BGP 协议原理与配置

BGP（Border Gateway Protocol） 边界网关协议

BGP 知识点

BGP 基础配置，BGP 5 种报文，6 种邻居状态，4 大类 细分 10 种属性，IBGP EBGP （环回口 物理接口）建立邻居，BGP 认证，fake-as ,路由传递原则，IBGP 防环，EBGP 防环，RR 防环，BGP 路由自动聚合，手工聚合（detail-suppressed，suppress-policy，attribute-policy，origin-policy），BGP 5 种 community 属性，BGP 选路，BGP 联盟，路由反射器，BGP 路由过滤，引入，下放默认路由

边界网关协议 BGP（Border Gateway Protocol）是一种实现自治系统 AS（Autonomous System）之间的路由可达，并选择最佳路由的 高级路径矢量路由协议。

早期发布的三个版本分别是 BGP-1（RFC1105）、BGP-2（RFC1163）和 BGP-3（RFC1267），1994 年开始使用 BGP-4（RFC1771），2006 年之后单播 IPv4 网络使用的版本是 BGP-4（RFC4271），其他网络（如 IPv6 等）使用的版本是 MP-BGP（RFC4760）。

MP-BGP 是对 BGP-4 进行了扩展，来达到在不同网络中应用的目的，BGP-4 原有的消息机制和路由机制并没有改变。MP-BGP 在 IPv6 单播网络上的应用称为 BGP4+，在 IPv4 组播网络上的应用称为 MBGP（Multicast BGP）。
为方便管理规模不断扩大的网络，网络被分成了不同的自治系统。1982 年，外部网关协议 EGP（Exterior Gateway Protocol）被用于实现在 AS 之间动态交换路由信息。

但是 EGP 设计得比较简单，只发布网络可达的路由信息，而不对路由信息进行优选，同时也没有考虑环路避免等问题，很快就无法满足网络管理的要求。
BGP 是为取代最初的 EGP 而设计的另一种外部网关协议。不同于最初的 EGP，BGP 能够进行路由优选、避免路由环路、更高效率的传递路由和维护大量的路由信息。

虽然 BGP 用于在 AS 之间传递路由信息，但并不是所有 AS 之间传递路由信息都需要运行 BGP。比如在数据中心上行的连入 Internet 的出口上，为了避免 Internet 海量路由对数据中心内部网络的影响，设备采用静态路由代替 BGP 与外部网络通信。

BGP 从多方面保证了网络的安全性、灵活性、稳定性、可靠性和高效性：
1、BGP 采用认证和 GTSM 的方式，保证了网络的安全性。
2、BGP 提供了丰富的路由策略，能够灵活的进行路由选路。
3、BGP 提供了路由聚合和路由衰减功能用于防止路由振荡，有效提高了网络的稳定性。
4、BGP 使用 TCP 作为其传输层协议（端口号为 179），并支持 BGP 与 BFD 联动，提高了网络的可靠性。

BGP 按照运行方式分为 EBGP（External/Exterior BGP）和 IBGP（Internal/Interior BGP）。
1、EBGP：
运行于不同 AS 之间的 BGP 称为 EBGP。为了防止 AS 间产生环路，当 BGP 设备接收 EBGP 对等体发送的路由时，会将带有本地 AS 号的路由丢弃。
2、IBGP：
运行于同一 AS 内部的 BGP 称为 IBGP。为了防止 AS 内产生

环路，BGP 设备不将从 IBGP 对等体学到的路由通告给其他 IBGP 对等体，并与所有 IBGP 对等体建立全连接。为了解决 IBGP 对等体的连接数量太多的问题，BGP 设计了路由反射器和 BGP 联盟（详情见后面）。
如果在 AS 内一台 BGP 设备收到 EBGP 邻居发送的路由后，需要通过另一台 BGP 设备将该路由传输给其他 AS，此时推荐使用 IBGP。

BGP 的路由器号（Router ID）
BGP 的 Router ID 是一个用于标识 BGP 设备的 32 位值，通常是 IPv4 地址的形式，在 BGP 会话建立时发送的 Open 报文中携带。
对等体之间建立 BGP 会话时，每个 BGP 设备都必须有唯一的 Router ID，否则对等体之间不能建立 BGP 连接。
BGP 的 Router ID 在 BGP 网络中必须是唯一的，可以采用手工配置，也可以让设备自动选取。
缺省情况下，BGP 选择设备上的 Loopback 接口的 IPv4 地址作为 BGP 的 Router ID。如果设备上没有配置 Loopback 接口，系统会选择接口中最大的 IPv4 地址作为 BGP 的 Router ID

==========
BGP 工作原理

BGP 对等体的建立、更新和删除等交互过程主要有 5 种报文、6 种状态机、4 类属性和 5 个原则。

BGP 报文
BGP 对等体间通过以下 5 种报文进行交互，其中 Keepalive 报文为周期性发送，其余报文为触发式发送：

1、Open 报文：
用于建立 BGP 对等体连接。
2、Update 报文：
用于在对等体之间交换路由信息。需要在 BGP 中 network 才会有 Update 报文
3、Notification（通告）报文：
用于中断 BGP 连接。
4、Keepalive 报文：
用于保持 BGP 连接。
5、Route-refresh（刷新）报文：
用于在改变路由策略后请求对等体重新发送路由信息。只有支持路由刷新（Route-refresh）能力的 BGP 设备会发送和响应此报文。
可以抓取到　route-refresh 报文
refresh bgp all import

重置 BGP
reset bgp all

修改计时器，默认 60 ，180
bgp 100
timer keepalive 5 hold 15

BGP 邻居建立状态：
idle:初始状态
connect:BGP 等待 TCP 连接的建立
active:TCP 连接失败，重新建立 TCP 连接
opensent：TCP 建立成功，发送 open 报文
openconfirm:收到正确的 OPEN 报文
established:BGP 邻居建立成功

1、Idle 状态是 BGP 初始状态。
在 Idle 状态下，BGP 拒绝邻居发送的连接请求。只有在收到本设备的 Start 事件后，BGP 才开始尝试和其它 BGP 对等体进行 TCP 连接，并转至 Connect（连接）状态。Start 事件是由一个操作者配置一个 BGP 过程，或者重置一个已经存在的过程或者路由器软件重置 BGP 过程引起的。
任何状态中收到 Notification（通告）报文或 TCP 拆链通知等 Error 事件后，BGP 都会转至 Idle 状态。

2、在 Connect（连接）状态下，BGP 启动连接重传定时器（Connect Retry），等待 TCP 完成连接。
如果 TCP 连接成功，那么 BGP 向对等体发送 Open 报文，并转至 OpenSent 状态。
如果 TCP 连接失败，那么 BGP 转至 Active（活跃）状态。
如果连接重传定时器超时，BGP 仍没有收到 BGP 对等体的响应，那么 BGP 继续尝试和其它 BGP 对等体进行 TCP 连接，停留在 Connect 状态。

3、在 Active 状态下，BGP 总是在试图建立 TCP 连接。
如果 TCP 连接成功，那么 BGP 向对等体发送 Open 报文，关闭连接重传定时器，并转至 OpenSent 状态。
如果 TCP 连接失败，那么 BGP 停留在 Active 状态。
如果连接重传定时器超时，BGP 仍没有收到 BGP 对等体的响应，那么 BGP 转至 Connect 状态。

4、在 OpenSent 状态下，BGP 等待对等体的 Open 报文，并对收到的 Open 报文中的 AS 号、版本号、认证码等进行检查。
如果收到的 Open 报文正确，那么 BGP 发送 Keepalive 报文，并转至 OpenConfirm 状态。

如果发现收到的 Open 报文有错误，那么 BGP 发送 Notification 报文给对等体，并转至 Idle 状态。

5、在 OpenConfirm 状态下，BGP 等待 Keepalive 或 Notification 报文。如果收到 Keepalive 报文，则转至 Established 状态，如果收到 Notification 报文，则转至 Idle 状态。

6、在 Established 状态下，BGP 可以和对等体交换 Update、Keepalive、Route-refresh 报文和 Notification 报文。
如果收到正确的 Update 或 Keepalive 报文，那么 BGP 就认为对端处于正常运行状态，将保持 BGP 连接。
如果收到错误的 Update 或 Keepalive 报文，那么 BGP 发送 Notification 报文通知对端，并转至 Idle 状态。
Route-refresh 报文不会改变 BGP 状态。
如果收到 Notification 报文，那么 BGP 转至 Idle 状态。
如果收到 TCP 拆链通知，那么 BGP 断开连接，转至 Idle 状态。

常见的 三种状态 Idle， Active，Established

BGP 对等体之间的交互原则

BGP 设备将最优路由加入 BGP 路由表，形成 BGP 路由。
BGP 设备与对等体建立邻居关系后，采取以下交互原则：

1、从 IBGP 对等体获得的 BGP 路由，BGP 设备只发布给它的 EBGP 对等体。
2、从 EBGP 对等体获得的 BGP 路由，BGP 设备发布给它所有 EBGP 和 IBGP 对等体。
3、当存在多条到达同一目的地址的有效路由时，BGP 设备只将最优路由发布给对等体。

4、路由更新时，BGP 设备只发送更新的 BGP 路由。
5、所有对等体发送的路由，BGP 设备都会接收。

BGP 属性

4 类属性， 10 种

属性名称	类别
ORIGIN	公认必须遵循
AS_PATH	公认必须遵循
NEXT_HOP	公认必须遵循
LOCAL_PREF	公认可选
ATOMIC_AGGREGATE	公认可选
AGGREGATOR	可选过渡
COMMUNITY	可选过渡
MULTI_EXIT_DISC (MED)	可选非过渡
ORIGINATOR_ID	可选非过渡
CLUSTER_LIST	可选非过渡

Atomtic_aggregate 是一个公认可选属性，它只相当于一种预警标识，而并不承载任何信息。当路由器收到一条 BGP 路由更新明发现该条路由携带 Atomtic_aggregate 属性时，它便知道这条路由可能出现了路径属性的丢失，此时该路由器把这条路由通告给其他对等体时，需保留路由的 Atomtic_aggregate 属性。另外，收到该路由更新的路由器不通将这条路由再度明细化。

Aggregator 是一个可选过渡属性，用于标记路由汇总行为发生在哪个 AS 及哪台 BGP 路由器上。

BGP 防环机制

IBGP 防环：

路由器从它的一个 BGP 对等体那里接收到的路由条目不会将该路由器再传递给其它 IBGP 对等体，这个原则被称为 BGP 水平分割

路由反射器的防环：Originator_id , Cluster_list

Originator_id 可选非过渡属性，由 RR 产生，封装在 Update 消息中，使用 router-id 值标识路由的始发者，用于防止集群内路由环路。
Cluster_list 可选非过渡属性，记录路由经过的每个集群的 Cluster_id，用来在集群间避免环路。

