性。通过 AS_SET 属性来携带 AS 号，以避免环路

bgp 400
aggregate 192.168.0.0 21 detail-suppressed
as-set 可创建一条聚合路由，路径包含了具体路由的 AS 路径信息
detail-suppressed 抑制该聚合路由所包含的所有具体路由，只发布该聚合路由。
suppress-policy 能产生聚合路由，但抑制指定路由的通告。
attribute-policy 可设置聚合路由的属性。
origin-policy 抑制部分路由信息，并且当路由消失的时候也会使得聚合路由消失
仅在匹配 route-policy 时才生成聚合路由

对等体组

对等体组（Peer Group）是一些具有某些相同策略的对等体的集合。当一个对等体加入对等体组中时，此对等体将获得与所在对等体组相同的配置。当对等体组的配置改变时，组内成员的配置也相应改变。在大型 BGP 网络中，对等体的数量会很多，其中很多对等体具有相同的策略，在配置时会重复使用

一些命令，利用对等体组可以简化配置。对等体组中的单个对等体也可以配置自己的发布路由与接收路由的策略。
bgp 100
group 1
peer 2.2.2.2 group 1
peer 3.3.3.3 group 1
peer 1 reflect-client
peer 1 next-hop-local

BGP 按组打包
按组打包技术将所有拥有共同出口策略的 BGP 邻居当作是一个打包组
每条待发送的路由只被打包一次然后发给组内的所有邻居

团体属性
团体属性用来简化路由策略的应用和降低维护管理的难度，利用团体可以使多个 AS 中的一组 BGP 设备共享相同的策略。
团体是一个路由属性，在 BGP 对等体之间传播，且不受 AS 的限制。BGP 设备在将带有团体属性的路由发布给其它对等体之前，可以先改变此路由原有的团体属性。
团体属性用于标识具有相同特征的 BGP 路由，该属性为可选过渡
bgp 2001
peer 192.168.45.5 advertise-community
团体属性分为：自定义团体属性 ,公认团体属性

Internet
No_Advertise
No_Export
No_Export_Subconfed

internet : 可以向任何 BGP 对等体发布路由
no-export : 不会发给 EBGP 对等体，但可以发布给联盟（Confederation）EBGP 对等体
no-advertise : 不会发给任何 BGP 对等体
no-export-Subconfed : 不发给 EBGP 对等体，也不发布给联盟（Confederation）EBGP 对等体

路由反射器 RR（Route Reflector）

为保证 IBGP 对等体之间的连通性，需要在 IBGP 对等体之间建立全连接（Full-mesh）关系。假设在一个 AS 内部有 n 台路由器，那么应该建立的 IBGP 连接数就为 n(n-1)/2。当 IBGP 对等体数目很多时，对网络资源和 CPU 资源的消耗都很大。利用路由反射可以解决这一问题。

在一个 AS 内，其中一台路由器作为路由反射器 RR（Route Reflector），其它路由器作为客户机（Client）。客户机与路由反射器之间建立 IBGP 连接。路由反射器和它的客户机组成一个集群（Cluster）。路由反射器在客户机之间反射路由信息，客户机之间不需要建立 BGP 连接。
路由反射器 RR：允许把从 IBGP 对等体学到的路由反射到其他 IBGP 对等体的 BGP 设备。
客户机（Client）：与 RR 形成反射邻居关系的 IBGP 设备。在 AS 内部客户机只需要与 RR 直连。
非客户机（Non-Client）：既不是 RR 也不是客户机的 IBGP 设备。在 AS 内部非客户机与 RR 之间，以及所有的非客户机之间仍然必须建立全连接关系。
始发者（Originator）：在 AS 内部始发路由。Originator_ID 属性用于防止集群内产生路由环路。
集群（Cluster）：路由反射器及其客户机的集合。Cluster_Li

st 属性用于防止集群间产生路由环路。

在向 IBGP 邻居发布学习到的路由信息时，RR 按照以下规则发布路由
从 EBGP 对等体学到的路由，发布给所有的非客户机和客户机。
从非客户机 IBGP 对等体学到的路由，发布给此 RR 的所有客户机。
从客户机学到的路由，发布给此 RR 的所有非客户机和客户机（发起此路由的客户机除外）。
RR 的配置方便，只需要对作为反射器的路由器进行配置，客户机并不需要知道自己是客户机。
在某些网络中，路由反射器的客户机之间已经建立了全连接，它们可以直接交换路由信息，此时客户机到客户机之间的路由反射是没有必要的，而且还占用带宽资源。VRP 支持配置命令 undo reflect between-clients 来禁止 RR 将从客户机收到的路由再反射给其他客户机。

bgp 100
und reflect between-clients

路由反射器 RR 防环 ：Originator_ID Cluster_List
Originator ID 由 RR 产生，使用的 Router ID 的值标识路由发送者，用于防止集群内产生路由环路。
当一条路由第一次被 RR 反射的时候，RR 将 Originator_ID 属性加入这条路由，标识这条路由的发起设备。如果一条路由中已经存在了 Originator_ID 属性，则 RR 将不会创建新的 Originator_ID 属性。
当设备接收到这条路由的时候，将比较收到的 Originator ID 和本地的 Router ID，如果两个 ID 相同，则不接收此路由。

路由反射器和它的客户机组成一个集群（Cluster）。在一个AS 内，每个路由反射器使用唯一的 Cluster ID 作为集群标识。为了防止集群间产生路由环路，路由反射器使用 Cluster_List 属性，记录路由经过的所有集群的 Cluster ID。当 RR 在它的客户机之间或客户机与非客户机之间反射路由时，RR 会把本地 Cluster_ID 添加到 Cluster_List 的前面。如果 Cluster_List 为空，RR 就创建一个。
当 RR 接收到一条更新路由时，RR 会检查 Cluster_List。如果 Cluster_List 中已经有本地 Cluster_ID，丢弃该路由；如果没有本地 Cluster_ID，将其加入 Cluster_List，然后反射该更新路由。

备份 RR

相同集群中的路由反射器要共享相同的 Cluster_ID；Cluster_List 的应用保证了同一 AS 内的不同 RR 之间不出现路由循环。

RR1 接收到该更新路由后，它向其他的客户机（Client2、Client3）和非客户机（RR2）反射，同时将本地 Cluster_ID 添加到 Cluster_List 前面。
RR2 接收到该反射路由后，检查 Cluster_List，发现自己的 Cluster_ID 已经包含在 Cluster_List 中。因此，它丢弃该更新路由，不再向自己的客户机反射。

BGP 联盟

联盟将一个 AS 划分为若干个子 AS。每个子 AS 内部建立 IBGP 全连接关系，子 AS 之间建立联盟 EBGP 连接关系，但联盟外部 AS 仍认为联盟是一个 AS。配置联盟后，原 AS 号将作为每个路由器的联盟 ID。
联盟的防环机制

AS_CONFED_SEQUENCE
AS_CONFED_SET

AS_PATH 属性被定义为公认必遵属性，该属性由 AS 号所组成。AS_PATH 包含 4 种不同类型
AS_SET: 由一系列 AS 号无序地组成，包含在 UPDATE 消息里。当网络发生聚合时，可通过适当策略使 AS_PATH 使用类型 AS_SET 来避免路径信息丢失。
AS_SEQUENCE: 由一系列 AS 号顺序地组成，包含在 UPDATE 消息里。一般情况下，AS_PATH 类型为 AS_SEQUENCE。
AS_CONFED_SEQUENCE: 在本地联盟内由一系列成员 AS 号按顺序地组成，包含在 UPDATE 消息中，用法和 AS_SEQUENCE 相同，只能在本地联盟内传递。
AS_CONFED_SET: 在本地联盟内由一系列成员 AS 无序地组成，包含在 UPDATE 消息中，用法和 AS_SET 相同，只能在本地联盟内传递。
联盟内部的成员 AS 号对于其他非联盟 AS 是不可见的，所以路由在从联盟内部发送到其他非联盟 AS 时，联盟成员 AS 号被剥离。

反射器和联盟的比较
联盟需要重新划分区域，对现网改动较大。
反射器在配置时，只需要对 RR 进行配置，客户机不需要做任何其他的操作；联盟需要在所有路由器上进行配置。
RR 与 RR 间需要 IBGP 全互联。路由反射器应用较为广泛；联盟应用较少。

==========
BGP 扩展特性 - 安全特性

MD5
GTSM
限制从对等体接收的路由数量
AS_Path 长度保护

1　MD5：BGP 使用 TCP 作为传输层协议，为提高 BGP 的安全性，可以在建立 TCP 连接时进行 MD5 认证。但 BGP 的 MD5 认证并不能对 BGP 报文认证，它只是为 TCP 连接设置 MD5 认证密码，由 TCP 完成认证。如果认证失败，则不建立 TCP 连接。
2　GTSM（Generalized TTL Security Mechanism）：通用 TTL 安全保护机制
GTSM 通过检测 IP 报文头中的 TTL 值是否在一个预先定义好的特定范围内，对 IP 层以上业务进行保护，增强系统的安全性。使能 BGP 的 GTSM 策略后，接口板对所有 BGP 报文的 TTL 值进行检查。根据实际组网的需要，对于不符合 TTL 值范围的报文，GTSM 可以设置为通过或丢弃。配置 GTSM 缺省动作为丢弃时，可以根据网络拓扑选择合适的 TTL 有限值范围，不符合 TTL 值范围的报文会被接口板直接丢弃，这样就避免了网络攻击者模拟的“合法”BGP 报文占用 CPU。该功能与 EBGP 多跳互斥。

bgp 10
peer 10.1.1.2 valid-ttl-hops 1　　要求收到报文的 TTL 范围为[255,255]：

指定需要检测的 TTL 跳数值。 整数形式，取值范围是 1～255，缺省值是 255。如果配置为 hops，则被检测的报文的 TTL 值有效范围为[255–hops+1, 255]

gtsm default-action { drop | pass }
设置未匹配 GTSM 策略的报文的缺省动作，缺省情况下，默认动作是 pass，即未匹配 GTSM 策略的报文可以通过过滤。

3　限制从对等体接收的路由数量，防止资源耗尽性攻击。
当设备遭到恶意攻击或者网络中出现错误配置时，会导致 BGP 从邻居接收到大量的路由，从而消耗大量设备的资源。因此管理员必须根据网络规划和设备容量，对运行时所使用的资源进行制。BGP 提供了基于对等体的路由控制，限定邻居发来的路由数量，这样可以避免上述问题
bgp 100
peer 1.1.1.1 filter-policy 10 import
peer 1.1.1.1 filter-policy 10 export

peer 1.1.1.1 route-limit 50

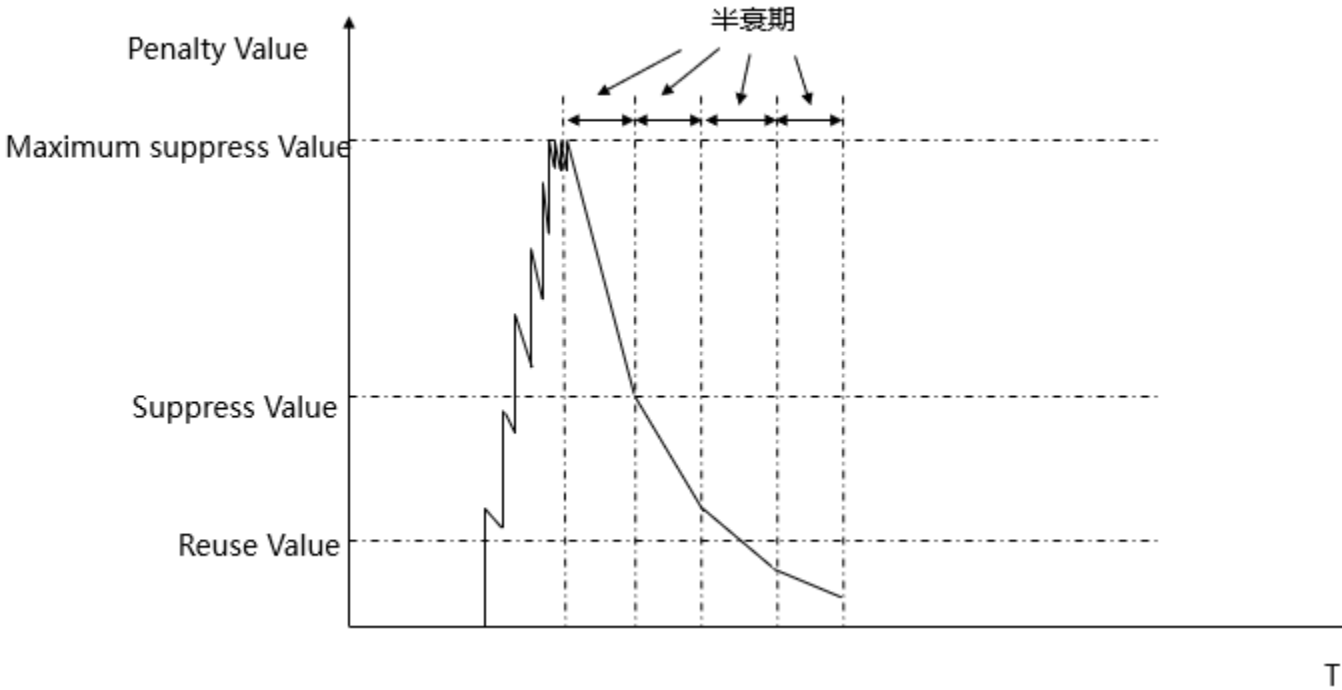
4　AS_Path 长度保护。通过在入口和出口两个方向对 AS_Path 的长度进行限定，直接丢弃 AS_Path 超限的报文。

========
BGP 扩展特性 - 路由衰减　　用来解决路由不稳定的问题

路由衰减（Route Dampening）用来解决路由不稳定的问题。多数情况下，BGP 协议都应用于复杂的网络环境中，路由变化十分频繁。为了防止持续的路由振荡带来的不利影响，BGP 使用路由衰减来抑制不稳定的路由。

BGP 衰减使用惩罚值（Penalty Value）来衡量一条路由的稳定性，惩罚值越高则说明路由越不稳定。路由每发生一次振荡（路由从激活状态变为未激活状态，称为一次路由振荡），BGP 便会给此路由增加一定的惩罚值（1000）。当惩罚值超过