EBGP 防环：

当路由器从 EBGP 邻居收到 BGP 路由时，如果该路由的 AS_Path 中包含了自己的 AS 编号，则该路由将会直接丢弃。

==========

BGP 是目前 Internet 骨干网上运行的核心路由协议，也是部署最广泛的路由协议之一。在过去的几十年里，Internet 的发展日新月异，新兴应用的不断涌现，对 Internet 网络的可靠性、扩展性提出了更高的要求。作为整个 Internet 稳定运行的基础，BGP 为了适应 Internet 的发展趋势，也推出了许多高级特性。

fake-as

RTB:
bgp 2000

peer 1.1.1.1 fake-as 200

==========

配置 BGP 负载分担

在大型网路中，到达同一目的地通常会存在多条有效路由，但是 BGP 只将最优路由发布给对等体，这一特点往往会造成很多流量负载不均衡的情况。通过配置 BGP 负载分担，可以流量负载均衡，减少网络拥塞。
一般情况下，只有“BGP 选择路由的策略”所描述的前 8 个属性完全相同，BGP 路由之间才能相互等价，实现 BGP 的负载分担。

bgp 100
maximum load-balancing 2

配置完成后，查看全局 ip 路由表
同一个 BGP 路由条目，是有两个下一条

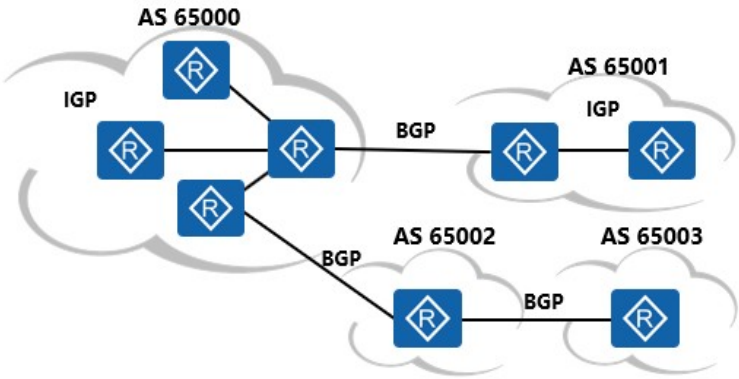
```
   44.44.44.0/24  IBGP    255  0          RD   1.1.1.1         GigabitEthernet
/0/0
                  IBGP    255  0          RD   2.2.2.2         GigabitEthernet
/0/1
   55.55.55.0/24  IBGP    255  0          RD   1.1.1.1         GigabitEthernet
/0/0
                  IBGP    255  0          RD   2.2.2.2         GigabitEthernet
/0/1
```

在 BGP 路由表中，还是只优选一个条目

前言

- 在EGP协议中，引入了AS（Autonomous System，自治系统）的概念。AS是指由同一个技术管理机构管理，使用统一选路策略的一些路由器的集合。
- AS的内部使用IGP来计算和发现路由，同一个AS内部的路由器之间是相互信任的，因此IGP的路由计算和信息泛洪完全处于开放状态，人工干预很少。
- 不同AS之间的连接需求推动了外部网关协议的发展，BGP作为一种外部网关协议，用于在AS之间进行路由控制和优选。

BGP的基本作用

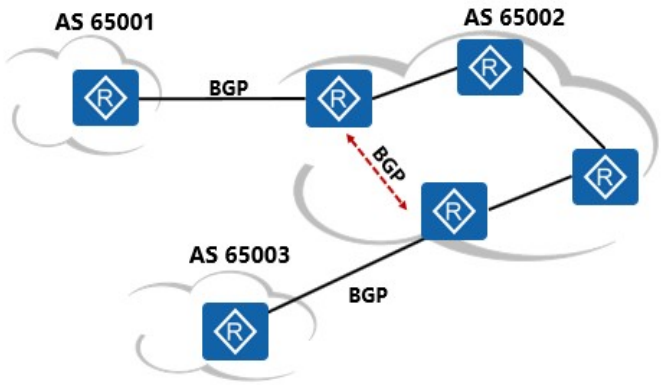


- AS内部使用IGP来计算和发现路由，如OSPF，ISIS，RIP等。
- AS之间使用BGP来传递和控制路由。

- BGP 的前身 EGP 设计得非常简单，只能在 AS 之间简单地传递路由信息，不会对路由进行任何优选，也没有考虑如何在 AS 之间避免路由环路等问题，因而 EBP 最终被 BGP 取代。
- 相比于 EGP，BGP 更具有路由协议的特征，如下：
- 邻居的发现与邻居关系的建立；

- 路由的获取，优选和通告；
- 提供路由环路避免机制，并能够高效传递路由，维护大量的路由信息；
- 在不完全信任的 AS 之间提供丰富的路由控制能力。
- 使用 BGP 作为传递路由的协议，则用户的路由域被作为一个整体和其他路由域进行路由交换，这个路由域即 AS。AS 的概念是若干台路由器以及这些路由器组成的网络集合，这些路由器均属于同一个管理机构，并执行统一的路由策略。
- 运行 BGP 协议需要一个统一的自治系统号来标识路由域，即 AS 编号。每个自治系统都有唯一的一个编号，这个编号由 IANA 分配。2009 年 1 月之前，只能使用最多 2 字节长度的 AS 号码，即 1-65535。其中 1-64511 为公有 AS，64512-65534 为私有 AS。在 2009 年 1 月之后，IANA 决定使用 4 字节长度 AS，范围是 65536-4294967295。

BGP协议特点

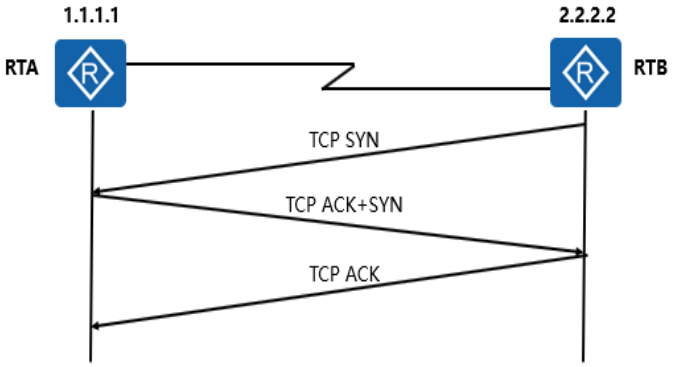


- 如图所示，BGP可以跨越多跳路由器建立邻居关系。
- 为实现路由按需求进行控制和优选，BGP设计了诸多属性携带在路由中。

- 因为是在 AS 之间传递路由，为保证数据的可靠性，BGP 使用 TCP 作为其承载协议建立连接。因此与 IGP 逐跳路由器建立邻居不同，BGP 可以跨越多跳路由器建立邻居关系。

- AS 之间的路由器是不完全相互信任的，为实现路由按需求进行控制和优选，BGP 设计了诸多属性。

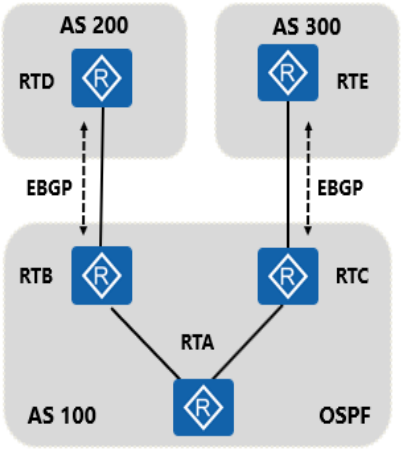
BGP邻居发现



- 先启动BGP的一端先发起TCP连接，如图所示，RTB先启动BGP协议，RTB使用随机端口号向RTA的179端口发起TCP连接。

- BGP 协议被设计运行在 AS 之间传递路由，AS 之间是广域网链路，数据包在广域网上传递是可能出现不可预测的链路拥塞或丢失等情况，因此 BGP 使用 TCP 作为其承载协议来保证可靠性。
- BGP 使用 TCP 封装建立邻居关系，端口号为 179，TCP 采用单播建立连接，因此 BGP 协议并不像 RIP 和 OSPF 一样使用组播发现邻居。单播建立连接也使 BGP 只能手动指定邻居。

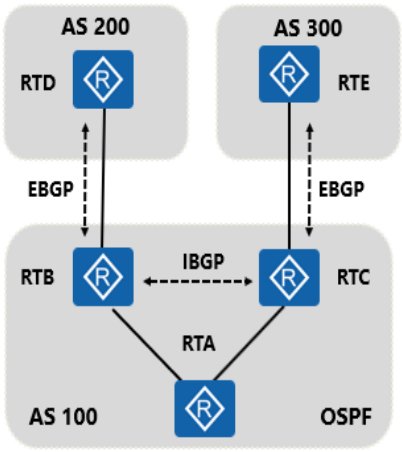
BGP邻居类型 - EBGP



- 运行在不同AS之间的BGP路由器建立的邻居关系为EBGP（External BGP）邻居关系。

- EBGP 只用于不同 AS 之间传递路由。如图，AS 100 内的 RTB 与 BTC 分别从 AS 200 与 AS 300 学习到不同的路由，怎么实现 AS 200 与 AS 300 之间路由在 AS 100 内的交换？
- 在 AS 100 内实现将学到的 AS 200 和 AS 300 路由进行交换，可以在拓扑中的 RTB 与 RTC 路由器上将 BGP 的路由引入 IGP 协议（图中为 OSPF 协议），再将 IGP 协议的路由在 RTB 与 RTC 路由器上引入回 BGP 协议，实现 AS 200 与 AS 300 路由的交换。
- 上述方法存在以下几个缺点：
- 公网上 BGP 承载的路由数目非常大，引入 IGP 协议后，IGP 协议无法承载大量的 BGP 路由；
- BGP 路由引入 IGP 协议时，需要做严格的控制，配置复杂，不易维护；
- BGP 携带的属性在引入 IGP 协议时，由于 IGP 协议不能识别，可能会丢失。
- 因此我们需要设计 BGP 在 AS 内部完成路由的传递。

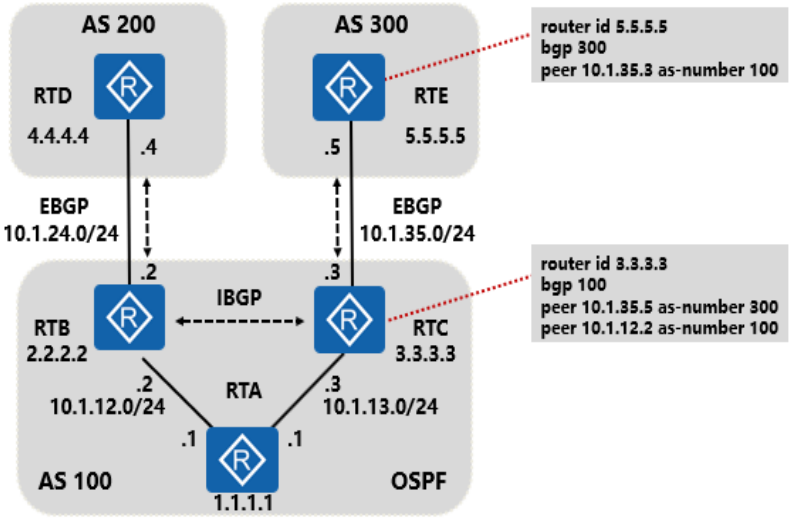
BGP邻居类型 - IBGP



- 运行在相同AS内的BGP路由器建立的邻居关系为IBGP（Internal BGP）邻居关系。

- 如上图，因为 BGP 使用 TCP 作为其承载协议，所以可以跨设备建立邻居关系。如图所示，RTB 与 RTC 之间建立 IBGP 邻居关系，并各自将从其他 AS 学到的路由传递给对端，实现 BGP 路由在 AS 内的传递。