抑制阈值（Suppress Value）时，此路由被抑制，不加入到路由表中，也不再向其他 BGP 对等体发布更新报文。
当某条路由的惩罚值到达最大抑制值（Maximum Suppress Value）,便不会再增加，这样就可以确保某路由在非常短的时间内翻动十几次之后，不会将惩罚值累加到一个很高的、使路由始终保持被抑制状态的值。



被抑制的路由每经过一段时间，惩罚值便会减少一半，这个时间称为半衰期（Half-life）。当惩罚值降到再使用阈值（Reuse Value）时，此路由变为可用并被加入到路由表中，同时向其他 BGP 对等体发布更新报文。上文提到的惩罚值、抑制阈值和半衰期都可以手动配置。

路由衰减只适用于 EBGP 路由。对于从 IBGP 收来的路由不能进行衰减，因为 IBGP 路由经常含有本 AS 的路由，内部网络路由要求转发表尽可能一致。如果衰减对 IBGP 路由起作用，不同设备的衰减参数不一致时，会导致转发表不一致。

dampening 命令只对 EBGP 路由生效。　dampening ibgp 命令只对 BGP VPNv4 路由生效

bgp 100
dampening 10 1000 2000 5000

half-life-reach　指定可达路由的半衰期。单位为分钟，取值范围为 1～45。缺省值为 15。
reuse　　指定路由解除抑制状态的阈值。取值范围为 1～20000。缺省值为 750。
suppress 指定路由进入抑制状态的阈值　取值范围为 1～20000，缺省值为 2000。
ceiling　　惩罚上限值。取值范围为 1001～20000。缺省值为 16000。
所指定的 reuse、suppress、ceiling 三个阈值是依次增大的，reuse<suppress<ceiling。

========

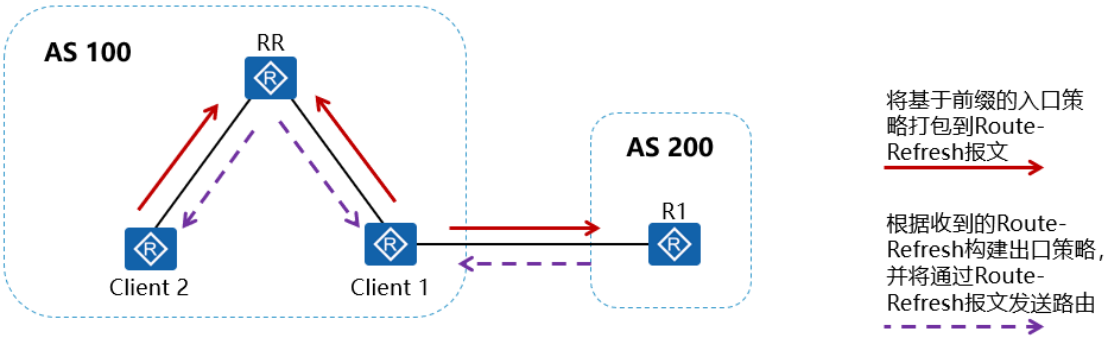
BGP 增强特性 - BGP ORF

BGP 基于前缀的 ORF（ Outbound Route Filtering ）功能：基于本地的入口策略构建对端的出口策略，实现 BGP 按需发布路由；

BGP 基于前缀的 ORF 能力，能将本端设备配置的基于前缀的入口策略通过路由刷新报文发送给 BGP 邻居。BGP 邻居根据这些策略构造出口策略，在路由发送时对路由条目进行过滤。这样不仅避免了本端设备接收大量无用的路由，降低了本端设备的 CPU 使用率，还有效减少了 BGP 邻居的配置工作，降低了链路带宽的占用率。
ORF (Outbound Route Filter)功能，是指将对端的入口策略作

为本端的出口策略，对发送的路由进行过滤。不符合对端入口策略的路由将不发送给对端。

bgp 100
peer 192.168.1.1 capability-advertise orf ip-prefix both
both 表示允许发送和接收 ORF 报文。



直连 EBGP 邻居中，Client1、R1 协商基于前缀的 ORF 能力后，Client1 将本地配置的基于前缀的入口策略打包到 Route-refresh 报文中发送给 R1。R1 根据接收到的路由刷新报文构造出口策略，通过 Route-refresh 报文发送路由给 Client1。Client1 只收到它需要的路由，而 R1 不必维护路由策略，减少了配置工作。
Client1、Client2 为 RR 的客户端，Client1 与 RR、Client2 与 RR，分别协商基于前缀的 ORF 能力，Client1、Client2 将本地配置的基于前缀的入口策略打包到 Route-refresh 报文中发送给 RR。RR 根据接收到的 Client1、Client2 基于前缀的入口策略，构造 RR 的出口策略，将路由反射给 Client1、Client2。Client1 和 Client2 只收到需要的路由，RR 不必维护路由策略，减少了配置工作。

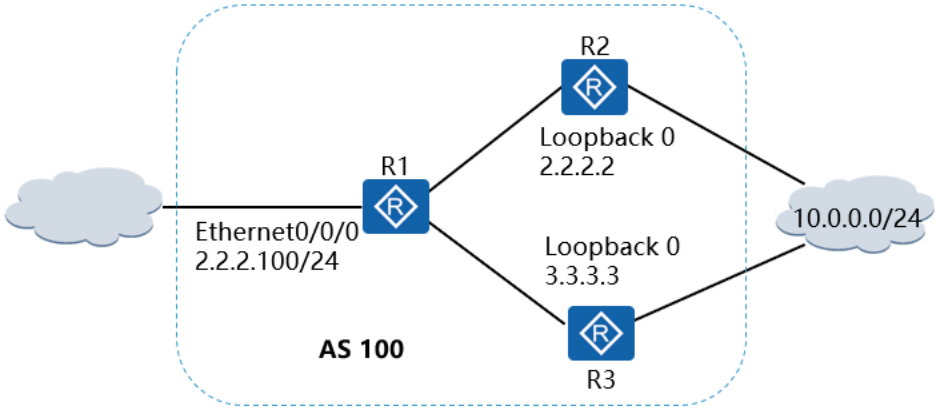
BGP 增强特性 - Active-Route-Advertise

只有当 BGP 路由被成功的安装进 IP 或 IPv6 路由表，该路由才能被发送给邻居。
默认情况下路由只需在 BGP 中优选即可向邻居发布。配置了此特性之后，路由必须同时满足在 BGP 协议层面优选与在路由管理层面活跃两个条件，才能向邻居发布。

bgp 100
active-route-advertise

BGP 增强特性 - 按策略进行下一跳迭代

通过配置路由策略来限制迭代到的路由。如果路由不能通过路由策略，则该路由迭代失败。
BGP 需要对非直连的下一跳进行路由迭代，但是如果不对迭代到的路由进行过滤的话，可能会迭代到一个错误的转发路径上。按策略进行下一跳迭代就是通过配置路由策略来限制迭代到的路由。如果路由不能通过路由策略，则该路由迭代失败。



R1 和 R2、R3 之间通过 Loopback 口建立 IBGP 邻居。R1 从 R2、R3 分别收到了前缀为 10.0.0.0/24 的 BGP 路由。其中从 R2 收到的 BGP 路由的原始下一跳为 2.2.2.2。另外，R1 上 Ethernet0/0/0 的接口地址为 2.2.2.100/24。

当 R2 正常运行时，R1 收到从 R2 发来的前缀为 10.0.0.0/24 的路由会迭代到 IGP 路由 2.2.2.2/32。但是当 R2 的 IGP 发生故障时，IGP 路由 2.2.2.2/32 被撤销，这样就导致下一跳重新迭代。在 R1 上会用原始下一跳 2.2.2.2 在 IP 路由表中进行最长匹配迭代，结果会迭代到 2.2.2.0/24 的路由上。但此时用户期望的是，当到 2.2.2.2 的路由不可达时，可以重新选路优选到 3.3.3.3 的路由。实际上该故障主要是由于 BGP 收敛引起的，从而产生了路由的瞬时黑洞。
配置下一跳迭代策略，可以通过到 BGP 路由原始下一跳所依赖路由的掩码长度来过滤迭代路由。可以通过配置下一跳迭代策略，使到原始下一跳 2.2.2.2 只能依赖于 2.2.2.2/32 的 IGP 路由。
bgp 100
nexthop recursive-lookup route-policy rp_nexthop

=========

什么样的网络需要 BGP

常见的企业网络拓扑类型
单归属自治系统（一个出口设备且连接到一个 ISP）
多归属到单自治系统（多个出口设备仅连接到一个 ISP）
多归属到多自治系统（多个出口设备且连接到多个 ISP）

单归属自治系统：仅有一个出口设备且只连接到一个 ISP。这种情况下，就可以不需要配置 BGP 协议。可在用户边界设备上添加一条默认路由，并宣告到用户自治系统内部。

多归属到单自治系统：增加了对链路和网络设备的冗余性，一般这种情况下用户网络用的会是私有 AS 号。若两条链路采用主备的方式，那么也不需要采用 BGP。两台出口设备分别向本自治系统内的设备宣告 metric 值不同的默认路由即可。

（若采用 OSPF 为 IGP，外部路由的 cost 应该采用 E2 方式，仅考虑外部开销（cost））

若两台路由器采用负载分担方式：
方式一：两台路由器分别向自治系统内（IGP 采用 OSPF）宣告 cost type 为 E1 的默认路由，使得自治系统其他路由器选择距离自己最近的出口路由器到达外部网络。这种情况也可以不使用 BGP。但是当两个出口路由器的物理间隔十分大，并且对时延有很高要求时，就可以考虑采取 BGP 来获取更精细的路由条目。
方式二：与 ISP 设备之间建立 BGP 连接，从 BGP 接收更为精细的路由条目，配合上路由策略工具的使用，来达到针对不同目的地址使用不同出口路由的目的。

多归属到多自治系统：不仅增加了对链路和网络设备的冗余性，同时使用了做到了 ISP 的冗余备份。
对于这种自治系统，需要充分考虑到地址空间是否独立于运营商，是否拥有公有 AS 号等问题。
理想情况下，当用户网络拥有独立于 ISP 的地址空间和公有 AS 号时，有三种部署方式
方式一：采取主备方式，出口路由器向内部宣告开销不一样的默认路由。
方式二：负载分担方式，出口路由器向内部宣告默认路由，仅使用 IGP 的开销计算机制，由 IGP 自行决定使用哪一台出口路由器。
方式三：部署 BGP。考虑与 ISP 签署的合约，企业本身的业务流量特点等因素，使用各种路由策略工具，如有必要也可以使用默认路由宣告等方式。充分控制企业进和出方向的流量。
一般情况下，多归属到多自治系统的网络会考虑部署 BGP 协议，因为前两种方法不利于路由的控制。但是不是绝对的，需

要仔细权衡所得到的好处与增加路由复杂度所带来的代价。

BGP 路由劫持

产生原因：BGP 协议里虽然有一些简单的安全认证的部分，但是对于两个已经成功建立 BGP 连接的 AS 来说，基本会无条件的相信对方 AS 所传来的信息，包括对方声称所拥有的 IP 地址范围。

潜在危害：若无条件相信对方发送过来的 Update 消息，不排除恶意的 AS 宣告不存在的 IP 网段，通过修改 AS_Path 等 BGP 属性，让其他 AS 认为这条路径才是到达这个目的网段的最短路径，那么该恶意的 AS 就能截获到数据流量。

不对称路由

产生原因：不恰当的属性使用或者是路由聚合不合理导致路由精准性不足，导致流量的出方向和入方向不同。

潜在危害：首先，不对称流量会使互联网络的流量模型变得难以预测，使得网络基准、容量规划、故障检测及排除变得困难；其次，不对称流量会使链路使用率出现不均衡，某些链路的带宽出现饱和，而其他链路的带宽却得不到有效利用；再次，不对称流量会使出流量和入流量的时延出现很大的差异，这种时延变化（即抖动）会对某些时延敏感型应用（如语音和直播视频）造成损害。

前言

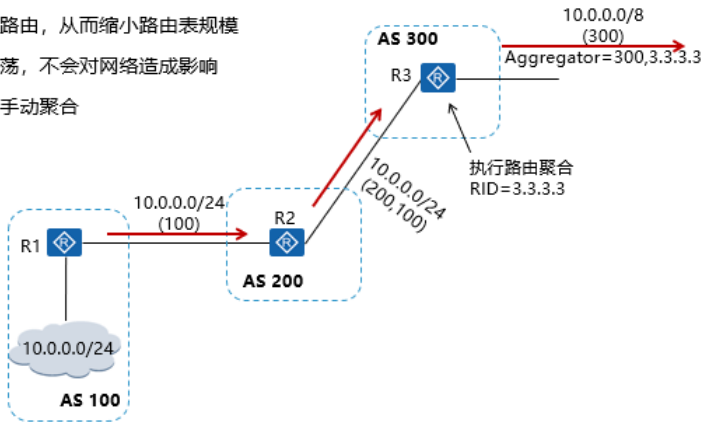
- BGP提供各种高级特性与技术供使用
- 合理运用各种特性与技术可提升网络性能
- BGP网络可能存在的问题

BGP大规模路由应用

- 在较大规模组网或者路由条目较多的情况下，出于简化配置，减少路由条目，提升设备性能等等因素的考虑，会需要用到以下几种工具或技术：
  - 路由聚合（Aggregation）
  - 对等体组（Peer Group）
  - 团体属性（Community ）
  - 路由反射（Route Reflection ）
  - BGP联盟（Confederations ）

路由聚合

- 路由聚合
  - 只向对等体发送聚合后的路由，从而缩小路由表规模
  - 明细路由如果发生路由振荡，不会对网络造成影响
  - 路由聚合分为自动聚合和手动聚合
  - 对于IPv4路由，BGP支持
    - 手工聚合
    - 自动聚合
  - 对于IPv6路由，BGP支持
    - 手工聚合