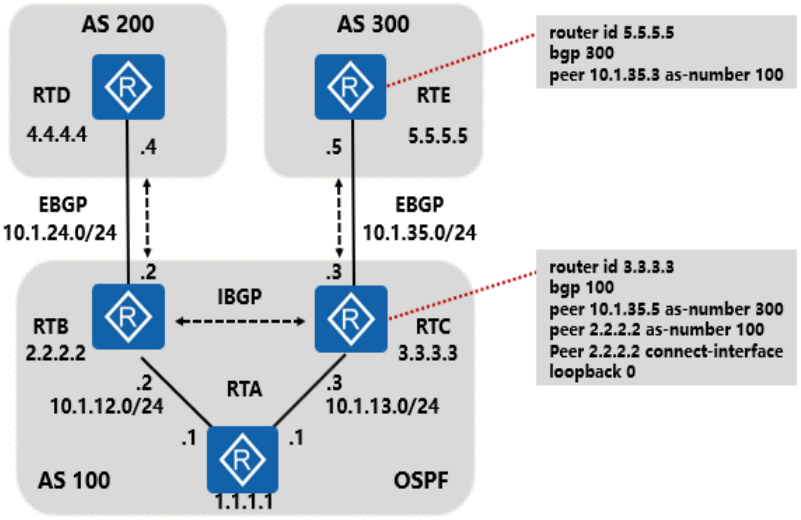
BGP邻居关系配置



- 配置步骤：

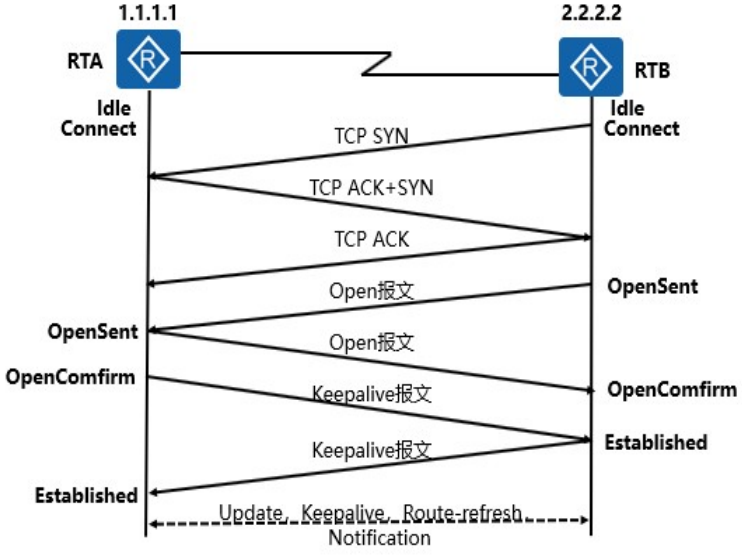
- 配置 Router ID（标识路由器）；
- 配置 EBGP 邻居关系（AS 之间传递路由）；
- 配置 IBGP 邻居关系（AS 内部传递路由）。
- 配置解释：
- 如果没有配置 Router ID，BGP 路由器会按一定规则自动选举 Router ID，选举规则如下：
- 路由器在它的所有 LoopBack 接口上选择数值最高的 IP 地址；
- 如果没有 LoopBack 接口，路由器会在它的所有物理接口上选择数值最高的 IP 地址。
- 配置命令：router id X.X.X.X
- BGP 邻居关系的类型主要靠配置的 AS 号区别，peer 关键字后面是对端邻居的接口 IP 地址，as-number 后面是邻居路由器所在的 AS 号，AS 号相同则为 IBGP 邻居关系；AS 号不同，则为 EBGP 邻居关系。
- peer 关键字后面是对端邻居的更新源 IP 地址，标识自己向对端邻居发起 TCP 连接的目的地址。该地址可以是对端邻居直连接口的 IP 地址，也可以是非直连 LoopBack 接口的 IP 地址（但必须保证该 IP 地址路由可达）。建立 IBGP 邻居关系时，一般使用 LoopBack 接口的 IP 地址，因为 LoopBack 接口开启后一直处于 UP 状态，只要保证路由可达，邻居关系一直处于稳定状态；而建立 EBGP 邻居关系时，一般使用直连接口的 IP 地址，因为 EBGP 是跨 AS 建立邻居关系，邻居关系建立之前非直连接口之间的路由不可达。

BGP邻居关系配置的优化



- 建立 EBGP 邻居关系时，一般使用直连接口的 IP 地址；建立 IBGP 邻居关系时，一般使用 Loopback 接口的 IP 地址。

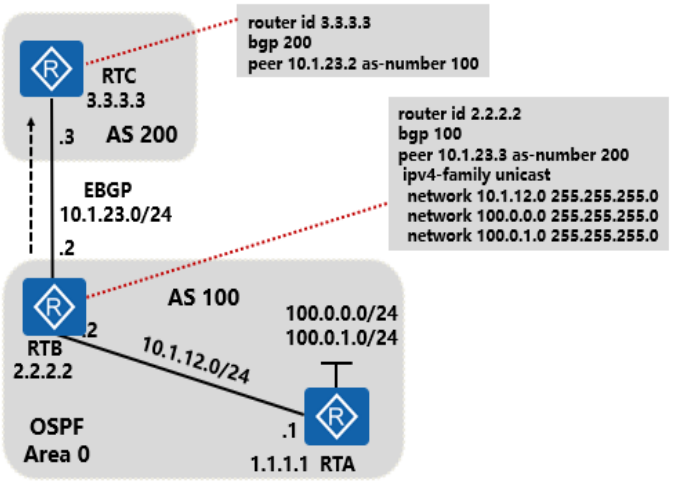
BGP邻居关系建立



- BGP 通过报文的交互完成邻居建立、路由更新等操作，共有 Open、Update、Notification、Keepalive 和 Route-refresh 等 5 种报文类型。

- Open 报文：是 TCP 连接建立后发送的第一个报文，用于建立 BGP 邻居之间的连接关系。BGP 邻居在接收到 Open 报文并协商成功后，将发送 Keepalive 报文确认并保持连接的有效性。确认后，BGP 邻居间可以进行 Update、Notification、Keepalive 和 Route-refresh 报文的交换。
- Update 报文：用于在 BGP 邻居之间交换路由信息。Update 报文可以发布多条属性相同的可达路由信息，也可以撤销多条不可达路由信息。
- 一条 Update 报文可以发布多条具有相同路由属性的可达路由，这些路由可共享一组路由属性。所有包含在一个给定的 Update 报文里的路由属性适用于该 Update 报文中的 NLRI（Network Layer Reachability Information）字段里的所有目的地（用 IP 前缀表示）。
- 一条 Update 报文可以撤销多条不可达路由。每一个路由通过目的地（用 IP 前缀表示），清楚地定义了 BGP 路由器之间先前通告过的路由。
- 一条 Update 报文可以只用于撤销路由，这样就不需要包括路径属性或者 NLRI。相反，也可以只用于通告可达路由，就不需要携带撤销路由信息了。
- Notification 报文：当 BGP 路由器检测到错误状态时，就向邻居发出 Notification 报文，之后 BGP 连接会立即中断。
- Keepalive 报文：BGP 路由器会周期性的向邻居发出 Keepalive 报文，用来保持连接的有效性。
- Route-refresh 报文：Route-refresh 用于在改变路由策略后请求对等体重新发送路由信息。

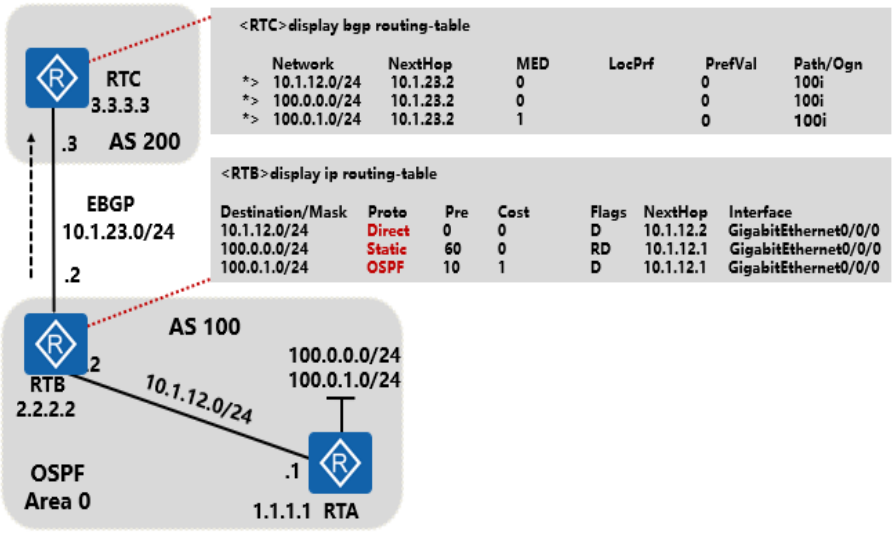
BGP路由的生成方式 – Network (1)



- Network命令是逐条将IP路由表中已经存在的路由引入到BGP路由表中。

- 生成 BGP 路由的方式有两种：第一种是使用配置命令 network，第二种是使用配置命令 import。
- 如图所示，RTA 上存在 100.0.0.0/24 与 100.0.1.0/24 的两个用户网段，RTB 上通过静态路由指定去往 100.0.0.0/24 网段的路由，通过 OSPF 学到去往 100.0.1.0/24 的路由。RTB 与 RTC 建立 EBGP 的邻居关系，RTB 通过 network 命令宣告 100.0.0.0/24,100.0.1.0/24 与 10.1.12.0/24 的路由，使对端 EBGP 邻居 RTC 学习到 RTB 路由表里的路由。

BGP路由的生成方式 – Network (2)



<RTC>display bgp routing-table

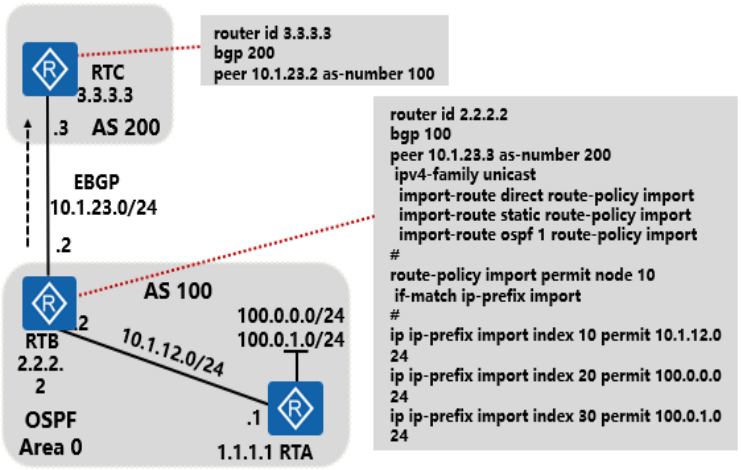
	Network	NextHop	MED	LocPrf	PrefVal	Path/Ogn
*>	10.1.12.0/24	10.1.23.2	0		0	100i
*>	100.0.0.0/24	10.1.23.2	0		0	100i
*>	100.0.1.0/24	10.1.23.2	1		0	100i

<RTB>display ip routing-table

Destination/Mask	Proto	Pre	Cost	Flags	NextHop	Interface
10.1.12.0/24	Direct	0	0	D	10.1.12.2	GigabitEthernet0/0/0
100.0.0.0/24	Static	60	0	RD	10.1.12.1	GigabitEthernet0/0/0
100.0.1.0/24	OSPF	10	1	D	10.1.12.1	GigabitEthernet0/0/0