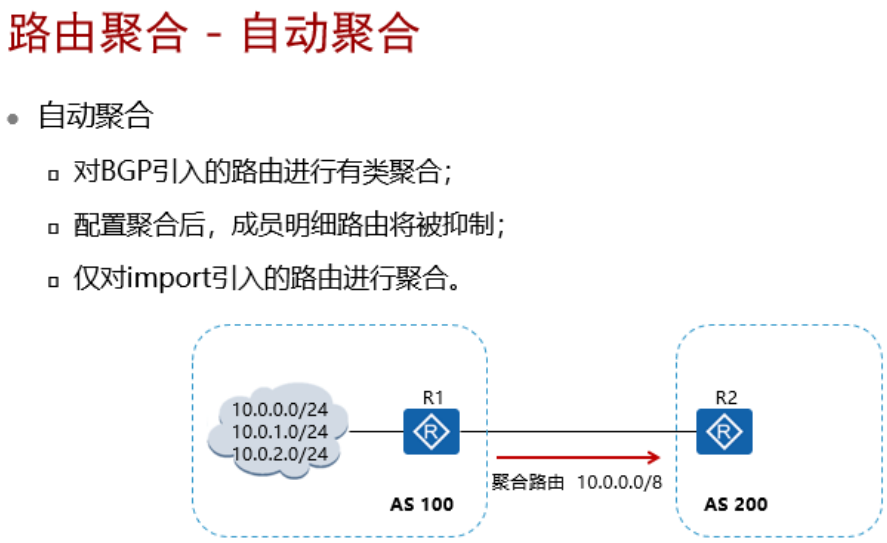


- 在大规模的网络中，BGP 路由表十分庞大，给设备造成了很大的负担，同时使发生路由振荡的几率也大大增加，影响

网络的稳定性。

- 路由聚合是将多条路由合并的机制，它通过只向对等体发送聚合后的路由而不发送所有的具体路由的方法，减小路由表的规模。并且被聚合的路由如果发生路由振荡，也不再对网络造成影响，从而提高了网络的稳定性。

- 路由聚合是会使用 Aggregator 属性（可选过渡属性），该属性标识发生聚合的节点，携带发生聚合节点的 router-id 和 AS 号。

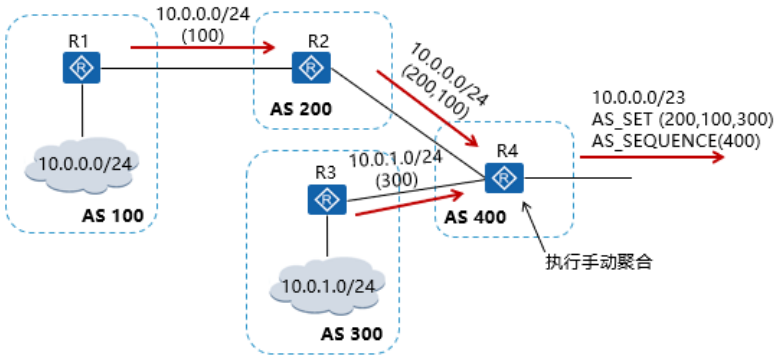


- 自动聚合注意事项
- 该命令对 BGP 引入的路由进行聚合，引入的路由可以是直连路由、静态路由、OSPF 路由、IS-IS 路由。配置聚合后，BGP 将按照自然网段聚合路由，明细路由在 BGP 路由更新中被抑制。该命令对 network 命令引入的路由无效。
- BGP 只向对等体发送聚合后的路由；
- 缺省情况下 BGP 不启用自动聚合；
- 聚合之后的路由将带有 atomic_aggregate 和 aggregator

属性。

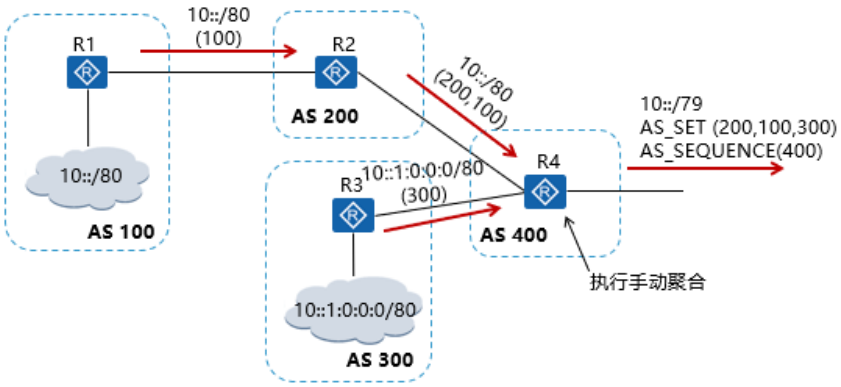
路由聚合 - 手工聚合 (1)

- 手工聚合
  - 对BGP本地路由表中的路由进行聚合。手动聚合可以控制聚合路由的属性，以及决定是否发布明细路由。



- 手动聚合
- 可通过命令决定是否抑制明细路由，抑制后该聚合后的路由会携带 atomic_aggregate 属性。
- 聚合路由不会携带成员明细路由的 AS_PATH 属性。
- 通过 AS_SET 属性来携带 AS 号，以避免环路。SET 和 SEQUENCE 的不同之处在于，SET 选项下的 AS 列表通常用于路由聚合，将来自不同 AS 的 AS 号无序排列在 AS 列表里；而 SEQUENCE 选项下的 AS 列表是有序的，每经过一个 AS 都会将其 AS 号排列在列表的前端。

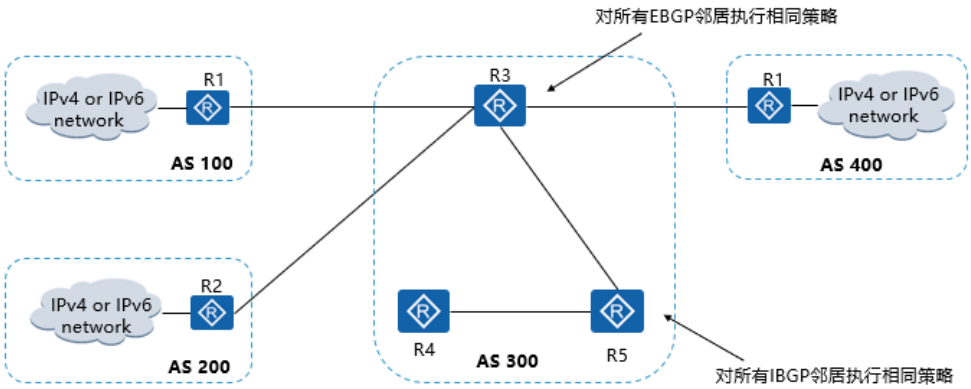
路由聚合 - 手工聚合 (2)



- 手动聚合
- 可通过命令决定是否抑制明细路由，抑制后该聚合后的路由会携带 atomic_aggregate 属性。
- 聚合路由不会携带成员明细路由的 AS_PATH 属性。
- 通过 AS_SET 属性来携带 AS 号，以避免环路。SET 和 SEQUENCE 的不同之处在于，SET 选项下的 AS 列表通常用于路由聚合，将来自不同 AS 的 AS 号无序排列在 AS 列表里；而 SEQUENCE 选项下的 AS 列表是有序的，每经过一个 AS 都会将其 AS 号排列在列表的前端。

对等体组

- 对等体组（Peer Group）是一些具有某些相同策略的对等体的集合。
- 此功能可以简化BGP的配置，同时减少路由性能损耗。



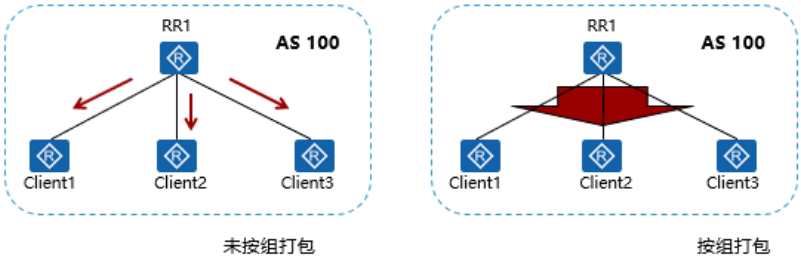
- 对等体组（Peer Group）是一些具有某些相同策略的对等体的集合。当一个对等体加入对等体组中时，此对等体将获得与所在对等体组相同的配置。当对等体组的配置改变时，组内成员的配置也相应改变。

- 在大型 BGP 网络中，对等体的数量会很多，其中很多对等体具有相同的策略，在配置时会重复使用一些命令，利用对等体组可以简化配置。

- 对等体组中的单个对等体也可以配置自己的发布路由与接收路由的策略。

按组打包

- BGP按组打包
  - 按组打包技术将所有拥有共同出口策略的BGP邻居当作是一个打包组
  - 每条待发送路由只被打包一次然后发给组内的所有邻居

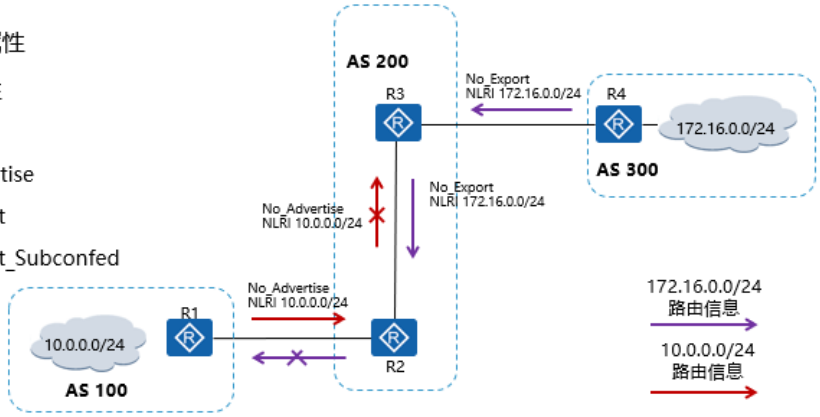


- BGP 按组打包
- 缺省情况下，BGP 会针对不同邻居（即使出口策略相同）单独打包路由。
- 应用按组打包功能后，每条待发送路由只被打包一次然后发给组内的所有邻居，使打包效率成倍提升。

- 拓扑描述
- 一个反射器有 3 个客户机，有 10 万条路由需要反射。如果按照每个邻居分别打包的方式，反射器 RR 在向 3 个客户机发送路由的时候，所有路由被打包的总次数是 10 万×3。而按组打包技术将这个过程变为 10 万×1，性能相当于提升了 3 倍。

BGP属性特点—团体（COMMUNITY）

- 团体属性用于标识具有相同特征的BGP 路由，该属性为可选过渡
  - 团体属性分为：
    - 自定义团体属性
    - 公认团体属性
      - Internet
      - No_Advertise
      - No_Export
      - No_Export_Subconfed



团体属性 - 公认属性

- 团体属性用一组以4字节为单位的列表来表示，格式有：
  - aa:nn：aa和nn的取值范围都是0～65535。
  - 团体号：团体号是0～4294967295的整数。
  - 标准协议中定义，0（0x00000000）～65535（0x0000FFFF）和4294901760（0xFFFF0000）～4294967295（0xFFFFFFFF）是预留的。

团体名称	团体标识	说明
Internet	0（0x00000000）	缺省情况下，所有路由器都属于Internet团体。具有此属性的路由可以被通告给所有的BGP对等体。
No_Export	4294967041（0xFFFFFF01）	具有此属性的路由在收到后，不能被发布到本地AS之外。
No_Advertise	4294967042（0xFFFFFF02）	具有此属性的路由在收到后，不能被通告给任何其他的BGP对等体。
No_Export_Subconfed	4294967043（0xFFFFFF03）	具有此属性的路由在收到后，不能被发布到本地AS之外，也不能被发布到其他子AS。

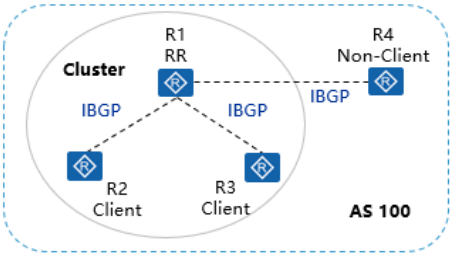
- 团体属性是一组有相同特征的目的地址的集合。团体属性用一组以 4 字节为单位的列表来表示，设备中团体属性的格式是 aa:nn 或团体号。
- aa:nn：aa 和 nn 的取值范围都是 0～65535，管理员可根据实际情况设置具体数值。通常 aa 表示自治系统 AS 编号，nn 是管理员定义的团体属性标识。例如，来自 AS100 的一条路由，管理员定义的团体属性标识是 1，则该路由的团体属性格式是 100:1。

- 团体号：团体号是 0～4294967295 的整数。RFC1997 中定义，0（0x00000000）～65535（0x0000FFFF）和 4294901760（0xFFFF0000）～4294967295（0xFFFFFFFF）是预留的。
- 团体属性用来简化路由策略的应用和降低维护管理的难度，利用团体可以使多个 AS 中的一组 BGP 设备共享相同的策略。团体是一个路由属性，在 BGP 对等体之间传播，且不受 AS 的限制。BGP 设备在将带有团体属性的路由发布给其它对等体之前，可以先改变此路由原有的团体属性。

- 公认团体属性
- Internet：缺省情况下，所有的路由都属于 Internet 团体。具有此属性的路由可以被通告给所有的 BGP 对等体。
- No_Advertise：具有此属性的路由在收到后，不能被通告给任何其他的 BGP 对等体。
- No_Export：具有此属性的路由在收到后，不能被发布到本地 AS 之外。如果使用了联盟，则不能被发布到联盟之外，但可以发布给联盟中的其他子 AS。
- No_Export_Subconfed：具有此属性的路由在收到后，不能被发布到本地 AS 之外，也不能发布到联盟中的其他子 AS。

路由反射器

- 路由反射器
  - 允许将从IBGP邻居学习到的路由发送给特定IBGP邻居，打破了IBGP邻居关系全互联的需求，减少IBGP会话数量。
  - 包括路由反射器（RR）和客户机（Client）。



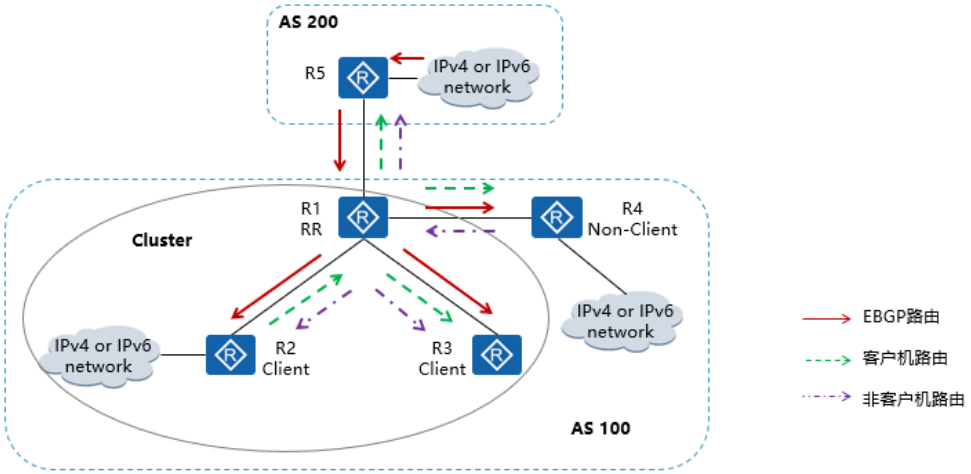
- 为保证 IBGP 对等体之间的连通性，需要在 IBGP 对等体之间建立全连接（Full-mesh）关系。假设在一个 AS 内部有 n 台路由器，那么应该建立的 IBGP 连接数就为 n(n-1)/2。当 IBGP 对等体数目很多时，对网络资源和 CPU 资源的消耗都很大。利用路由反射可以解决这一问题。

- 在一个 AS 内，其中一台路由器作为路由反射器 RR（Route Reflector），其它路由器作为客户机（Client）。客户机与路由反射器之间建立 IBGP 连接。路由反射器和它的客户机组成一个集群（Cluster）。路由反射器在客户机之间反射路由信息，客户机之间不需要建立 BGP 连接。

- 路由反射器概念
- 路由反射器 RR（Route Reflector）：允许把从 IBGP 对等体学到的路由反射到其他 IBGP 对等体的 BGP 设备。
- 客户机（Client）：与 RR 形成反射邻居关系的 IBGP 设备。在 AS 内部客户机只需要与 RR 直连。
- 非客户机（Non-Client）：既不是 RR 也不是客户机的 IBGP 设备。在 AS 内部非客户机与 RR 之间，以及所有的非客户机之间仍然必须建立全连接关系。

- 始发者（Originator）：在 AS 内部始发路由的设备。Originator_ID 属性用于防止集群内产生路由环路。
- 集群（Cluster）：路由反射器及其客户机的集合。Cluster_List 属性用于防止集群间产生路由环路。

路由反射器 - 反射规则



- 在向 IBGP 邻居发布学习到的路由信息时，RR 按照以下规则发布路由
- 从 EBGP 对等体学到的路由，发布给所有的非客户机和客户机。
- 从非客户机 IBGP 对等体学到的路由，发布给此 RR 的所有客户机。
- 从客户机学到的路由，发布给此 RR 的所有非客户机和客户机（发起此路由的客户机除外）。
- RR 的配置方便，只需要对作为反射器的路由器进行配置，客户机并不需要知道自己是客户机。
- 在某些网络中，路由反射器的客户机之间已经建立了全连接，它们可以直接交换路由信息，此时客户机到客户机之间的路由反射是没有必要的，而且还占用带宽资源。VRP 支持配置命令 undo reflect between-clients 来禁止 RR 将从客户机收到的路由再反射给其他客户机。

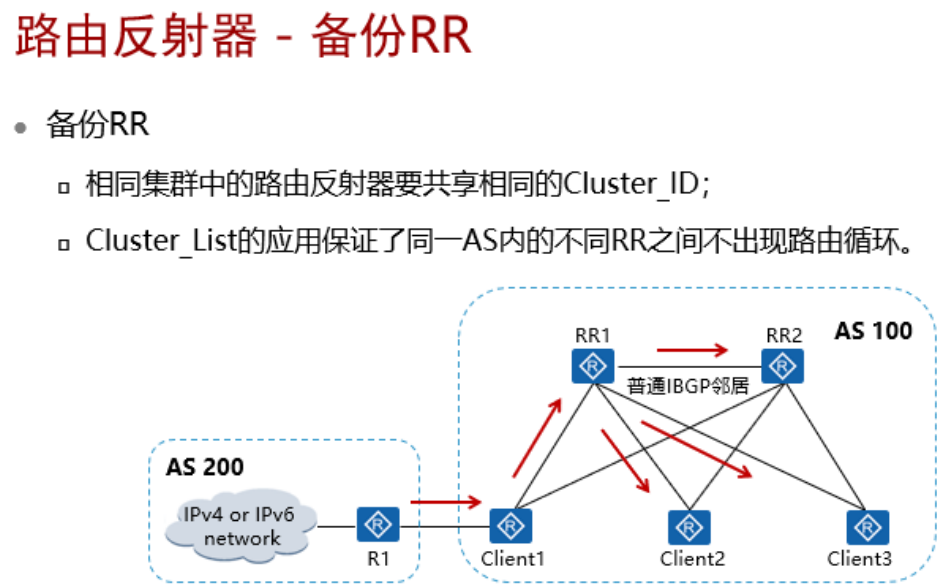
路由反射器 - 防环机制

- 路由反射器的防环机制
  - Originator_ID属性
    - 该属性为可选非过渡；
    - 用于集群内的防环；
    - 由路由反射器（RR）产生，携带了本地AS内该路由发送者的Router ID。
  - Cluster_List属性
    - 该属性为可选非过渡；
    - 用于集群间的防环；
    - 由每个路由反射器（RR）产生，记录反射路由经过的集群。

- Originator ID 由 RR 产生，使用的 Router ID 的值标识路由的发送者，用于防止集群内产生路由环路。
- 当一条路由第一次被 RR 反射的时候，RR 将 Originator_ID 属性加入这条路由，标识这条路由的发起设备。如果一条路由中已经存在了 Originator_ID 属性，则 RR 将不会创建新的 Originator_ID 属性。
- 当设备接收到这条路由的时候，将比较收到的 Originator ID 和本地的 Router ID，如果两个 ID 相同，则不接收此路由。

- 路由反射器和它的客户机组成一个集群（Cluster）。在一个 AS 内，每个路由反射器使用唯一的 Cluster ID 作为集群标识。
- 为了防止集群间产生路由环路，路由反射器使用 Cluster_List 属性，记录路由经过的所有集群的 Cluster ID。
- 当 RR 在它的客户机之间或客户机与非客户机之间反射路由时，RR 会把本地 Cluster_ID 添加到 Cluster_List 的前面。如果 Cluster_List 为空，RR 就创建一个。
- 当 RR 接收到一条更新路由时，RR 会检查 Cluster_List。

如果 Cluster_List 中已经有本地 Cluster_ID，丢弃该路由；如果没有本地 Cluster_ID，将其加入 Cluster_List，然后反射该更新路由。



- 备份 RR 主要是为了解决单点故障。

- 备份 RR
- VRP 需要使用命令 reflector cluster-id 给所有位于同一个集群内的路由反射器配置相同的 Cluster_ID。
- 在冗余的环境里，客户机会收到不同反射器发来的到达同一目的地的多条路由，这时客户机应用 BGP 选择路由的策略来选择最佳路由。
- Cluster_List 的应用保证了同一 AS 内的不同 RR 之间不出现路由循环。

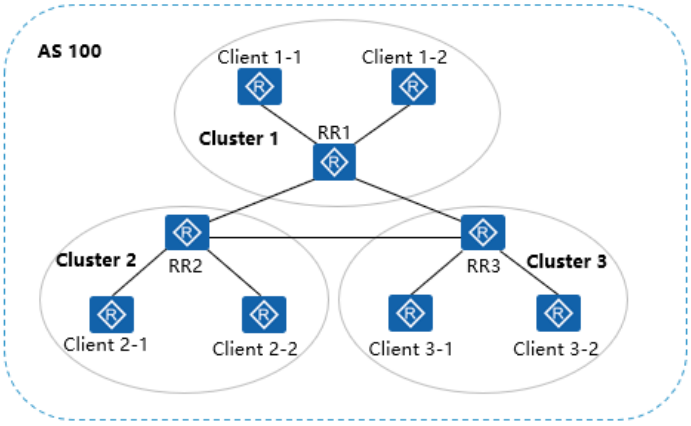
- 拓扑描述
- 当客户机 Client1 从外部对等体接收到一条更新路由（10.0.0.0/24）后，它通过 IBGP 向 RR1 和 RR2 通告这条路由。
- RR1 接收到该更新路由后，它向其他的客户机（Client2、

Client3）和非客户机（RR2）反射，同时将本地 Cluster_ID 添加到 Cluster_List 前面。

- RR2 接收到该反射路由后，检查 Cluster_List，发现自己的 Cluster_ID 已经包含在 Cluster_List 中。因此，它丢弃该更新路由，不再向自己的客户机反射。

路由反射器 - 同级反射器

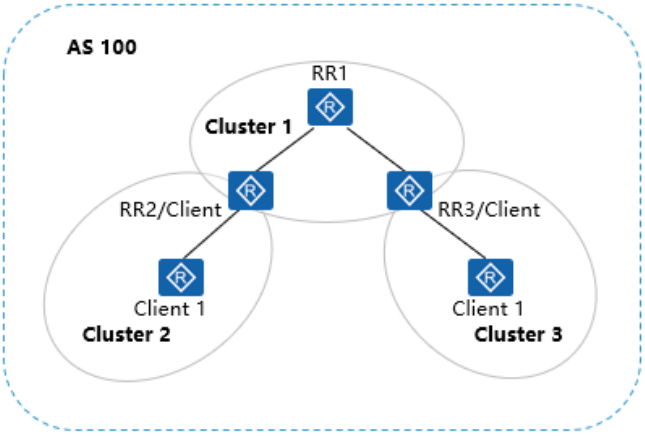
- 同级反射器
  - 一个AS中可能存在多个集群，各个RR之间是IBGP对等体的关系。



- 一个骨干网被分成多个反射集群，每个 RR 将其它的 RR 配置成非客户机，各 RR 之间建立全连接。每个客户机只与所在集群的 RR 建立 IBGP 连接。这样该自治系统内的所有 BGP 路由器都会收到反射路由信息。

路由反射器 - 分级反射器

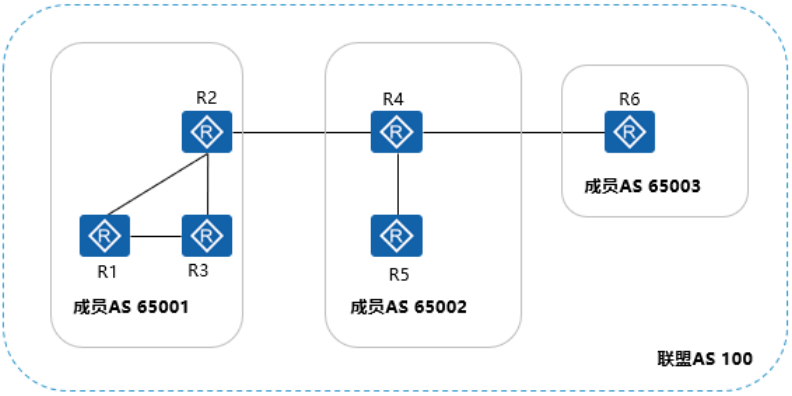
- 分级反射器
  - 将较低网络层次的RR配成更高网络层次中RR的客户机。



- Cluster1 中部署了一个一级 RR（RR-1），Cluster2 和 Cluster3 中的 RR(RR-2 和 RR-3)作为 RR-1 的客户端。

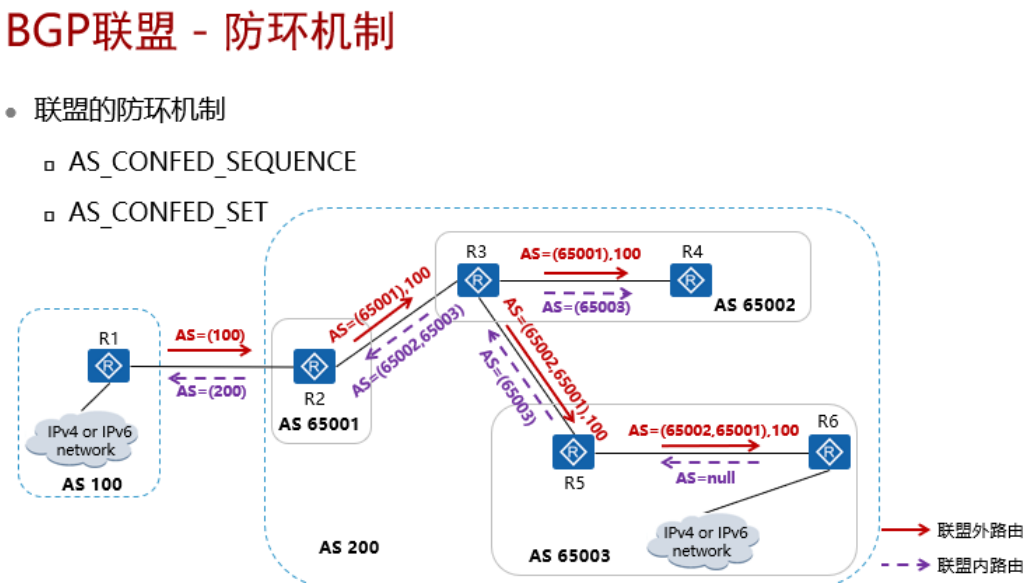
BGP联盟

- 将一个AS划分为若干个子AS，每个子AS内部建立全连接的IBGP邻居，子AS之间建立EBGP连接关系。



- 联盟
- 联盟将一个 AS 划分为若干个子 AS。每个子 AS 内部建立 IBGP 全连接关系，子 AS 之间建立联盟 EBGP 连接关系，但联盟外部 AS 仍认为联盟是一个 AS。

- 配置联盟后，原 AS 号将作为每个路由器的联盟 ID。
- 原有的 IBGP 属性，包括 Local Preference 属性、MED 属性和 NEXT_HOP 属性等；联盟相关的属性在传出联盟时会自动被删除，即管理员无需在联盟的出口处配置过滤子 AS 号等信息的操作。