- 通过display命令在RTC上查看是否学到BGP发布的路由条目。

BGP路由的生成方式 – Import (1)




```
router id 3.3.3.3
bgp 200
peer 10.1.23.2 as-number 100
```

```
router id 2.2.2.2
bgp 100
peer 10.1.23.3 as-number 200
 ipv4-family unicast
  import-route direct route-policy import
  import-route static route-policy import
  import-route ospf 1 route-policy import
#
route-policy import permit node 10
 if-match ip-prefix import
#
ip ip-prefix import index 10 permit 10.1.12.0
24
ip ip-prefix import index 20 permit 100.0.0.0
24
ip ip-prefix import index 30 permit 100.0.1.0
24
```

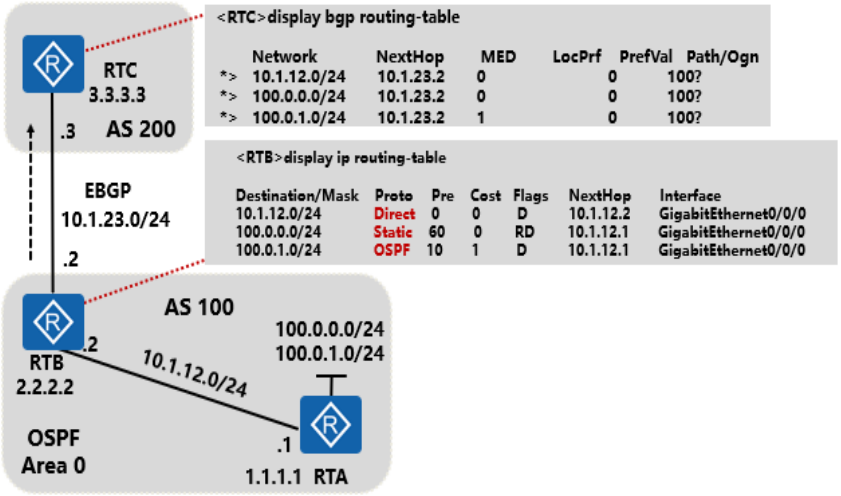
- Import命令是根据运行的路由协议（RIP，OSPF，ISIS等）将路由引入到BGP路由表中，同时import命令还可以引入直连和静态路由。

- RTA 上存在 100.0.0.0/24 与 100.0.1.0/24 的两个用户网段，RTB 上通过静态路由指定去往 100.0.0.0/24 网段的路由，通过 OSPF 学到去往 100.0.1.0/24 的路由。RTB 与 RTC 建立 EBGP 的邻居关系，RTB 通过 import 命令宣告 100.0.0.0/24,100.0.1.0/24 与 10.1.12.0/24 的路由，使对端 EBGP 邻居学习

到本 AS 内的路由。

- 为了防止其他路由被引入到 BGP 中，需要配置 ip-prefix 进行精确匹配，调用 route-policy 在 BGP 引入路由时进行控制。

BGP路由的生成方式 – Import (2)

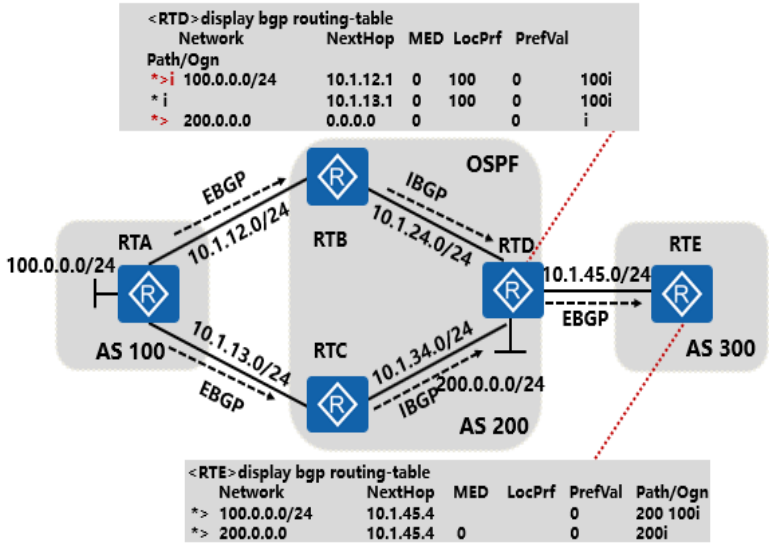


- 通过display命令在RTC上查看是否学到BGP引入的路由条目。

BGP的Update报文

- BGP通过Network和Import两种方式生成BGP路由，BGP路由封装在Update报文中通告给邻居。BGP在邻居关系建立后才开始通告路由信息。
- Update消息主要用来公布可用路由和撤销路由，Update中包含以下信息：
  - 网络层可达信息（NLRI）：用来公布IP前缀和前缀长度。
  - 路径属性：为BGP提供环路检测，控制路由优选。
  - 撤销路由：用来描述无法到达且从业务中撤销的路由前缀和前缀长度。
- 在通告BGP路由时，由于各种因素的影响，为了避免路由通告过程中出现问题，BGP路由通告需要遵守一定的规则，下面进行详细介绍。

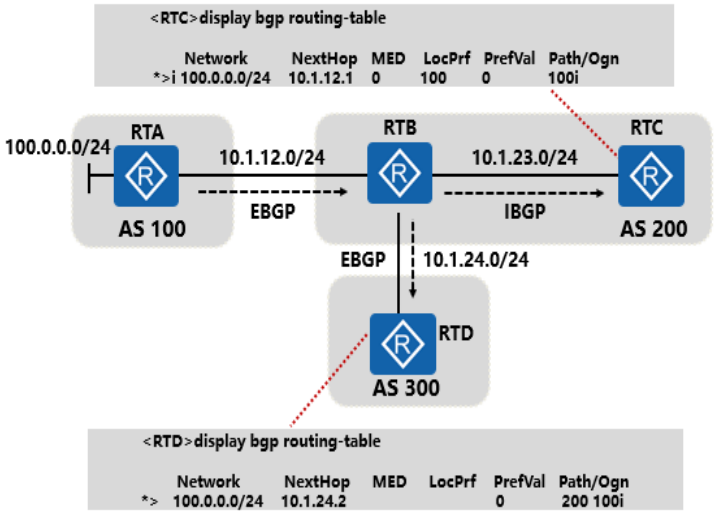
BGP通告原则之一：仅将自己最优的路由发布给邻居



- 存在多条有效路由时，BGP 路由器只将自己最优的路由发布给邻居。
- RTD 可以从 BGP 邻居 RTB 与 RTC 学习到 100.0.0.0/24 的路由，同时 RTD 将自己的直连路由 200.0.0.0/24 发布到 BGP 中。在 RTD 上使用命令 display bgp routing-table 查看如图所示；
- 在 RTE 上使用命令 display bgp routing-table 查看如图所示。可以发现，RTD 将自己标为有效且最优的路由发布给了 BGP 邻居 RTE。
- BGP 路由表中的状态含义：
- Status codes: * - valid, > - best, d - damped, h - history, i - internal, s - suppressed, S - Stale
- Origin : i - IGP, e - EGP, ? – incomplete
- Network：显示 BGP 路由表中的网络地址
- NextHop：报文发送的下一跳地址
- MED：路由度量值
- LocPrf：本地优先级
- PrefVal：协议首选值

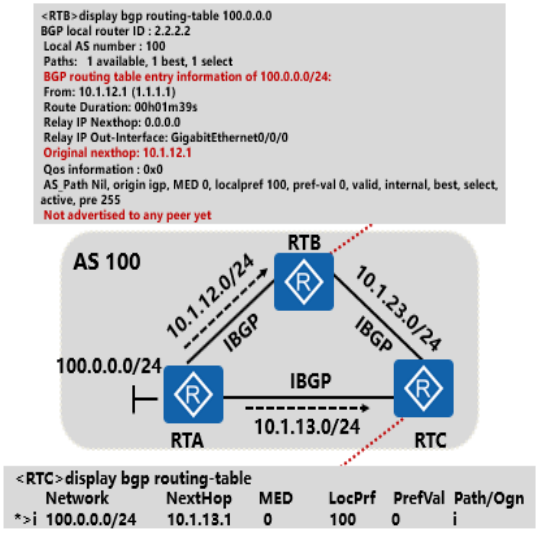
- Path/Ogn：显示 AS 路径号及 Origin 属性
- Community：团体属性信息

BGP通告原则之二：通过EBGP获得的最优路由发布给所有BGP邻居



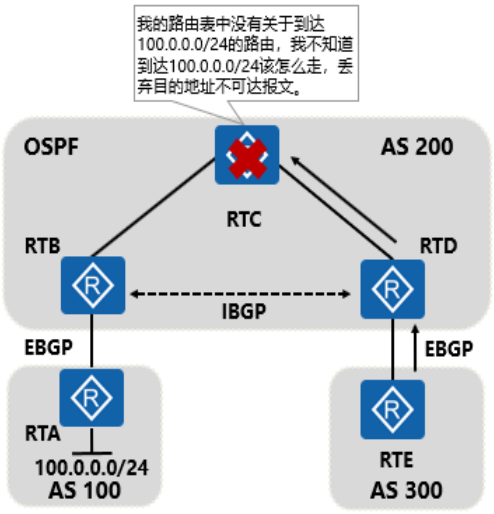
- BGP 路由器通过 EBGP 获得的最优路由会发布给所有的 BGP 邻居（包括 EBGP 邻居和 IBGP 邻居）。
- 如图所示，RTA 上有一个 100.0.0.0/24 的用户网段，并通过 EBGP 将该网段发布给 BGP 邻居 RTB。RTB 收到 EBGP 邻居发送来的 100.0.0.0/24 的路由后，将会通告给自己的 IBGP 邻居 RTC 与 EBGP 邻居 RTD。

BGP通告原则之三：通过IBGP获得的最优路由不会发布给其他的IBGP邻居



- BGP 路由器通过 IBGP 获得的最优路由不会发布给其他的 IBGP 邻居。
- 如图所示，RTA 上存在一个 100.0.0.0/24 的用户网段，RTA、RTB 与 RTC 之间互为 IBGP 邻居，RTA 通过 IBGP 将 100.0.0.0/24 的路由发布给 RTB 与 RTC，但是 RTB 并不会将收到的 IBGP 路由发布给自己的 IBGP 邻居 RTC。
- 这样设计的目的是防止在 AS 内部形成路由环路。根据规定，BGP 路由在同一个 AS 内进行传递时，AS_Path 属性不会发生变化。如图所示，RTA 将 100.0.0.0/24 的路由发布给 RTB 时，AS_Path 属性不变，为空。如果 RTB 能将 IBGP 路由 100.0.0.0/24 发布给 RTC，AS_Path 依旧为空。则 RTC 也有可能将 100.0.0.0/24 的路由发布给 RTA，因为 AS_Path 为空，RTA 并不会拒收该 IBGP 路由，路由环路产生。因此，上述通告原则是为了防止在 AS 内部形成路由环路。