- AS_PATH 属性被定义为公认必遵属性，该属性由 AS 号所组成。AS_PATH 包含 4 种不同类型
- AS_SET: 由一系列 AS 号无序地组成，包含在 UPDATE 消息里。当网络发生聚合时，可通过适当策略使 AS_PATH 使用类型 AS_SET 来避免路径信息丢失。
- AS_SEQUENCE: 由一系列 AS 号顺序地组成，包含在 UPDATE 消息里。一般情况下，AS_PATH 类型为 AS_SEQUENCE。
- AS_CONFED_SEQUENCE: 在本地联盟内由一系列成员 AS 号按顺序地组成，包含在 UPDATE 消息中，用法和 AS_SEQUENCE 相同，只能在本地联盟内传递。
- AS_CONFED_SET: 在本地联盟内由一系列成员 AS 无序地组成，包含在 UPDATE 消息中，用法和 AS_SET 相同，

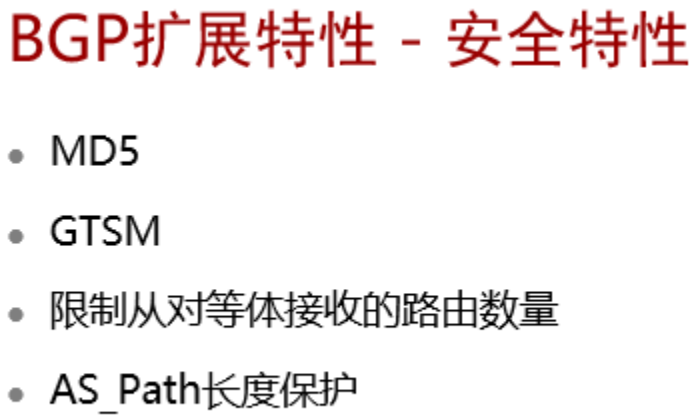
只能在本地联盟内传递。

- 联盟内部的成员 AS 号对于其他非联盟 AS 是不可见的，所以路由在从联盟内部发送到其他非联盟 AS 时，联盟成员 AS 号被剥离。

BGP路由反射器和联盟的比较

反射器	联盟
不需要更改现有的网络拓扑，兼容性好	需要修改逻辑拓扑
配置方便，客户机不知道自己是客户机	所有设备需要重新进行配置，且所有设备必须支持联盟功能
集群与集群之间仍然需要全连接	联盟的子AS之间是特殊的EBGP连接，不需要全连接
在大型网络中应用广泛	应用较少

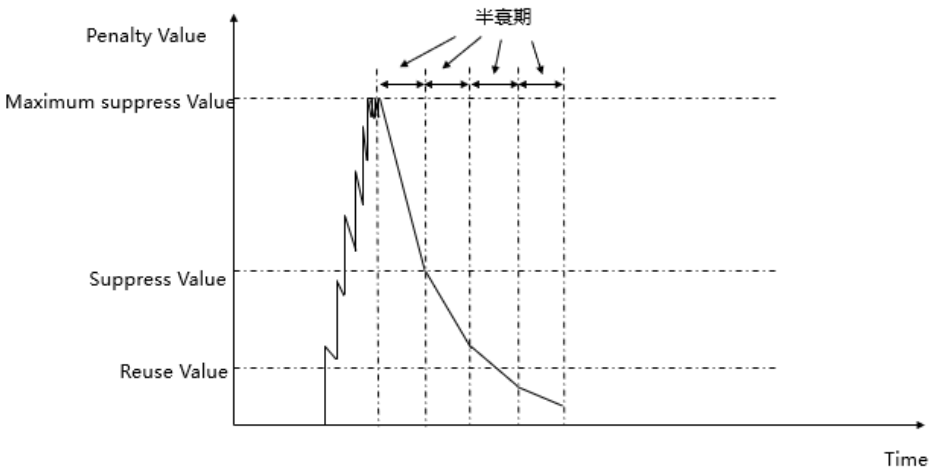
- 反射器和联盟的比较
- 联盟需要重新划分区域，对现网改动较大。
- 反射器在配置时，只需要对 RR 进行配置，客户机不需要做任何其他的操作；联盟需要在所有路由器上进行配置。
- RR 与 RR 间需要 IBGP 全互联。
- 路由反射器应用较为广泛；联盟应用较少。



- BGP 安全特性：
- MD5：BGP 使用 TCP 作为传输层协议，为提高 BGP 的安全性，可以在建立 TCP 连接时进行 MD5 认证。但 BGP 的 MD5 认证并不能对 BGP 报文认证，它只是为 TCP 连接设置 MD5 认证密码，由 TCP 完成认证。如果认证失败，则不建立 TCP 连接。
- GTSM（Generalized TTL Security Mechanism）：GTSM 通过检测 IP 报文头中的 TTL 值是否在一个预先定义好的特定范围内，对 IP 层以上业务进行保护，增强系统的安全性。使能 BGP 的 GTSM 策略后，接口板对所有 BGP 报文的 TTL 值进行检查。根据实际组网的需要，对于不符合 TTL 值范围的报文，GTSM 可以设置为通过或丢弃。配置 GTSM 缺省动作为丢弃时，可以根据网络拓扑选择合适的 TTL 有限值范围，不符合 TTL 值范围的报文会被接口板直接丢弃，这样就避免了网络攻击者模拟的“合法”BGP 报文占用 CPU。该功能与 EBGP 多跳互斥。
- 限制从对等体接收的路由数量，防止资源耗尽性攻击。
- AS_Path 长度保护。通过在入口和出口两个方向对 AS_Path 的长度进行限定，直接丢弃 AS_Path 超限的报文。

BGP扩展特性 - 路由衰减

- 路由衰减用来解决路由不稳定的问题



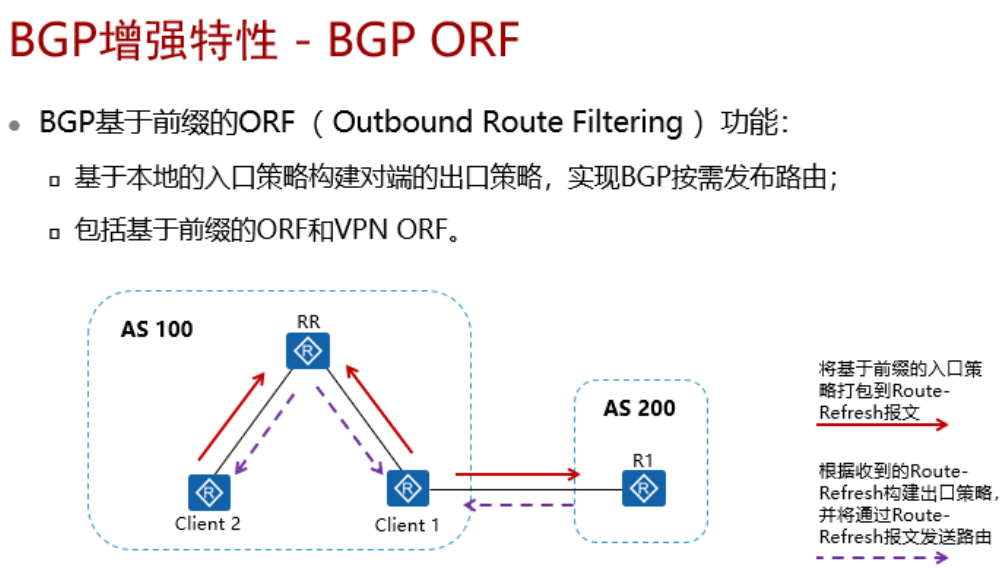
- 路由衰减（Route Dampening）用来解决路由不稳定的问题。多数情况下，BGP 协议都应用于复杂的网络环境中，路由变化十分频繁。为了防止持续的路由振荡带来的不利影响，BGP 使用路由衰减来抑制不稳定的路由。

- BGP 衰减使用惩罚值（Penalty Value）来衡量一条路由的稳定性，惩罚值越高则说明路由越不稳定。路由每发生一次振荡（路由从激活状态变为未激活状态，称为一次路由振荡），BGP 便会给此路由增加一定的惩罚值（1000）。当惩罚值超过抑制阈值（Suppress Value）时，此路由被抑制，不加入到路由表中，也不再向其他 BGP 对等体发布更新报文。
- 当某条路由的惩罚值到达最大抑制值（Maximum Suppress Value）,便不会再增加，这样就可以确保某路由在非常短的时间内翻动十几次之后，不会将惩罚值累加到一个很高的、使路由始终保持被抑制状态的值。

- 被抑制的路由每经过一段时间，惩罚值便会减少一半，

这个时间称为半衰期（Half-life）。当惩罚值降到再使用阈值（Reuse Value）时，此路由变为可用并被加入到路由表中，同时向其他 BGP 对等体发布更新报文。上文提到的惩罚值、抑制阈值和半衰期都可以手动配置。

- 路由衰减只适用于 EBGP 路由。对于从 IBGP 收来的路由不能进行衰减，因为 IBGP 路由经常含有本 AS 的路由，内部网络路由要求转发表尽可能一致。如果衰减对 IBGP 路由起作用，不同设备的衰减参数不一致时，会导致转发表不一致。



- RFC5291、RFC5292 规定了 BGP 基于前缀的 ORF 能力，能将本端设备配置的基于前缀的入口策略通过路由刷新报文发送给 BGP 邻居。BGP 邻居根据这些策略构造出口策略，在路由发送时对路由条目进行过滤。这样不仅避免了本端设备接收大量无用的路由，降低了本端设备的 CPU 使用率，还有效减少了 BGP 邻居的配置工作，降低了链路带宽的占用率。

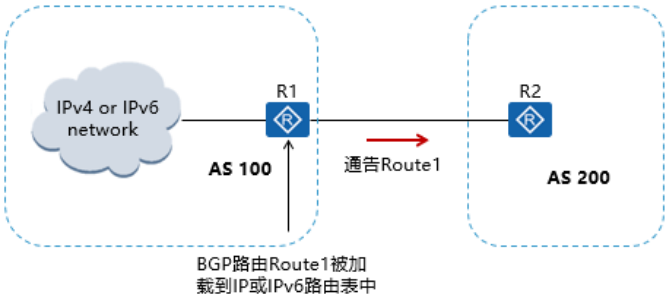
- 拓扑描述
- 直连 EBGP 邻居中，Client1、R1 协商基于前缀的 ORF

能力后，Client1 将本地配置的基于前缀的入口策略打包到 Route-refresh 报文中发送给 R1。R1 根据接收到的路由刷新报文构造出口策略，通过 Route-refresh 报文发送路由给 Client1。Client1 只收到它需要的路由，而 R1 不必维护路由策略，减少了配置工作。

- Client1、Client2 为 RR 的客户端，Client1 与 RR、Client2 与 RR，分别协商基于前缀的 ORF 能力，Client1、Client2 将本地配置的基于前缀的入口策略打包到 Route-refresh 报文中发送给 RR。RR 根据接收到的 Client1、Client2 基于前缀的入口策略，构造 RR 的出口策略，将路由反射给 Client1、Client2。Client1 和 Client2 只收到需要的路由，RR 不必维护路由策略，减少了配置工作。

BGP增强特性 - Active-Route-Advertise

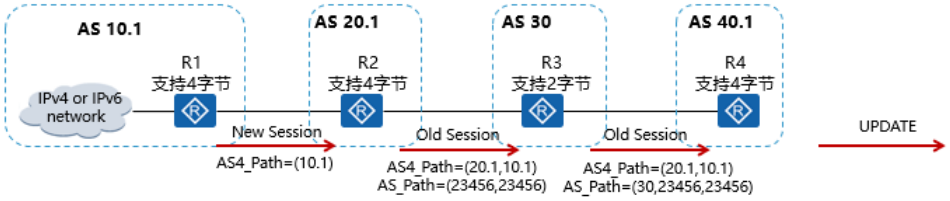
- Active-Route-Advertise
  - 只有当BGP路由被成功的安装进IP或IPv6路由表，该路由才能被发送给邻居。



- Active-Route-Advertise
- 默认情况下路由只需在 BGP 中优选即可向邻居发布。配置了此特性之后，路由必须同时满足在 BGP 协议层面优选与在路由管理层面活跃两个条件，才能向邻居发布。
- 与命令 bgp-rib-only（用来禁止 BGP 路由下发到 IP 路由表）互斥。

BGP增强特性 - 4字节AS号概念

- 4字节AS号
  - 4字节AS号特性是将AS号的编码范围由2字节扩大为4字节。
- 协议扩展
  - 定义了一种新的Open能力码用于进行BGP连接的能力协商；
  - 2种新的可选过渡属性，AS4_Path和AS4_Aggregator属性；
  - 定义AS_TRANS（保留值为23456）用于衔接2字节AS和4字节AS。

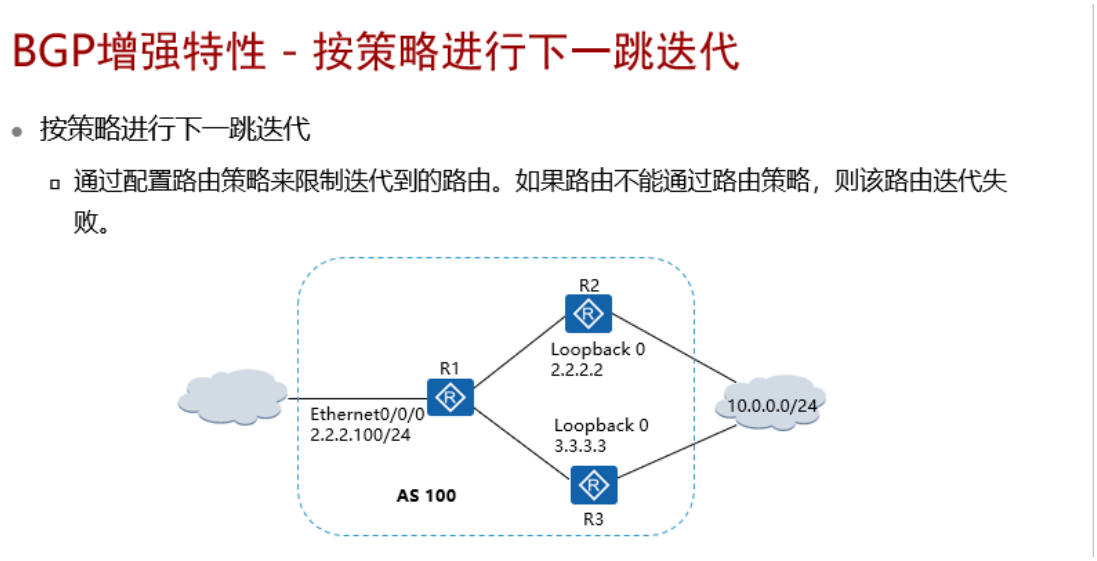


- 4 字节 AS 号定义的角色
- New Speaker：支持 4 字节 AS 号扩展能力的对等体。
- Old Speaker：不支持 4 字节 AS 号扩展能力的对等体。
- New Session：New Speaker 之间建立的 BGP 连接。
- Old Session：New Speaker 和 Old Speaker 之间或者 Old Speaker 之间建立的 BGP 连接。