BGP通告原则之四：BGP与IGP同步

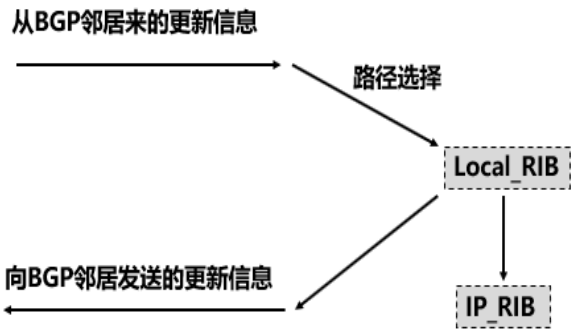


- RTA 上存在一个 100.0.0.0/24 的用户网段，通过 EBGP 发布给 RTB。RTB 与 RTD 建立了 IBGP 邻居关系，RTD 通过 IBGP 学习到该 BGP 路由，并将该路由发布给 EBGP 邻居 RTE。
- 当 RTE 访问 100.0.0.0/24 的路由时，查找路由表，发现到达 100.0.0.0/24 路由的下一跳是 RTD，RTE 查找出接口后，将数据包发送给 RTD；RTD 收到数据包后，查找路由表，发现到达 100.0.0.0/24 路由的下一跳是 RTB，出接口是 RTD 上与 RTC 相连的接口，于是将数据包发给 RTC，RTC 查找路由表，发现没有到达 100.0.0.0/24 的路由，于是将数据丢弃，形成“路由黑洞”。
- BGP 的通告原则：一条从 IBGP 邻居学来的路由在发布给一个 BGP 邻居之前，通过 IGP 必须知道该路由，即 BGP 与 IGP 同步。
- 如图所示，RTD 在收到 RTB 发来的 IBGP 路由之后，如果要发布给 BGP 邻居 RTE，则在发布之前先检查 IGP 协议（即 OSPF 协议）能否学到该条路由。如果能，则将 IBGP 路

由发布给 RTE。

- 在华为路由器上，默认是将 BGP 与 IGP 的同步检查关闭的，原因是为了实现 IBGP 路由的正常通告。但关闭了 BGP 与 IGP 的同步检查后会出现“路由黑洞”的问题。因此，有两种解决方案解决上述问题：
- 将 BGP 路由引入到 IGP，从而保证 IGP 与 BGP 的同步。但是，因为 Internet 上的 BGP 路由数量十分庞大，一旦引入到 IGP，会给 IGP 路由器带来巨大的处理和存储负担，如果路由器负担过重，则可能瘫痪。
- IBGP 路由器必须是全互联，确保所有的路由器都能学习到通告的路由。这样可以解决关闭同步后导致的“路由黑洞”问题。

BGP路由信息处理



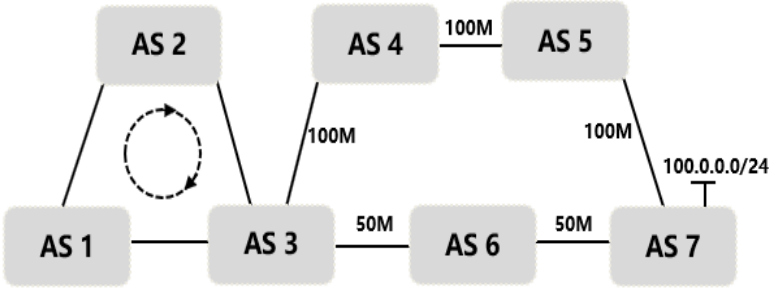
- 当从BGP邻居接收到Update报文时，路由器将会执行路径选择算法，来为每一条前缀确定最佳路径；
- 得出的最佳路径被存储到本地BGP路由表（Local_RIB）中，然后被提交给本地IP路由表（IP_RIB），以用作安装考虑；
- 被选出的有效的最佳路径路由将会被封装在Update报文中，发送给对端的BGP邻居。

- IP 路由表（IP_RIB）：全局路由信息库，包括所有的 IP 路由信息。
- BGP 路由表（Local_RIB）：BGP 路由信息库，包括本地 BGP 路由器选择的路由信息，邻居表，邻居清单列表。
- 收到 BGP 邻居发来的 Update 报文，路由器会执行算法

进行路径选择，确定每一条前缀的最佳路径，并将计算出的最佳路径存储到本地 BGP 路由表（Local_RIB）中。

- 如果启用了多路径特性，最佳路径和所有等值路径都被提交给 IP_RIB，考虑是否安装。除了从 BGP 邻居接收的最佳路径外，Local_RIB 也包含当前路由器注入的路由（被称为本地发起的路由）。
- 在 Local_RIB 中，只有被选为最优的前缀才会被封装到 Update 报文中通告给自己的 BGP 邻居。

BGP选路遇到的问题



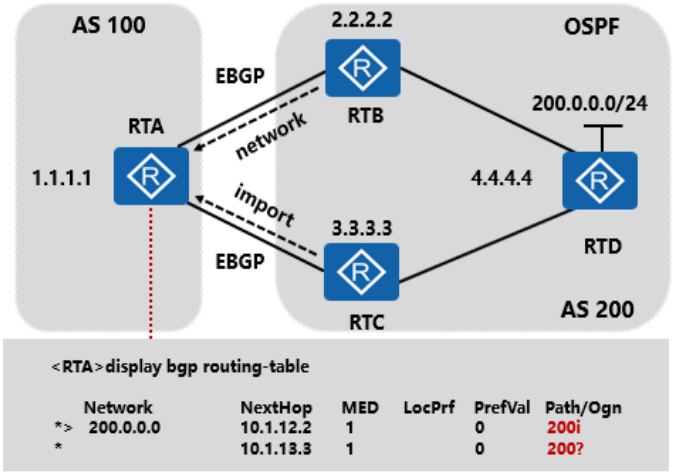
- 如图，AS 7中有一个100.0.0.0/24的用户网段，通过BGP发布给各个AS，各个AS都能学到100.0.0.0/24的路由，但是路由在传递过程中存在两个主要的问题：
  - AS 3可以从AS 4与AS 6两个AS收到100.0.0.0/24的路由，但AS 3与AS 4之间的链路带宽较大，有哪些方法可以影响AS 3选择AS 4访问100.0.0.0/24的网段?
  - AS 1，AS 2与AS 3之间存在拓扑上的环路，因此数据包在传递的过程中可能出现环路，怎么解决类似的环路问题?

- 以上两个问题的解决方案：
- 在 AS 之间交换路由可达信息时，设计 BGP 能够提供丰富的属性，实现对路由的灵活控制和优选。
- 修改路由表，调整 AS 之间的链路 Metric；2.不修改路由表， 使用策略修改路由下一跳。但是这些方法在某些情况下具有局限性，不能满足网络的丰富需求。
- 路由在 AS 之间传递时记录传播路径，防止环路的产生。

BGP的丰富属性

- 公认属性：所有 BGP 路由器都必须识别并支持的属性。
- 公认必遵：BGP 的 Update 消息中必须包含的属性。
- 公认任意：不必存在于 BGP 的 Update 消息中，可以根据需求自由选择的属性。
- 可选属性：不要求所有的 BGP 路由器都能够识别的属性。
- 可选过渡：BGP 不能识别该属性，但可以接收该属性并将其发布给它的邻居的属性。
- 可选非过渡：BGP 可以忽略包含该属性的消息并且不向它的邻居发布。

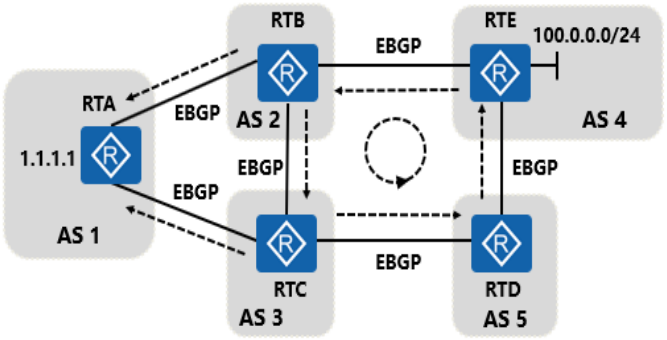
BGP属性 - Origin



- Origin属性定义路径信息的来源，标记一条路由是怎么成为BGP路由的。

- 如图所示，AS 200 内运行 OSPF 协议，200.0.0.0/24 网段宣告到 OSPF 中。RTB 通过 network 方式将 200.0.0.0/24 的路由变为 BGP 路由通告给 RTA，RTC 通过 import 方式将 200.0.0.0/24 的路由变为 BGP 路由通告给 RTA。
- BGP 在 AS 之间传递信息，承载大量的路由。如果到达同一目的 IP 有多条路径，且 BGP 学到这些路由通过不同的方式，则 Origin 属性是决定最优路径的一个因素，用于标明路由的起源。
- Origin 的 3 种属性：
- i 表明 BGP 路由通过 network 命令注入；
- e 表明 BGP 路由是从 EGP 学来的，EGP 协议在现网中很难见到，但可以通过路由策略将路由的 Origin 属性修改为 e；
- ？即 Incomplete 表明 BGP 路由通过其它方式学到路由信息，如使用 import 命令引入的路由。
- 3 种 Origin 属性的优先级为：i>e>Incomplete（ ？ ）。

BGP属性 - AS_Path

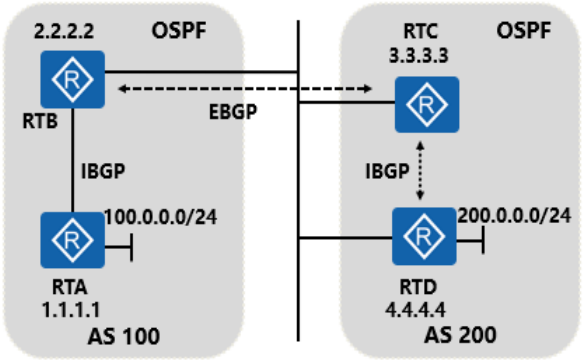


- 如图所示:
  - AS 1内的RTA能够从RTB与RTC收到100.0.0.0/24的路由，RTA如何进行自动优选?
  - RTA->RTB->RTC之间在拓扑上存在环路，RTB->RTC->RTD->RTE之间在拓扑上也存在环路，因此BGP在路由传递的过程中也可能存在路由环路，BGP如何防止环路呢?

- BGP 针对以上 2 个问题，设计了 AS_Path 属性，该属性记录了路由经过的所有 AS 的编号：
- 图中 RTA 从 RTB 收到 100.0.0.0/24 的路由时，AS_Path 为（2，4），RTA 从 RTC 收到 100.0.0.0/24 的路由时，AS_Path 为（3，5，4）。规定 AS_Path 越短（记录的 AS 编号越少），路径越优，因此 RTA 会优选从 RTB 收到的 100.0.0.0/24 的路由。
- 以 RTE 为例，通过 BGP 发布 100.0.0.0/24 的路由，路由可能通过 RTE->RTB->RTC->RTD->RTE 形成环路。为了防止环路的产生，RTE 在收到 RTD 发来的路由时会检查 AS_Path（该路由携带的）属性，如果发现该路由的 AS_Path 中包含自己的 AS 号，则丢弃该路由。
- AS_Path 的 4 种类型：
- AS_Sequence（后续讲解 BGP 路由聚合时会详细说明）；
- AS_Set（后续讲解 BGP 路由聚合时会详细说明）；
- AS_Confed_Sequence（应用于联盟，本课程不涉及）；
- AS_Confed_Set（应用于联盟，本课程不涉及）。

BGP属性 - Next_hop

- 如图所示：

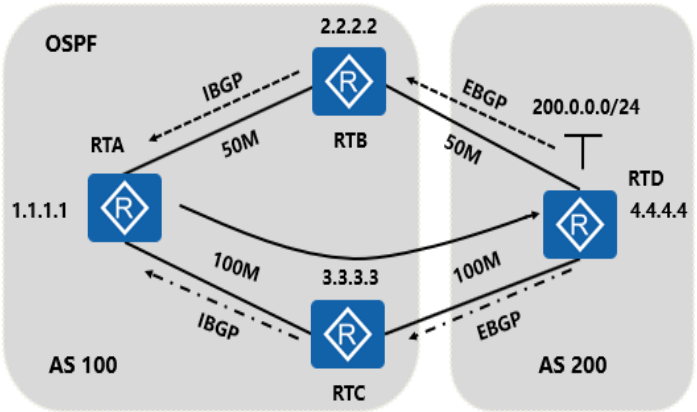


  - RTA将100.0.0.0/24的网段发布给RTB时，Next_hop的IP地址是多少?
  - RTB将100.0.0.0/24的网段发布给RTC时，Next_hop的IP地址是多少?
  - RTA从RTB学到RTC发布的200.0.0.0/24的网段时，Next_hop的IP地址是多少?

- BGP 路由器将本端始发路由发布给 IBGP 邻居时，会把该路由信息的 Next_hop 设为本端建立邻居关系所使用的接口 IP。
- 如图所示，RTA 将 100.0.0.0/24 的网段发布给 RTB 时，如果 RTA 与 RTB 使用直连接口建立 IBGP 邻居，则 Next_hop 为 RTA 上与 RTB 直连的接口 IP；如果 RTA 与 RTB 使用 Loopback 接口建立 IBGP 邻居，则 Next_hop 为 RTA 的 Loopback 接口 IP。
- BGP 路由器在向 EBGP 邻居发布路由时，会把路由信息的 Next_hop 设置为本端与对端建立 BGP 邻居关系的接口 IP。
- 如图所示，RTB 将 100.0.0.0/24 的网段发布给 RTC 时，Next_hop 为 RTB 上与 RTC 直连的接口 IP。
- BGP 路由器在向 IBGP 邻居通告从 EBGP 学来的路由时，不改变该路由下一跳属性。
- 特例：如图所示，RTA 从 RTB 学到 RTC 发布的 200.0.0.0/24 的网段时，Next_hop 为 RTD 的出接口 IP，因为 RTB 与 RTD 在同一网段，RTC 通告给 RTB 的 Next_hop 为 RTD 的出接口 IP。

- 对于上述三种情况的解释：
- EBGP 邻居之间一般采用直连接口建立邻居关系，EBGP 邻居在相互通告路由时会修改 Next_hop 为自己的出接口 IP；
- IBGP 邻居通常采用 Loopback 接口建立邻居，当路由是本路由器起源的，在发送给邻居之后 Next_hop 改为自己的更新源地址，这样即使网络中出现链路故障，只要 Next_hop 可达，同样可以访问目的网段，提高网络稳定性；
- 相对于 IGP，如 RIP 在发布路由时，每经过一个路由器都会修改下一跳，发布路由的路由器都宣称自己能够到达目标地址，并采用逐跳传递的方式将数据包发送给目标网络，但网络中的路由器并不知道谁是真正的始发路由器，因此会造成环路。BGP 在 EBGP 之间传递时才修改 Next_hop，IBGP 发送从 EBGP 学来的路由给 IBGP 邻居时并不修改下一跳，在一定程度上起到了防环作用。

BGP属性 - Local_Preference

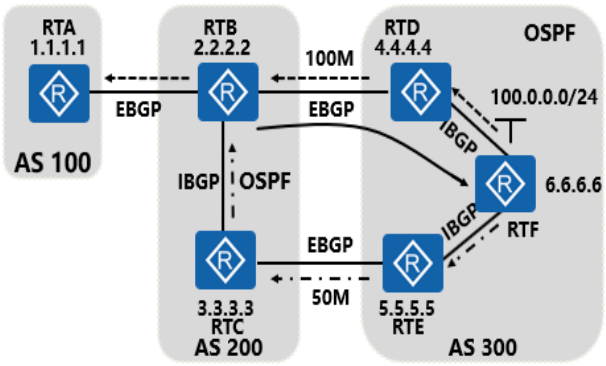


- Local_Pref属性仅在IBGP邻居之间有效，不通告给其他AS。它表明路由器的BGP优先级，用于判断流量离开AS时的最佳路由。

- 如图所示，AS 200 内有一个 200.0.0.0/24 的用户网段，通过 BGP 发布给 AS 100。AS 100 内的管理员如何设置才可以实现通过高带宽链路访问 200.0.0.0/24 的网络？
- 解决办法：

- 在 RTC 上设置 ip-prefix 匹配 200.0.0.0/24 的路由，使用 route-policy 调用该 ip-prefix，并设置 Local_Preference 为 200，将策略应用在对 RTA 发布路由的 export 方向。
- Local_Pref 属性仅在 IBGP 邻居之间有效，不通告给其他 AS。它表明路由器的 BGP 优先级，值越大越优。
- Local_Pref 属性用于判断流量离开 AS 时的最佳路由。当 BGP 路由器通过不同的 IBGP 邻居获得目的地址相同但下一跳不同的多条路由时，将优先选择 Local_Pref 属性值较高的路由，其默认值为 100。

BGP属性 - MED



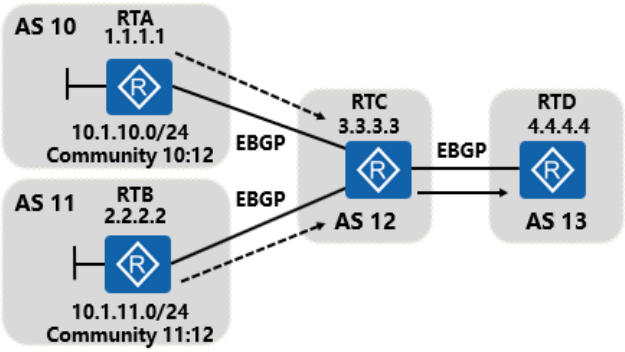
- MED（Multi-Exit-Discriminator）属性仅在相邻两个AS之间传递，收到此属性的AS不会再将其通告给任何其他第三方AS，用于判断流量进入AS时的最佳路由。

- 如图所示，AS 300 内的管理员希望在 AS 300 内操作影响 AS 200 通过高带宽链路访问 100.0.0.0/24，如何实现？
- 解决方法：
- 在 RTE 上设置 ip-prefix 匹配 100.0.0.0/24 的路由，再设置 route-policy 调用该 ip-prefix，并设置 MED 为 100，将策略应用在对 RTC 发布路由的 export 方向。
- MED（Multi-Exit-Discriminator）属性仅在相邻两个 AS 之间传递，收到此属性的 AS 不会再将其通告给任何其他第三

方 AS。如图所示，AS100 内并不会收到 AS 300 内设置的 MED 值，但是 AS 200 内会收到 AS 300 内设置的 MED 值，因此 AS 200 内可以选择高带宽的路由。

- MED 属性相当于 IGP 使用的度量值（Metric），它用于判断流量进入 AS 时的最佳路由。当一个运行 BGP 的路由器通过不同的 EBGP 邻居获得目的地址相同但下一跳不同的多条路由时，在其它条件相同的情况下，将优先选择 MED 值较小者作为最佳路由，其默认值为 0。

BGP属性 - Community



- BGP的Community属性的两个作用：
  - 限定路由的传播范围。
  - 打标记，便于对符合相同条件的路由进行统一处理。

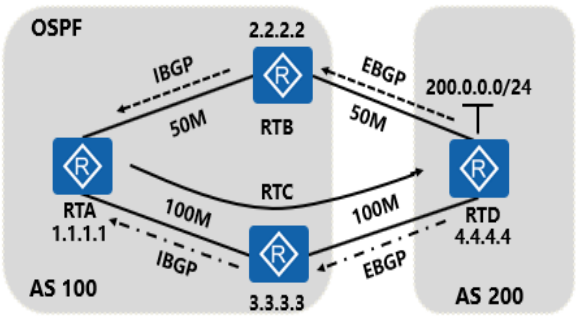
- 如图所示，AS 10 内有 10.1.10.0/24 的用户网段，AS 11 内有 10.1.11.0/24 的用户网段。为了区分用户网段，AS 10 内的 10.1.10.0/24 设置了 10:12 的 Community，AS 11 的 10.1.11.0/24 设置了 11:12 的 Community，通过 BGP 发送给 AS 12 后，AS 12 希望汇总后屏蔽掉明细路由再发送给 AS 13，并且希望 AS 13 收到路由后不再传递给其他 AS，如何实现？
- 解决方法：
- 在 RTC 上设置 Community-filter，匹配 Community 为 10:12 和 11:12 的路由，再设置 route-policy 匹配 Community-filter，将两条路由聚合成 10.1.10.0/23 的路由并调用 route-policy。

- 在 RTC 上设置 route-policy，设置团体属性为 no-export，在 RTC 通告给 RTD 的 export 方向调用该 route-policy。

BGP路由优选原则

- BGP路由器将路由通告给邻居后，每个BGP邻居都会进行路由优选，路由选择有三种情况：
  - 该路由是到达目的地的唯一路由，直接优选。
  - 对到达同一目的地的多条路由，优选优先级最高的。
  - 对到达同一目的地且具有相同优先级的多条路由，必须用更细的原则去选择一条最优的。
- 一般来说，BGP计算路由优先级的规则如下：
  - 丢弃下一跳不可达的路由。
  - 优选Preference_Value值最高的路由（私有属性，仅本地有效）。
  - 优选本地优先级（Local_Preference）最高的路由。
  - 优选手动聚合>自动聚合>network>import>从对等体学到的。
  - 优选AS_Path短的路由。
  - 起源类型IGP>EGP>Incomplete。
  - 对于来自同一AS的路由，优选MED值小的。
  - 优选从EBGP学来的路由（EBGP>IBGP）。
  - 优选AS内部IGP的Metric最小的路由。
  - 优选Cluster_List最短的路由。
  - 优选Orginator_ID最小的路由。
  - 优选Router_ID最小的路由器发布的路由。
  - 优选具有较小IP地址的邻居学来的路由。

Preference_Value对选路的影响



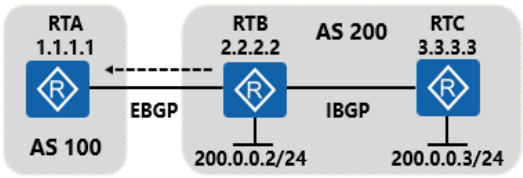
- Preference_Value是BGP的私有属性（华为私有属性），Preference_Value相当于BGP选路规则中Weight值，仅在本地路由器生效。Preference_Value值越大，越优先。

- 如图所示，AS 200 内有一个 200.0.0.0/24 的用户网段，AS 100 内的管理员希望通过高带宽链路访问 AS 200 内的 200.0.0.0/24 网段，并希望在 RTA 上的策略只能影响自己的选

路，不能影响其他设备，如何实现？

- 解决办法：
- 在 RTA 上设置 ip-prefix 匹配 200.0.0.0/24 的路由，再设置 route-policy 调用该 ip-prefix，并设置 Preference_Value 为 100，将策略应用在对 RTC 发布路由的 import 方向。
- 验证：RTC 上使用 Tracert 命令，查看访问 200.0.0.0/24 网段经过的路由器。