- 协议扩展
- 定义了 2 种新的可选过渡属性 AS4_Path（属性码为 0x11）和 AS4_Aggregator 属性（属性码为 0x12）用于在 Old Session 上传递 4 字节 AS 信息。
- 如果 New Speaker 和 Old Speaker 建立连接，定义 AS_TRANS（保留值为 23456）用于衔接 2 字节 AS 和 4 字节 AS。
- 新的 AS 号有三种写法：
- splain：就是一个十进制的数字
- asdot+:写成（2 字节）.（2 字节）的形式，所以旧的 2 字节 ASN123 可以写成 0.123，ASN65536 是 1.0；最大为 65535.65535；

- asdot：旧的 2 字节写法照旧，新的 4 字节写成 asdot+的形式；（1 - 65535；1.0 - 65535.65535）
- 华为支持 asdot 写法。

- 拓扑描述
- R2 收到 R1 的一条四字节 AS 的路由,AS 号码为 10.1；
- R2 与 R3 建立邻居，需要令 R3 认为 R2 的 AS 号为 AS_TRANS；
- R2 发送路由给 R3 的时候把 AS_TRANS 记录在 AS_Path 里面，把 10.1 与自己的 AS 号码 20.1 按照 BGP 要求的顺序记录在 AS4_Path；
- R3 对于不识别的属性 AS4_Path 不作处理依然保留，它只按照 BGP 的规则来发送路由给 RD。当然它认为 R4 的 AS 号码也是 AS_TRANS；
- 这样当 R4 收到从 R3 来的路由会把 AS_PATH 中的 AS_TRANS 按照顺序替换为 AS4_Path 里所记录的相应的地址，在 R4 上把 AS_PATH 属性还原为 30 20.1 10.1。



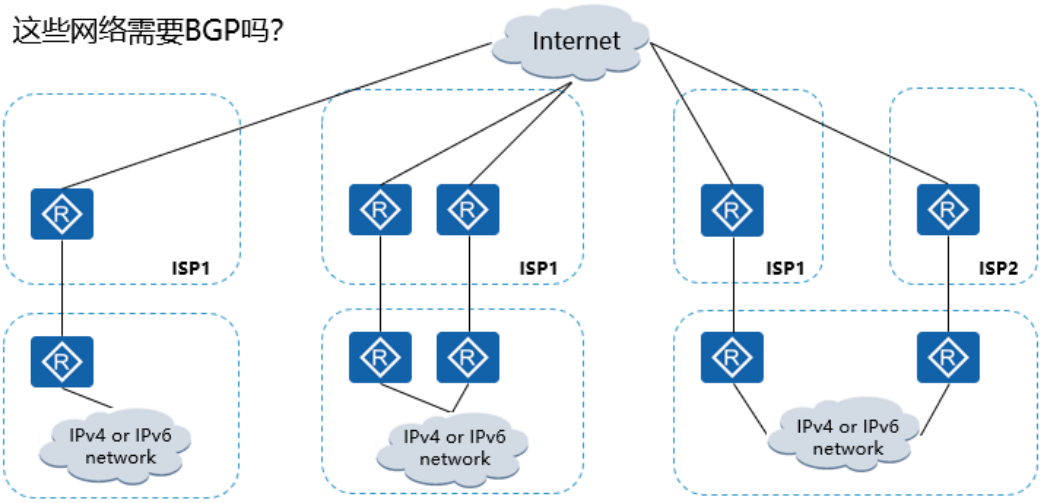
- 按策略进行下一跳迭代

- BGP 需要对非直连的下一跳进行路由迭代，但是如果不对迭代到的路由进行过滤的话，可能会迭代到一个错误的转发路径上。按策略进行下一跳迭代就是通过配置路由策略来限制迭代到的路由。如果路由不能通过路由策略，则该路由迭代失败。

- 拓扑描述
- R1 和 R2、R3 之间通过 Loopback 口建立 IBGP 邻居。R1 从 R2、R3 分别收到了前缀为 10.0.0.0/24 的 BGP 路由。其中从 R2 收到的 BGP 路由的原始下一跳为 2.2.2.2。另外，R1 上 Ethernet0/0/0 的接口地址为 2.2.2.100/24。
- 当 R2 正常运行时，R1 收到从 R2 发来的前缀为 10.0.0.0/24 的路由会迭代到 IGP 路由 2.2.2.2/32。但是当 R2 的 IGP 发生故障时，IGP 路由 2.2.2.2/32 被撤销，这样就导致下一跳重新迭代。在 R1 上会用原始下一跳 2.2.2.2 在 IP 路由表中进行最长匹配迭代，结果会迭代到 2.2.2.0/24 的路由上。但此时用户期望的是，当到 2.2.2.2 的路由不可达时，可以重新选路优选到 3.3.3.3 的路由。实际上该故障主要是由于 BGP 收敛引起的，从而产生了路由的瞬时黑洞。
- 配置下一跳迭代策略，可以通过到 BGP 路由原始下一跳所依赖路由的掩码长度来过滤迭代路由。可以通过配置下一跳迭代策略，使到原始下一跳 2.2.2.2 只能依赖于 2.2.2.2/32 的 IGP 路由。

什么样的网络需要BGP

- 这些网络需要BGP吗?



- 常见的企业网络拓扑类型
- 单归属自治系统（一个出口设备且连接到一个 ISP）
- 多归属到单自治系统（多个出口设备仅连接到一个 ISP）
- 多归属到多自治系统（多个出口设备且连接到多个 ISP）
- 单归属自治系统：仅有一个出口设备且只连接到一个 ISP。
- 这种情况下，就可以不需要配置 BGP 协议。可在用户边界设备上添加一条默认路由，并宣告到用户自治系统内部。
- 多归属到单自治系统：增加了对链路和网络设备的冗余性，一般这种情况下用户网络用的会是私有 AS 号。
- 若两条链路采用主备的方式，那么也不需要采用 BGP。两台出口设备分别向本自治系统内的设备宣告 metric 值不同的默认路由即可。（若采用 OSPF 为 IGP，外部路由的 cost 应该采用 E2 方式，仅考虑外部开销（cost））
- 若两台路由器采用负载分担方式：
- 方式一：两台路由器分别向自治系统内（IGP 采用 OSPF）宣告 cost type 为 E1 的默认路由，使得自治系统其他路由

器选择距离自己最近的出口路由器到达外部网络。这种情况也可以不使用 BGP。但是当两个出口路由器的物理间隔十分大，并且对时延有很高要求时，就可以考虑采取 BGP 来获取更精细的路由条目。

- 方式二：与 ISP 设备之间建立 BGP 连接，从 BGP 接收更为精细的路由条目，配合上路由策略工具的使用，来达到针对不同目的地址使用不同出口路由的目的。
- 多归属到多自治系统：不仅增加了对链路和网络设备的冗余性，同时使用了做到了 ISP 的冗余备份。
- 对于这种自治系统，需要充分考虑到地址空间是否独立于运营商，是否拥有公有 AS 号等问题。
- 理想情况下，当用户网络拥有独立于 ISP 的地址空间和公有 AS 号时，有三种部署方式
- 方式一：采取主备方式，出口路由器向内部宣告开销不一样的默认路由。
- 方式二：负载分担方式，出口路由器向内部宣告默认路由，仅使用 IGP 的开销计算机制，由 IGP 自行决定使用哪一台出口路由器。
- 方式三：部署 BGP。考虑与 ISP 签署的合约，企业本身的业务流量特点等因素，使用各种路由策略工具，如有必要也可以使用默认路由宣告等方式。充分控制企业进和出方向的流量。
- 一般情况下，多归属到多自治系统的网络会考虑部署 BGP 协议，因为前两种方法不利于路由的控制。但是不是绝对的，需要仔细权衡所得到的好处与增加路由复杂度所带来的代价。

什么样的网络需要BGP - 几点考虑

- 路由规模庞大且无法聚合；
- 跨经营实体时（沟通不便利、信息保密）；
- 一个路由选择域内运行了多种IGP协议；
- IGP无法提供相应的工具来实施所需策略；
- 多ISP归属网络；
- .....

- 当符合以上的情况，且充分考虑工程实施，网络规划等内容之后可以考虑部署BGP。

什么样的网络需要BGP - 潜在危险

- BGP路由劫持
- 不对称路由
- .....

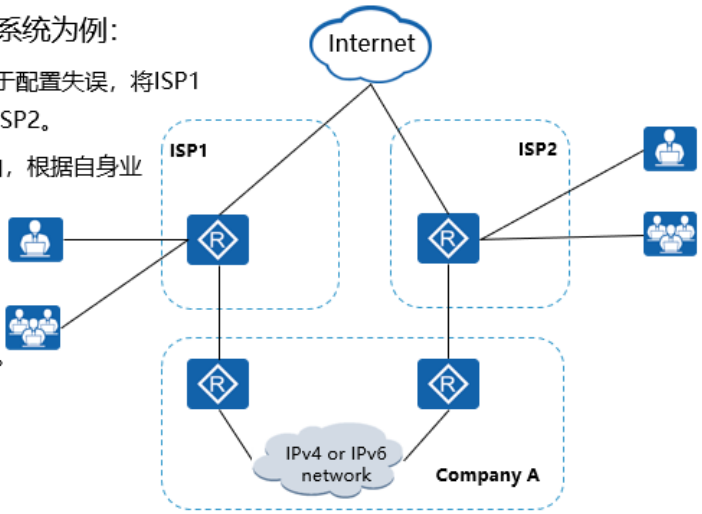
- BGP 路由劫持
- 产生原因：BGP 协议里虽然有一些简单的安全认证的部分，但是对于两个已经成功建立 BGP 连接的 AS 来说，基本会无条件的相信对方 AS 所传来的信息，包括对方声称所拥有的 IP 地址范围。
- 潜在危害：若无条件相信对方发送过来的 Update 消息，不排除恶意的 AS 宣告不存在的 IP 网段，通过修改 AS_Path 等 BGP 属性，让其他 AS 认为这条路径才是到达这个目的网段的最短路径，那么该恶意的 AS 就能截获到数据流量。

- 不对称路由
- 产生原因：不恰当的属性使用或者是路由聚合不合理导致路由精准性不足，导致流量的出方向和入方向不同。

- 潜在危害：首先，不对称流量会使互联网络的流量模型变得难以预测，使得网络基准、容量规划、故障检测及排除变得困难；其次，不对称流量会使链路使用率出现不均衡，某些链路的带宽出现饱和，而其他链路的带宽却得不到有效利用；再次，不对称流量会使出流量和入流量的时延出现很大的差异，这种时延变化（即抖动）会对某些时延敏感型应用（如语音和直播视频）造成损害。

BGP网络部署及维护注意事项

- 以一个多ISP归属自治系统为例：
  - 公司A连接两个AS，由于配置失误，将ISP1宣告的路由宣告进入了ISP2。
  - 注意筛选ISP发布的路由，根据自身业务情况进行路由选择。
  - 时刻关注对端AS对网络的操作，防止因对端误操作而对本网造成危害。



Internet设计理念 - 优化BGP能力

- 优化BGP能力
  - 建立对等体会话
  - 路由更新起源
  - 优化路由策略
  - 路由过滤和属性控制
  - 路由聚合

Internet设计理念 - 提高BGP可用性

- 提高BGP可用性
  - 冗余
  - 流量对称
  - 负载均衡

Internet设计理念 - 控制AS内部路由

- 控制AS内部路由
  - 非BGP路由与BGP路由之间的交互
  - 默认路由
  - 策略路由

- 非 BGP 路由和 BGP 路由之间的交互
- 一般情况下，IGP 与 BGP 会有路由的引入。应采用合理的过滤策略，使合适的路由在 IGP 与 BGP 之间互相引入。
- 默认路由的控制
- 对于默认路由的发放，可以通过策略使默认路由根据某些具体条件来下发默认路由。
- 策略路由
- 通过策略路由来优化流量路径。

Internet设计理念 - 控制大型AS

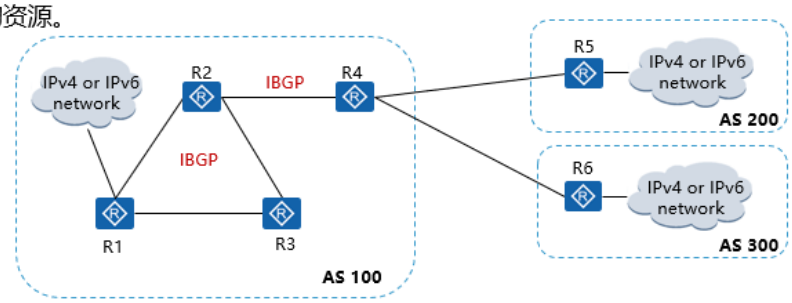
- 控制大型AS
  - BGP按组打包
  - 路由反射器
  - 联盟

Internet设计理念 - 设计稳定的Internet

- 设计稳定的Internet
  - 减少不稳定路由的产生
  - 提升BGP稳定性

配置对等体组 (1)

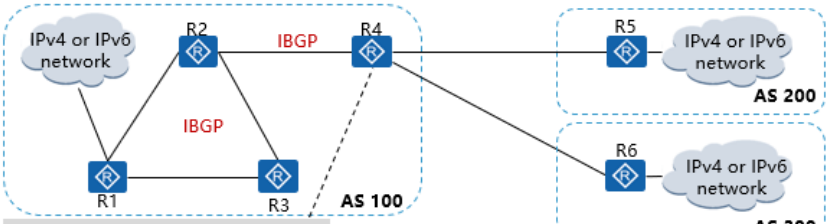
- 配置需求：
  - 底层IGP协议已经部署完成，所有设备均处于OSPF骨干区域中；
  - R4、R5、R6之间使用loopback接口建立EBGP邻居关系，要求R4的配置方式能够尽量节约资源。