聚合方式对选路的影响



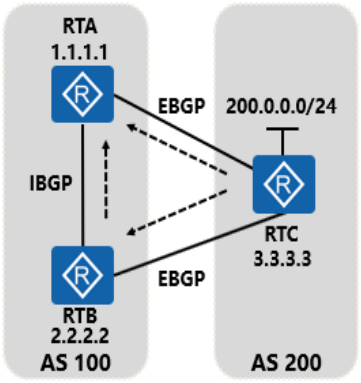

```
<RTB>display bgp routing-table 200.0.0.0
BGP local router ID : 2.2.2.2
 Local AS number : 200
 Paths:   2 available, 1 best, 1 select
 BGP routing table entry information of 200.0.0.0/24:
 Aggregated route.
 ......
 Aggregator: AS 200, Aggregator ID 2.2.2.2, Atomic-aggregate
 Advertised to such 2 peers:
    10.1.12.1
    10.1.23.3
 BGP routing table entry information of 200.0.0.0/24:
 Summary automatic route
 ......
 Aggregator: AS 200, Aggregator ID 2.2.2.2
 Not advertised to any peer yet
```

- 聚合路由的优先级：手动聚合>自动聚合。

- 如图所示，在 AS 200 内，RTB 与 RTC 上存在 200.0.0.0/24 网段的用户，RTB 与 RTC 将 200.0.0.0/24 的网段通过 import 方式变为 BGP 路由，在 RTB 上将路由聚合后发给 RTA，同时开启自动聚合与手动聚合，RTB 如何优选聚合路由？
- 如图所示，在 RTB 上同时使能自动聚合与手动聚合，使用命令查看，可以发现，手动聚合的路由条目被发送给 RTA，自动聚合的路由条目则没有通告，说明手动聚合的优先级高于自动聚合。
- 在使用路由聚合时需要注意，自动聚合只能对引入的 BGP 路由进行聚合，手动聚合可以对存在于 BGP 路由表中的路由进行聚合，后续在 BGP 路由聚合中详细介绍。上述场景中，

因为需要聚合的路由都是引入的路由，所以使用自动聚合与手动聚合都可以实现聚合的目的。如果 BGP 路由表中既有引入的路由又有 network 宣告的路由时，只能采用手动聚合实现。

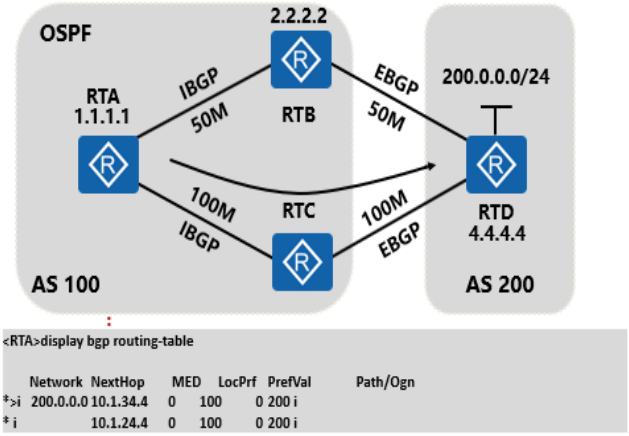
EBGP邻居的路由优于IBGP邻居的路由



- 根据选路原则，RTA会优选从EBGP邻居学来的路由。

- 如图所示，在 AS 200 内有一个 200.0.0.0/24 的网段，通过 EBGP 邻居关系通告给 RTA 与 RTB，RTB 会通过 IBGP 邻居关系将 200.0.0.0/24 的网段通告给 RTA，于是 RTA 会收到两条到达 200.0.0.0/24 的路由，RTA 会如何优选？
- 根据选路原则，RTA 会优选从 EBGP 邻居学来的路由。

AS内部IGP Metric对BGP选路的影响



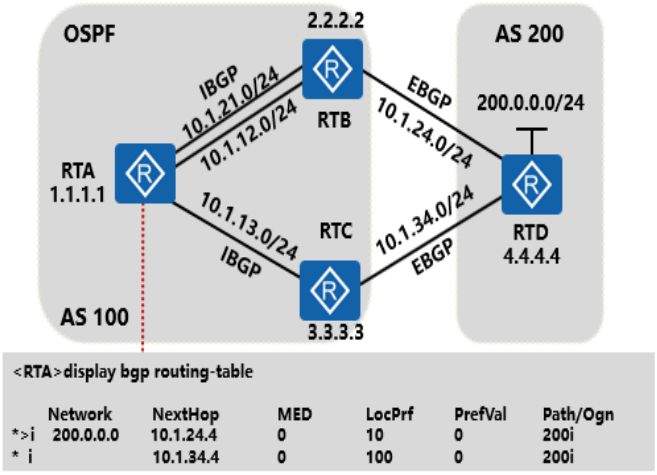
<RTA>display bgp routing-table

	Network	NextHop	MED	LocPrf	PrefVal	Path/Ogn
*>i	200.0.0.0	10.1.34.4	0	100	0	200 i
* i		10.1.24.4	0	100	0	200 i

- 如图所示，通过调整OSPF Cost，使RTA选择高带宽路径访问200.0.0.0/24网段。

- 如图所示，AS 200 内有一个 200.0.0.0/24 的用户网段，通过 EBGP 发布给 RTB 与 RTC，RTB 与 RTC 通过 IBGP 将路由发布给 RTA。AS 100 内的管理员希望通过高带宽链路访问 AS 200 内的 200.0.0.0/24 网段， RTA 上该如何实现？
- 将 RTA 与 RTB 所连接口的 OSPF Cost 值调为 100，RTA 则将选择 RTA->RTC->RTD 的路径访问 200.0.0.0/24 网段：
- 原因是 RTA 访问 200.0.0.0/24 时，到 Next_hop 10.1.34.4 的 Cost（2）小于到 Next_hop 10.1.24.4 (101)的 Cost。

Router-ID与IP地址对BGP选路的影响



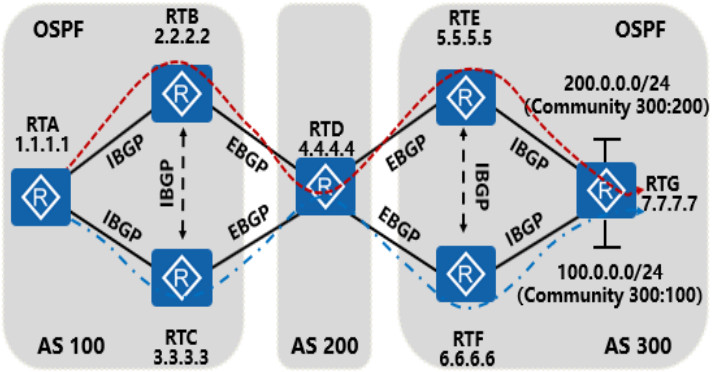
- 如图所示，RTA选择通过RTB访问AS内的200.0.0.0/24的网段，出接口为10.1.12.1地址所在的接口。

- 如图所示，AS 200 内有一个 200.0.0.0/24 的用户网段，通过 EBGP 发布给 RTB 和 RTC，RTB 和 RTC 通过 IBGP 将路由发布给 RTA。RTA 和 RTB 之间通过 2 条链路相连，RTA 会如何优选？
- RTA 会选择下一跳为 10.1.12.2 作为下一跳访问 200.0.0.0/24 的网段：
- RTA 选择 RTA->RTB->RTD 的路径访问 200.0.0.0/24 网段，原因是 RTB 的 Router-ID 比 RTC 小，BGP 优选 Router-ID 较小的路由器发布的路由；
- RTA 选择下一跳为 10.1.12.2 地址所在的接口为出接口，原因是 BGP 优选 IP 地址较小的邻居学来的路由。
- 在 RTA 上使用命令 display bgp routing-table 200.0.0.0 查看如下：

```
<RTA>display bgp routing-table 200.0.0.0
 BGP local router ID : 1.1.1.1
 Local AS number : 100
 Paths:    2 available, 1 best, 1 select
 BGP routing table entry information of 200.0.0.0/24:
```

From: 2.2.2.2 (2.2.2.2)
Route Duration: 00h02m10s
Relay IP Nexthop: 10.1.12.2
Relay IP Out-Interface: GigabitEthernet0/0/0
Original nexthop: 10.1.24.4
Qos information : 0x0
AS-path 200, origin igp, MED 0, localpref 100, pref-val 0, valid, internal, pre255, IGP cost 2, not preferred for router ID
……

BGP路由策略配置实例



- 如图所示，AS 300内有两个用户网段，AS 100内用户访问这两个网段时，希望在RTB和RTC上实现流量分担。AS 200访问这两个网段时，希望在RTE和RTF上实现流量分担。请用尽可能多的方法来实现上述需求。

- 如图所示，AS 300 内有两个用户网段，一个是 200.0.0.0/24，一个是 100.0.0.0/24。为了区分不同网段的用户，在 AS 300 内为 100.0.0.0/24 的网段分配 Community 属性为 300:100，为 200.0.0.0/24 的网段分配 Community 属性为 300:200。AS 100 内用户访问这两个网段时，希望在 RTB 和 RTC 上实现流量分担。AS 200 访问这两个网段时，希望在 RTE 和 RTF 上实现流量分担。请用尽可能多的方法来实现上述需求。
- 根据需求，在 AS 100 访问这两个网段时，希望在 RTB 和 RTC 上实现流量分担；在 AS 200 访问这两个网段时，希

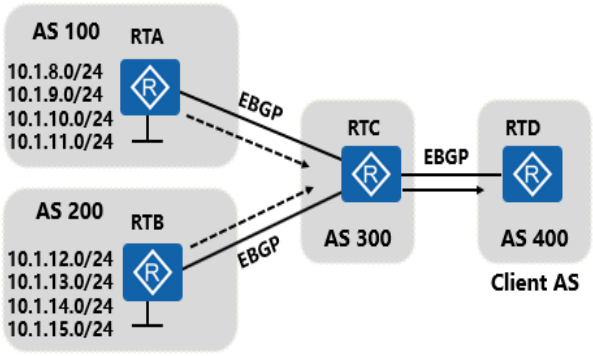
望在 RTE 和 RTF 上实现流量分担。假设 RTA 访问 100.0.0.0/24 时的路径为 RTA->RTB->RTD->RTE->RTG，访问 200.0.0.0/24 时的路径为 RTA->RTC->RTD->RTF->RTG，根据所学路径属性的知识，可供参考解决方案如下：

- RTE 和 RTF 向 RTD 通告携带团体属性的路由；
- RTD 收到携带团体属性的路由后，使用两个 Community-filter 分别匹配不同的团体属性，再使用两个 route-policy 分别调用 Community-filter，将匹配团体属性 300:100 的路由的下一跳设为 RTE 上的出接口地址；将匹配团体属性 300:200 的路由的下一跳设为 RTF 上的出接口地址；
- RTD 上再设置两个 route-policy，一个是将匹配团体属性为 300:100 的路由设置其 MED 值为 100，在对 RTC 的 export 方向调用；另一个是匹配团体属性为 300:200 的路由并设置其 MED 值为 100，在对 RTB 的 export 方向调用。

BGP路由聚合概述

- BGP在AS之间传递路由信息，随着AS数量的增多，单个AS规模的扩大，BGP路由表将变得十分庞大，因此带来如下两类问题：
  - 存储路由表将占用大量的内存资源，传输和处理路由信息需要消耗大量的带宽资源；
  - 如果传输的路由条目出现频繁的更新和撤销，对网络的稳定性会造成影响。
- 本节将介绍BGP的路由聚合对上述两种问题的处理，下面我们将从以下三个方面进行具体介绍：
  - BGP路由聚合的必要性——解决BGP网络存在的问题；
  - BGP路由聚合的配置方法；
  - BGP路由聚合带来的问题讨论。

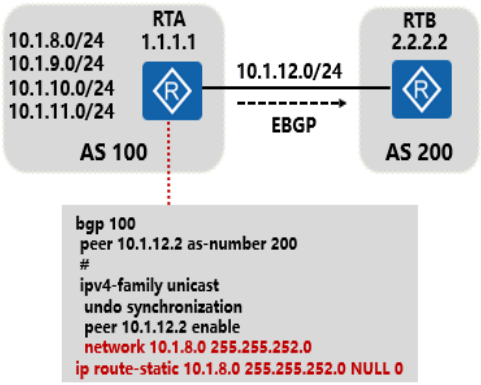
BGP路由聚合的必要性



- 如图所示，AS 100内有4个用户网段，AS 200内有4个用户网段。AS 300连接了一个Client AS，该AS内的路由器比较低端，处理能力较低，因此既希望能访问AS 100与AS 200内的网段，又不希望接收过多的明细路由，如何解决该问题?