- 案例描述
- 本案例中设备互联地址规则如下：
- 如 RX 与 RY 互联，则互联地址为 XY.1.1.X 与 XY.1.1.Y，掩码长度为 24 位。

- OSPF 和 OSPFv3 已经运行正常，且设备互联地址和环回口地址已经宣告进了 OSPF 或 OSPFv3。
- 案例分析
- EBGP 邻居之间使用环回口建立邻居关系。

配置对等体组 (2)




```
bgp 100
group 100 external
 peer 100 ebgp-max-hop 2
 peer 100 connect-interface LoopBack0
 peer 5.5.5.5 as-number 200
 peer 5.5.5.5 group 100
 peer 6.6.6.6 as-number 300
 peer 6.6.6.6 group 100
#
ipv4-family unicast
 undo synchronization
 peer 100 enable
 peer 5.5.5.5 enable
 peer 5.5.5.5 group 100
 peer 6.6.6.6 enable
 peer 6.6.6.6 group 100
```

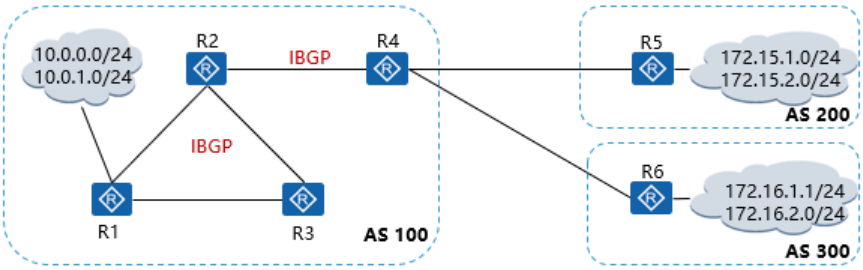
```
[R4]display bgp peer
BGP local router ID : 4.4.4.4
 Local AS number : 100
 Total number of peers : 2                 Peers in established state : 2
  Peer      V   AS  MsgRcvd  MsgSent  OutQ  Up/Down       State PrefRcv
  5.5.5.5   4   200       3        3     0  00:01:06 Established     0
  6.6.6.6   4   300       2        2     0  00:00:50 Established     0
```

- 命令含义
- peer as-number 命令用来配置指定对等体（组）的对端 AS 号。
- peer connect-interface 命令用来指定发送 BGP 报文的源接口，并可指定发起连接时使用的源地址。
- peer next-hop-local 命令用来设置向 IBGP 对等体（组）通告路由时，把下一跳属性设为自身的 IP 地址。
- group 命令可以用来创建对等体组。
- 具体用法
- 上述命令均为 BGP 进程视图下的命令
- 参数意义
- peer ipv4-address as-number as-number
- ip-address：对等体的 IPv4 地址。
- as-number：对等体的对端 AS 号。

- peer ipv4-address connect-interface interface-type interface-number [ipv4-source-address]
- ip-address：对等体的 IPv4 地址。
- interface-type interface-number：接口类型和接口号。
- ipv4-source-address：建立连接时的 IPv4 源地址。
- peer ipv4-address next-hop-local
- ip-address：对等体的 IPv4 地址。
- group group-name [external | internal]
- group-name:对等体组的名称。
- external:创建 EBGP 对等体组。
- internal:创建 IBGP 对等体组。
- 注意事项
- 在使用 Loopback 接口作为 BGP 报文的源接口时，必须注意以下事项：
- 确认 BGP 对等体的 Loopback 接口的地址是可达的。
- 如果是 EBGP 连接，还要配置 peer ebgp-max-hop 命令，允许 EBGP 通过非直连方式建立邻居关系。
- peer next-hop-local 和 peer next-hop-invariable 是两条互斥命令。
- Display bgp peer 中的 Rec 表示本端从对等体上收到路由前缀的数目。
- IPv6 的配置与 IPv4 配置一致。

配置BGP自动聚合和缺省路由

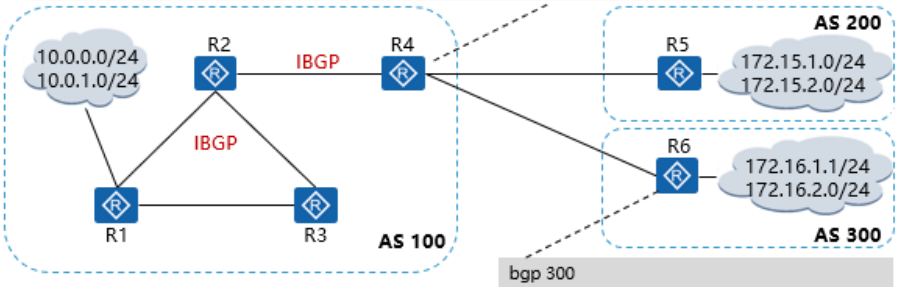
- 对AS 300进行路径优化，需求如下：
  - 将网络172.16.X.0/24引入到BGP中，并做自动汇总；
  - 当R4上存在172.16.0.0/16路由的时候，需要往AS300下发缺省路由。



- 案例描述
- 本案例中的需求是对之前案例的扩展，在原案例的基础上进行配置。
- 需求 2 中，要求缺省路由的下发需要关联路由 172.16.0.0/16，如果 172.16.0.0/16 消失，该缺省路由也消失。

配置BGP自动聚合和缺省路由

```
bgp 100
 ipv4-family unicast
  peer 6.6.6.6 default-route-advertise conditional-route-match-all 172.16.0.0 255.255.0.0
```




```
[R6]display bgp routing-table
     Network         NextHop    MED    LocPrf    PrefVal Path/Ogn
 *>  0.0.0.0         4.4.4.4    0                0       100i
 *>  10.0.0.0/24     4.4.4.4                     0       100?
 *>  172.16.0.0      127.0.0.1                   0       ?
 s>  172.16.1.0/24   0.0.0.0    0                0       ?
 s>  172.16.2.0/24   0.0.0.0    0                0       ?
```

```
bgp 300
 #
 ipv4-family unicast
  summary automatic
  import-route direct route-policy DIRECT
#
route-policy DIRECT permit node 10
 if-match ip-prefix 10
#
ip ip-prefix 10 index 10 permit 172.16.1.0 24
ip ip-prefix 10 index 20 permit 172.16.2.0 24
```

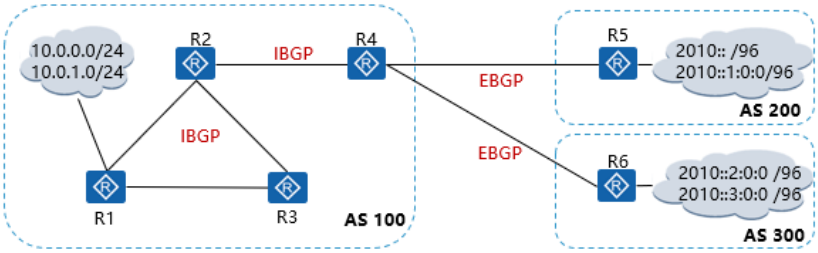
- 命令含义

- peer route-policy 命令用来对来自对等体（组）的路由或向对等体（组）发布的路由指定 Route-Policy，对接收或发布的路由进行控制。
- peer default-route-advertise 命令用来设置给对等体（组）发布缺省路由。
- 具体用法
- peer route-policy 命令为 BGP 视图命令
- peer default-route-advertise 命令 BGP 视图命令
- 参数意义
- peer ipv4-address route-policy route-policy-name { import | export }
- ipv4-address：对等体的 IPv4 地址。
- route-policy-name：Route-Policy 的名称。
- import：对从对等体（组）来的路由应用 Route-Policy。
- export：对向对等体（组）发布的路由应用 Route-Policy。
- peer { group-name | ipv4-address } default-route-advertise [route-policy route-policy-name] [conditional-route-match-all{ ipv4-address1 { mask1 | mask-length1 } } &<1-4> | conditional-route-match-any { ipv4-address2 { mask2 | mask-length2 } } &<1-4>]
- ipv4-address：对等体的 IPv4 地址。
- route-policy route-policy-name：指定 Route-Policy 名称。
- conditional-route-match-all ipv4-address1{ mask1 | mask-length1 }：指定条件路由的 IPv4 地址，以及掩码/掩码长度。匹配所有条件路由则发送缺省路由。
- conditional-route-match-any ipv4-address2{ mask2 | mask-length2 }：指定条件路由的 IPv4 地址，以及掩码/掩码长度。匹配任一条件路由则发送缺省路由。
- 实验现象
- 我们通过命令 display ip routing-table 命令用来显示路由

表中包含的信息。

配置BGP手动聚合 (1)

- 配置需求:
  - 在R1上引入网络10.0.X.0/24 同时将其community标记为200:200;
  - 在R1上对community标记200:200对路由进行手工聚合，聚合后路由掩码为/23，并抑制明细路由，需充分考虑环路避免。
  - 在R4上对2010::x:0:0/96网段进行聚合，聚合后掩码为94。



- 案例描述
- 本案例中的需求是对之前案例的扩展，在原案例的基础上进行配置。

配置BGP手动聚合 (2)

```
[R1]
bgp 100
 ipv4-family unicast
  aggregate 10.0.0.0 255.255.254.0 as-set detail-suppressed origin-policy COMM
  import-route direct route-policy DIRECT
#
route-policy DIRECT permit node 10
 if-match ip-prefix 10
 apply community 200:200
#
route-policy COMM permit node 10
 if-match community-filter 200:200
#
ip ip-prefix 10 index 10 permit 10.0.1.0 24
ip ip-prefix 10 index 20 permit 10.0.0.0 24
```

```
[R4]
bgp 100
#
 ipv6-family unicast
 aggregate 2010:: 94 detail-suppressed
 peer 2011::24:2 next-hop-local
#
```

```
[R4]dis bgp routing-table

 Total Number of Routes: 5
     Network        NextHop        MED        LocPrf    PrefVal Path/Ogn

*>   10.0.0.0/23     1.1.1.1         0                    0                ?
```

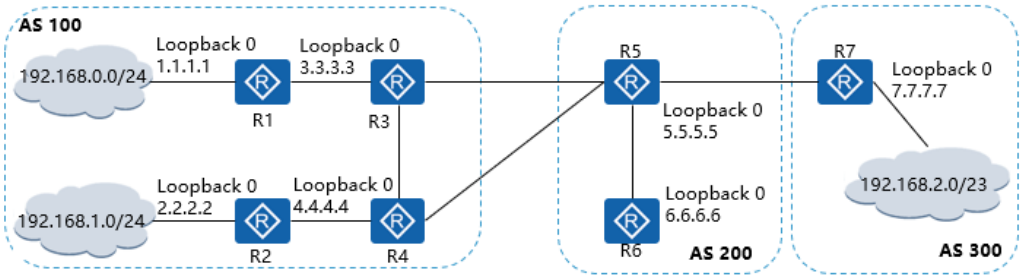
```
[R2]dis ipv6 routing-table protocol bgp
 Destination  : 2010::                PrefixLength : 94
 NextHop      : 2011::24:4            Preference   : 255
 Cost         : 0                     Protocol     : IBGP
 RelayNextHop : ::                    TunnelID     : 0x0
 Interface    : GigabitEthernet0/0/0  Flags        : D
```

- 命令含义：
- aggregate 命令用来在 BGP 路由表中创建一条聚合路由。

- 具体用法
- 命令 aggregate 为 BGP 视图命令。
- 参数意义
- aggregate ipv4-address { mask | mask-length } [as-set | attribute-policy route-policy-name1 | detail-suppressed | origin-policy route-policy-name2 | suppress-policyroute-policy-name3] *
- ipv4-address：指定聚合路由的 IPv4 地址。
- mask：仅 IBGP 路由参与负载分担。
- mask-length：指定聚合路由的网络掩码长度。
- as-set：指定生成具有 AS-SET 的路由。
- attribute-policy route-policy-name1：指定聚合后路由的属性策略名称。
- detail-suppressed：指定仅通告聚合路由。
- origin-policy route-policy-name2：指定允许生成聚合路由的策略名称。
- suppress-policy route-policy-name3：指定抑制指定路由通告的策略名称。
- 注意事项
- 手工聚合和自动聚合本地均会产生指向 NULL0 的路由。
- IPv6 的配置与 IPv4 配置一致。
- 实验结果
- 命令 display ip routing-table protocol bgp 可以查看 BGP 学到的路由。

BGP需求实现及故障排除 - 案例一 (1)

- 需求：
  - AS100内运行OSPF，所有路由器均属于AREA 0 ，直连路由器之间用LOOPBACK 0口建立IBGP，同时尽量减少BGP连接数；
  - R5和R6用LOOPBACK0口建立IBGP关系，R5与R3/R4用直连口建立EBGP邻居关系，用LOOPBACK 0 口和R7建立EBGP邻居关系；
  - R1直连192.168.0.0/24，R2直连192.168.1.0/24，R7直连192.168.2.0/23，要求这三个网段能互访，且R6上的路由条目尽可能的少。



- 路由器 Rx 和 Ry（X<Y）之间的接口网段为 10.0.xy.0/24 网段，Rx 地址=10.0.xy.x　Ry 地址=10.0.xy.y 。
- 所有接口地址均已经配置完成。

BGP需求实现及故障排除 - 案例一 (2)

- AS 100内设备的配置：

```
[R1]
ospf 1
 area 0.0.0.0
  network 1.1.1.1 0.0.0.0
  network 10.0.12.1 0.0.0.0
#
bgp 100
 peer 2.2.2.2 as-number 100
 peer 2.2.2.2 connect-interface LoopBack0
 #
 ipv4-family unicast
 network 192.168.0.0
  peer 2.2.2.2 enable
```

```
[R2]
ospf 1
 area 0.0.0.0
  network 2.2.2.2 0.0.0.0
  network 10.0.12.2 0.0.0.0
  network 10.0.23.2 0.0.0.0
#
bgp 100
 peer 1.1.1.1 as-number 100
 peer 1.1.1.1 connect-interface
LoopBack0
 peer 3.3.3.3 as-number 100
 peer 3.3.3.3 connect-interface
LoopBack0
 peer 10.0.25.5 as-number 200
 #
 ipv4-family unicast
  undo synchronization
  peer 1.1.1.1 enable
  peer 1.1.1.1 reflect-client
  peer 1.1.1.1 next-hop-local
  peer 3.3.3.3 enable
  peer 10.0.25.5 enable
#
```

```
[R3]
ospf 1
 area 0.0.0.0
  network 3.3.3.3 0.0.0.0
  network 10.0.34.3 0.0.0.0
  network 10.0.23.3 0.0.0.0
#
bgp 100
 peer 2.2.2.2 as-number 100
 peer 2.2.2.2 connect-interface
LoopBack0
 peer 4.4.4.4 as-number 100
 peer 4.4.4.4 connect-interface
LoopBack0
 peer 10.0.35.5 as-number 200
 #
 ipv4-family unicast
  undo synchronization
  peer 2.2.2.2 enable
  peer 4.4.4.4 enable
  peer 4.4.4.4 reflect-client
  peer 1.1.1.1 next-hop-local
  peer 10.0.35.5 enable
#
```

```
[R4]
ospf 1
 area 0.0.0.0
  network  0.0.0.0
  network 10.0.34.4 0.0.0.0
#
bgp 100
 peer 3.3.3.3 as-number 100
 peer 3.3.3.3 connect-interface LoopBack0
 #
 ipv4-family unicast
 network 192.168.1.0
  peer 3.3.3.3 enable
```

- 使用命令 display bgp peer 可以查看是否已经建立 BGP 邻居关系。
- 使用命令 display bgp routing-table 可以查看是否已经获取到路由信息。

BGP需求实现及故障排除 - 案例一 (3)

- AS 200内设备的配置：

```
[R5]
#
bgp 200
 peer 10.0.25.2 as-number 100
 peer 10.0.35.3 as-number 100
 peer 10.0.56.6 as-number 200
peer 7.7.7.7 as-number 300
peer 7.7.7.7 ebgp-max-hop 2
 #
 ipv4-family unicast
  undo synchronization
  aggregate 192.168.0.0 255.255.252.0 as-set detail-suppressed
  peer 10.0.25.2 enable
  peer 10.0.35.3 enable
  peer 10.0.56.6 enable
  peer 7.7.7.7 enable
#
ip route-static 7.7.7.7 255.255.255.255 10.0.57.7
```

```
[R6]
#
bgp 200
peer 10.0.56.6 as-number 200
#
 ipv4-family unicast
  undo synchronization
  peer 10.0.56.5 enable
#
```

- AS 300内设备的配置：

```
[R7]
#
bgp 300
peer 5.5.5.5 as-number 200
#
 ipv4-family unicast
  undo synchronization
  peer 5.5.5.5enable
#
ip route-static 5.5.5.5 255.255.255.255 10.0.57.5
```

- 路由器 Rx 和 Ry（X<Y）之间的接口网段为 10.0.xy.0/24 网段，Rx 地址=10.0.xy.x　Ry 地址=10.0.xy.y 。
- 所有接口地址均已经配置完成。

BGP需求实现及故障排除 - 案例一 (4)

- 配置完成之后，所有BGP邻居建立正常，但是AS100中的两个网段无法访问AS300中的网段。
  - 查看R1和R3的路由表：

```
<R1>dis bgp routing-table
Network          NextHop      MED      LocPrf    PrefVal Path/Ogn

 *>  192.168.0.0      0.0.0.0        0                 0     i
 *>i 192.168.1.0      4.4.4.4        0        100      0     i

[R3-bgp]dis bgp routing-table
Network          NextHop      MED      LocPrf    PrefVal Path/Ogn

 *>i 192.168.0.0      1.1.1.1        0        100      0     i
 *>i 192.168.1.0      4.4.4.4        0        100      0     i
```

```
[R5]dis bgp routing-table

Total Number of Routes: 6
     Network          NextHop      MED      LocPrf    PrefVal Path/Ogn

 *>  192.168.0.0/22   127.0.0.1                         0       {100 300}i
 s>  192.168.0.0      10.0.25.2                         0       100i
 *                    10.0.35.3                         0       100i
 s>  192.168.1.0      10.0.25.2                         0       100i
 *                    10.0.35.3                         0       100i
 s>  192.168.2.0/23   10.0.56.6      0                  0       300i
```

- 可以看到，因为 AS_Set 中包含本自治系统的 AS 号，所以导致无法接受这条聚合后的路由，可以考虑关闭明细抑制或取消 AS_Set。

BGP需求实现及故障排除 - 案例一 (5)

- 修改后的R5配置：

```
#
bgp 200
#
 ipv4-family unicast
  undo synchronization
  aggregate 192.168.0.0 255.255.252.0 detail-suppressed
```

- 查看R1和R3的路由表：

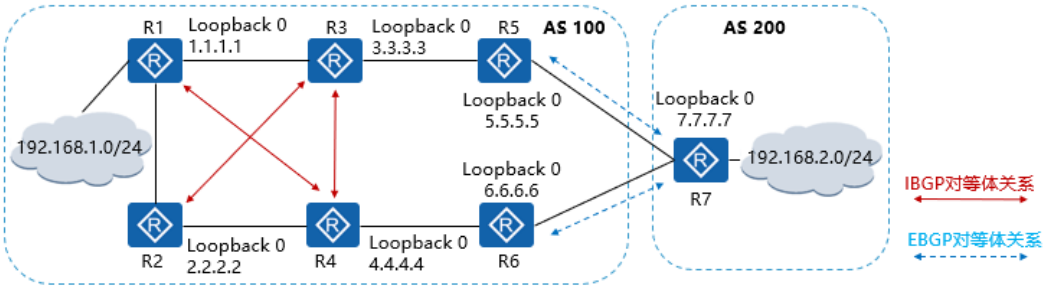
```
<R1>dis bgp routing-table
      Network         NextHop      MED    LocPrf    PrefVal     Path/Ogn
*>i   192.168.0.0/22  2.2.2.2                100       0         200i
 *>   192.168.0.0     0.0.0.0      0                   0          i
 *>i  192.168.1.0     4.4.4.4      0         100       0          i

[R3-bgp]dis bgp routing-table
      Network         NextHop      MED    LocPrf    PrefVal     Path/Ogn
*>i   192.168.0.0/22  10.0.35.5              100       0         200i
 *>i  192.168.0.0     1.1.1.1      0         100       0          i
 *>i  192.168.1.0     4.4.4.4      0         100       0          i
```

- 案例总结：
- 配置路由聚合时需要谨慎，聚合配置不当会出现以下问题：
- 无法学习到正确的路由；
- 可能导致环路的产生。

BGP需求实现及故障排除 - 案例二 (1)

- 需求：
  - AS100内运行OSPF，所有路由器均属于AREA 0，路由器之间用loopback0口建立IBGP邻居关系，对等体关系如图中标识
  - R7和R5及R6通过直连接口建立EBGP邻居关系。
  - 为了所有路由器能学习到AS200的路由器，设置RR反射器，R5为R3的client，R6为R4的client



- 路由器 Rx 和 Ry（X<Y）之间的接口网段为 10.0.xy.0/24 网段，Rx 地址=10.0.xy.x　Ry 地址=10.0.xy.y 。

- 所有接口地址均已经配置完成。
- R5 是 R3 的 client ， R6 是 R4 的 client 。

BGP需求实现及故障排除 - 案例二 (2)

- R1与R4配置:

```
[R1]
ospf 1 router-id 1.1.1.1
 area 0.0.0.0
  network 1.1.1.1 0.0.0.0
  network 10.0.12.1 0.0.0.0
  network 10.0.13.1 0.0.0.0
#
bgp 100
 peer 4.4.4.4 as-number 100
 peer 4.4.4.4 connect-interface LoopBack0
 #
 ipv4-family unicast
  undo synchronization
  network 192.168.1.0
 255.255.255.0
  peer 4.4.4.4 enable
```

```
[R4]
#
ospf 1 router-id 4.4.4.4
 area 0.0.0.0
  network 4.4.4.4 0.0.0.0
  network 10.0.24.4 0.0.0.0
  network 10.0.46.4 0.0.0.0
#
#
bgp 100
 peer 1.1.1.1 as-number 100
 peer 1.1.1.1 connect-interface LoopBack0
 peer 6.6.6.6 as-number 100
 peer 6.6.6.6 connect-interface LoopBack0
 #
 ipv4-family unicast
  undo synchronization
  peer 1.1.1.1 enable
  peer 6.6.6.6 enable
  peer 6.6.6.6 reflect-client
#
```

- 配置完成之后所有 BGP 邻居关系建立正常，且 OSPF 学习到的路由也都完整。
- R2 与 R1 配置类似，R3 与 R4 配置类似，R5 与 R6 配置类似。
- 建立完成之后 R1 通告直连的 192.168.1.0/24 进入 BGP，R7 通告直连的 192.168.2.0/24 进入 BGP。

BGP需求实现及故障排除 - 案例二 (3)

- R6与R7配置:

```
[R6]
ospf 1 router-id 6.6.6.6
 area 0.0.0.0
  network 6.6.6.6 0.0.0.0
  network 10.0.46.6 0.0.0.0
  network 10.0.67.6 0.0.0.0
#
bgp 100
 peer 4.4.4.4 as-number 100
 peer 4.4.4.4 connect-interface LoopBack0
 peer 4.4.4.4 next-hop-local
 peer 10.0.67.7 as-number 200
 #
 ipv4-family unicast
  undo synchronization
  peer 4.4.4.4 enable
  peer 10.0.67.7 enable
```

```
[R7]
#
bgp 200
 router-id 7.7.7.7
 peer 10.0.57.5 as-number 100
 peer 10.0.67.6 as-number 100
 #
 ipv4-family unicast
  undo synchronization
  peer 10.0.57.5 enable
  network 192.168.2.0 255.255.255.0
  peer 10.0.67.6 enable
#
```

- 配置完成之后所有 BGP 邻居关系建立正常，且 OSPF 学习到的路由也都完整。
- R2 与 R1 配置类似，R3 与 R4 配置类似，R5 与 R6 配置类似。
- 建立完成之后 R1 通告直连的 192.168.1.0/24 进入 BGP，R7 通告直连的 192.168.2.0/24 进入 BGP。

BGP需求实现及故障排除 - 案例二 (4)

- 查看R1与R7路由表项:

```
[R1]dis bgp routing-table

 Total Number of Routes: 2
      Network         NextHop        MED        LocPrf    PrefVal Path/Ogn
*>i  192.168.2.0     6.6.6.6          0          100       0          200i

[R1]display ip routing-table
Destination/Mask    Proto   Pre  Cost      NextHop         Interface
192.168.2.0/24      IBGP    255  0           6.6.6.6       Ethernet0/0/1
```

```
[R1]ping -a 192.168.1.1 192.168.2.1
  PING 192.168.2.1: 56  data bytes, press CTRL_C to break
    Request time out
    Request time out
    Request time out
    Request time out
    Request time out

  --- 192.168.2.1 ping statistics ---
    5 packet(s) transmitted
    0 packet(s) received
    100.00% packet loss
```

```
<R7>dis bgp routing-table

 Total Number of Routes: 2
      Network         NextHop        MED        LocPrf    PrefVal Path/Ogn

 *>   192.168.1.0     10.0.67.6                  0                    100i
 *>   192.168.2.0     0.0.0.0          0         0                    i

<R7>display ip routing-table
Destination/Mask    Proto   Pre  Cost      Flags NextHop         Interface
127.0.0.1/32        Direct  0    0           D   127.0.0.1       InLoopBack0
192.168.1.0/24      EBGP    255  0           D   10.0.67.6       Ethernet0/0/1
```

- 配置完成之后所有 BGP 邻居关系建立正常，且 OSPF 学习到的路由也都完整。

- R2 与 R1 配置类似，R3 与 R4 配置类似，R5 与 R6 配置类似。
- 建立完成之后各路由器通告自己的 loopback 0 口的地址

BGP需求实现及故障排除 - 案例二 (5)

- 查看R1与R2路由表项:

```
[R1]display ip routing-table
Destination/Mask  Proto   Pre  Cost      NextHop         Interface
192.168.2.0/24    IBGP    255  0         6.6.6.6         Ethernet0/0/1
 6.6.6.6/32       OSPF    10   3         10.0.12.2       Ethernet0/0/1
```

```
[R2]display ip routing-table
Destination/Mask  Proto   Pre  Cost      NextHop         Interface
192.168.2.0/24    IBGP    255  0         5.5.5.5         Ethernet0/0/1
 5.5.5.5/32       OSPF    10   3         10.0.12.1       Ethernet0/0/1
```

- 故障分析：
- R7 向 R5 和 R6 发送 192.168.2.0/24 前缀。
- R5,R6 收到，分别向自己的 IBGP 邻居 R3,R4 发送。
- 这里分析 R4 的情况。R4 收到后会有一个路径决策过程，这里 R3 也会向它发送 192.168.2.0/24 的前缀，根据 BGP 路径决策的 13 个原则，R4 最总选择 IGP 度量值最小的，即选择 R6 作为下一跳。然后它将这个最佳路径发往 R3 和 R1。
- 同理，R3 最总选择的下一跳是 R5。
- 关键在于 R1 和 R2。因为 R1 只能收到 R4 发来的更新，所以，它去往 192.168.1.0/24 的下一跳也是 R4；同理 R2 去往 192.168.1.0/24 的下一跳是 R5。
- 经过 IGP 路由的递归查询，从 192.168.1.1 到 192.168.2.1 的数据包会在 R1 和 R2 之间来回转发，直到 IP 报文的 TTL 被减至 0 。

BGP需求实现及故障排除 - 案例二 (6)

- 案例总结:
  - 配置RR时，应遵循以下原则:
    - 不要跨越非客户端建立客户端
    - 不要跨越客户端建立非客户端对等体
    - 客户端与非客户端之间不要建立IBGP会话

思考题

1. RR在反射从IBGP发来的路由信息时，不允许修改其中任何的属性。（ ）

   A. T

   B. F

2. 简要说明一下路由聚合的分类以及注意事项等。

- 参考答案：
- A
- 路由聚合分为自动聚合和手工聚合
- 自动聚合：只能聚合通过 import 命令引入的路由，只能按照自然掩码进行聚合，IPv6 不支持自动聚合。
- 手工聚合：IPv4 与 IPv6 路由均能聚合，可设置明细路由抑制，添加 AS_set 等功能。