- 解决方案：
- 在 RTC 上将 AS 100 和 AS 200 内的明细路由聚合成 10.1.8.0/21 的一条路由，并将此聚合路由发布给 Client AS。
- 现在 Internet 上的路由条目数量众多，处理这些路由时存在以下问题：
- 存储路由条目的路由表将占用大量的内存资源，传输路由信息需要占用大量的带宽资源；
- 明细路由频繁震荡造成网络不稳定。
- 因此，通过路由聚合来节省内存和带宽资源，减少路由震荡带来的影响成为必然。

BGP路由聚合方法 - 静态

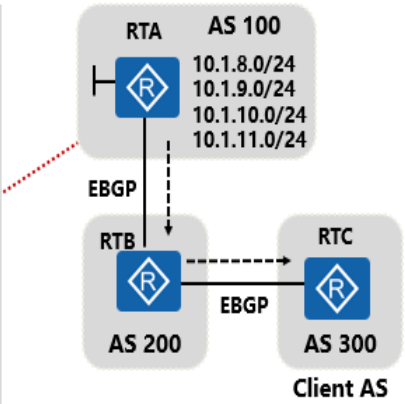


- AS 100内有4个用户网段，RTA通过路由聚合屏蔽明细路由，只将一条聚合后的路由10.1.8.0/22发布给AS 200内的RTB。

- 使用静态路由配置路由聚合的思路：
- 使用静态路由将明细路由聚合成 10.1.8.0/22，下一跳指向 NULL 0，因为聚合路由并不是具体的地址，发送给 AS 200 时只是明细路由的替代，为了防止路由环路，所以将下一跳指向 Null 0；
- 由于使用静态路由，路由表中产生了一条 10.1.8.0/22 的路由，下一跳为 Null 0。使用 network 命令将 IP 路由表中的 10.1.8.0/22 路由变为 BGP 路由，并通告给对端 BGP 邻居，达到聚合的目的。

BGP路由聚合方法 - 自动聚合

```
bgp 100
 peer 10.1.12.2 as-number 200
 #
 ipv4-family unicast
  undo synchronization
  summary automatic
  import-route direct route-policy r1
  peer 10.1.12.2 enable
#
route-policy r1 permit node 10
 if-match ip-prefix r1
#
ip ip-prefix r1 index 10 permit 10.1.11.0 24
ip ip-prefix r1 index 20 permit 10.1.10.0 24
ip ip-prefix r1 index 30 permit 10.1.9.0 24
ip ip-prefix r1 index 40 permit 10.1.8.0 24
```



- 如图所示，AS 100 内有 4 个用户网段，通过 import 的方式变为 BGP 路由，AS 200 连接了一个 Client AS，该 AS 内的路由器处理能力较低，因此既希望能访问 AS 100 与 AS 200 内的网段，又不希望接收过多路由，如何解决该问题？
- 配置如图所示，在 RTB 与 RTC 路由器上使用命令 display bgp routing-table 查看，输出如下：

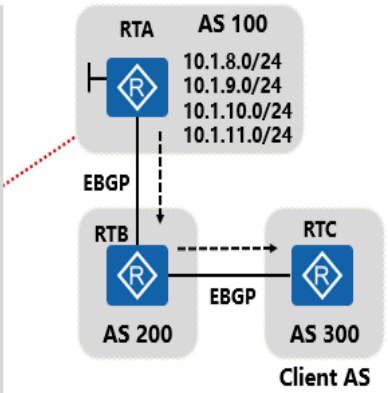
```
<RTB>display bgp routing-table

         Network            NextHop          MED
LocPrf      PrefVal   Path/Ogn
*>    10.0.0.0              10.1.12.1
0           100?
<RTC>display bgp routing-table

         Network              NextHop          MED
LocPrf      PrefVal   Path/Ogn
*>    10.0.0.0              10.1.23.2
0           200 100?
```

- 自动聚合只对引入 BGP 的路由进行聚合，聚合到自然网

段后发送给邻居。

BGP路由聚合方法 - 手动聚合

```
bgp 100
 peer 10.1.12.2 as-number 200
 #
 ipv4-family unicast
  undo synchronization
  aggregate 10.1.8.0 255.255.252.0
detail-suppressed
  network 10.1.8.0 255.255.255.0
  network 10.1.9.0 255.255.255.0
  import-route direct route-policy r1
  peer 10.1.12.2 enable
#
route-policy r1 permit node 10
 if-match ip-prefix r1
#
ip ip-prefix r1 index 10 permit 10.1.11.0 24
ip ip-prefix r1 index 20 permit 10.1.10.0 24
```



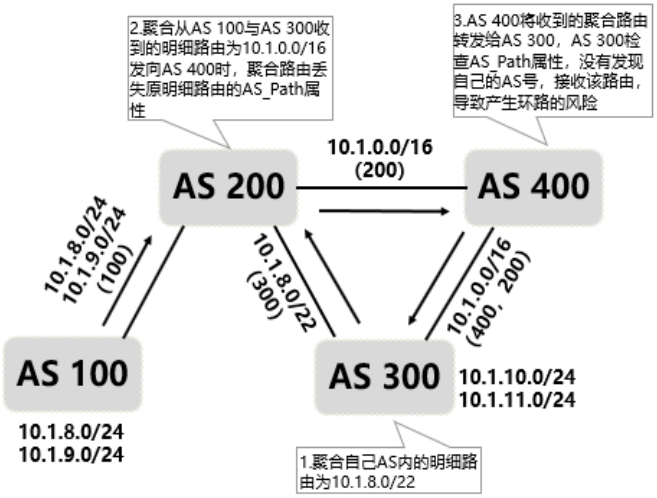
- 如图所示，AS 100 内有 4 个用户网段，既有通过 import 的方式引入 BGP 的路由，又有通过 network 方式引入 BGP 的路由。AS 200 连接了一个 Client AS，该 AS 内的路由器处理能力较低，因此既希望能访问 AS 100 与 AS 200 内的网段，又不希望接收过多路由，如何解决该问题？
- 配置如图所示，在 RTB 与 RTC 路由器上使用命令 display bgp routing-table 查看，输出如下：

```
<RTB>display bgp routing-table

         Network              NextHop          MED
LocPrf      PrefVal Path/Ogn
*>     10.1.8.0/22          10.1.12.1
0           100?
<RTC>display bgp routing-table

          Network              NextHop          MED
LocPrf      PrefVal Path/Ogn
 *>    10.1.8.0/22          10.1.23.2
```

0　　200 100?

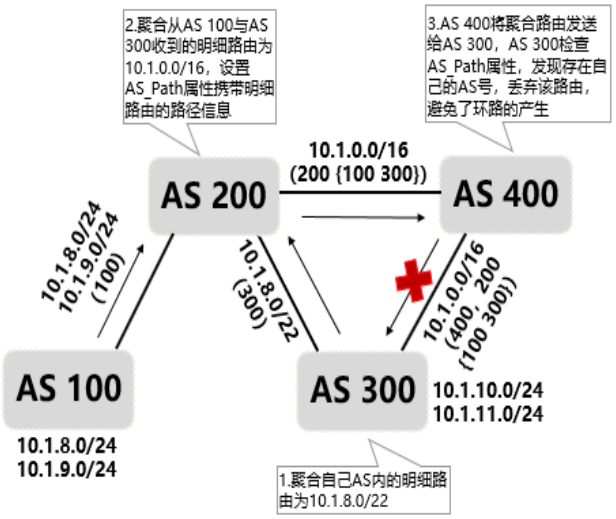
- 手动聚合对 BGP 本地路由表里存在的路由进行聚合，并且能指定聚合路由的掩码。

BGP路由聚合带来的问题 - 潜在环路



- 如何解决BGP路由聚合带来的潜在环路问题?

BGP路由聚合带来的问题 - 解决方法



- 为了解决 BGP 路由聚合带来的问题，设置了两个 AS_Path 属性：

- Atomic-Aggregate：公认任意属性，用于警告下游路由器出现了信息丢失，如图所示，AS 200 内设置了路由聚合的路由器在聚合后发生了路径丢失的现象，此时该路由器通过 Update 报文携带该属性通知自己的邻居发生了路径丢失。
- Aggregator：可选过度属性，该属性包含发起聚合的路由器的 AS 号和 Router-ID，表明发生聚合的位置。
- AS_Path 属性有两种类型：
- AS_Sequence：表示 AS_Path 内的 AS 号是一个有序的列表。
- AS_Set：表示 AS_Path 内的 AS 号是一个无序的列表。
- AS_Path 本身是一个有序的列表，因为 AS_Path 每经过一个 AS 都会将 AS 号添加到 AS_Path 中，并且按经过的顺序从左到右排列。
- 如图所示，AS 400 向 AS 300 通告聚合路由时，AS_Path 属性（大括号的除外）表示该聚合路由依次经过了 AS 200 和 AS 400。
- 当发生聚合后，如果需要聚合路由携带所有明细路由经过的 AS 号来防止环路，则在配置聚合的命令后添加 as-set 参数。
- 如图所示，AS 200 内发生了聚合并配置了 as-set 参数，则聚合路由会将明细路由的 AS_Path 信息用一个 AS-Set 集表示（放在中括号里的 AS 号信息，该集合的 AS 号没有先后顺序），携带在聚合路由后用以防止环路。
- 路由聚合解决了两类问题，一是减轻了设备传输和计算路由所需资源的负担，二是隐藏了具体的路由信息，减少了路由震荡的影响。但是路由聚合后，AS_Path 属性丢失，存在产生环路的风险。
- 如果路由聚合后携带所有明细路由经过的 AS 信息，当明细路由发生频繁震荡时，聚合路由也可能受其影响频繁刷新。
- 因此，聚合路由是否携带丢失的 AS_Path 信息，需要设

计者综合考虑网络环境。

思考题

1. BGP公认必遵属性有哪些？（ ）

   A. Origin

   B. AS_Path

   C. Next_hop

   D. Local_preference

2. BGP使用的端口号为多少？（ ）

   A. TCP 21

   B. TCP 179

   C. TCP 80

- 答案：ABC。
- 答案：B。
-