HCRSE106-BGP 双栈原理 Basic

BGP 知识点：

BGP 概述，基本配置，报文类型，邻居状态，邻居建立因素，对等体交互原则，路由属性，BGP 选路，BGP4+新增，BGP 特性（路由聚合，对等体组，团体，反射器，联盟）

BGP（Border Gateway Protocol）
BGP 是一种外部网关协议（EGP），与 OSPF、RIP 等内部网关协议（IGP）不同，其着眼点不在于自动发现网络拓扑，而在于在 AS 之间选择最佳路由和控制路由的传播。

BGP 使用 TCP 作为其传输层协议（监听端口号为 179），提高了协议的可靠性，且不需要专门的机制来确保连接的可控性。BGP 进行域间的路由选择，对协议的稳定性要求非常高。因此用 TCP 协议的高可靠性来保证 BGP 协议的稳定性。BGP 的对等体之间必须在逻辑上连通，并进行 TCP 连接。目的端口号为 179，本地端口号任意。

NLRI（Network Layer Reachability Information）网络层可达信息

BGP 防环

IBGP：BGP 在 AS 内学到的路由不再通告给 AS 内的 BGP 邻居，避免了 AS 内产生环路。
EBGP：BGP 携带 AS 路径信息标记途经 AS，带有本地 AS 号的路由将被丢弃，从而避免域间产生环路。

BGP 5 种报文

Open 消息：是 TCP 连接建立后发送的第一个消息，用于建

立 BGP 对等体之间的连接关系。对等体在接收到 Open 消息并协商成功后，将发送 Keepalive 消息确认并保持连接的有效性。
Update 消息：用于在对等体之间交换路由信息。一条 Update 消息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。
Keepalive 消息：BGP 会周期性的向对等体发出 Keepalive 消息，用来保持连接的有效性。
Notification 消息：当 BGP 检测到错误状态时，向对等体发 Notification 消息，之后 BGP 连接立即中断。
Route-Refresh 消息：通过 OPEN 消息告知 BGP peer 本地支持路由刷新能力（Route-Refresh capability）。在所有 BGP 路由器使能 Route-Refresh 能力的情况下，如果 BGP 的入口路由策略发生了变化，本地 BGP 路由器会向对等体发布 Route-Refresh 消息，收到此消息的对等体会将其路由信息重新发给本地 BGP 路由器。这样，可以在不中断 BGP 连接的情况下，对 BGP 路由表进行动态刷新，并应用新的路由策略。

BGP 报文应用：
BGP 使用 TCP 建立连接，本地监听端口为 179。和 TCP 连接建立相同，BGP 连接的建立也要经过一系列的对话和握手。TCP 通过握手协商通告其端口等参数，BGP 的握手协商的参数有：BGP 版本、BGP 连接保持时间、本地的路由器标识（Router ID）、授权信息等。这些信息都在 Open 消息中携带。
BGP 连接建立后，如果有路由需要发送则发送 Update 消息通告对端。Update 消息发布路由时，还要携带此路由的路由属性，用以帮助对端 BGP 协议选择最优路由。在本地 BGP 路由变化时，要通过 Update 消息来通知 BGP 对等体。
经过一段时间的路由信息交换后，本地 BGP 和对端 BGP 都无新路由通告，趋于稳定状态。此时要定时发送 KEEPALIVE

消息以保持 BGP 连接的有效性。对于本地 BGP，如果在保持时间内，未收到任何对端发来的 BGP 消息，就认为此 BGP 连接已经中断，将断开此 BGP 连接，并删除所有从该对等体学来的 BGP 路由。
当本地 BGP 在运行中发现错误时（如对端 BGP 版本本地不支持、本地 BGP 收到了结构非法的 Update 消息等），要发送 Notification 消息通告 BGP 对等体。本地 BGP 退出 BGP 连接时，也需发送 Notification 报文。

BGP6 种状态：
Idle、Connect、Active、OpenSent、OpenConfirm 和 Established。

Idle：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 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。
Connect：状态下，BGP 启动连接重传定时器（Connect Retry，缺省为 32 秒），等待 TCP 完成连接。
此阶段主动发起 TCP 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 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。
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。
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。
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。
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。

1、BGP open 包影响邻居关系建立的参数？
（1）版本： 一般都是 V4，不一致不能建立。
（2）Router-id： router-id 不能一样，否则不可以建立邻居。
（3）AS 号: 建立邻居的时候会检测 open 报文中 as 号是否与本端配置的邻居 as 号一致，如果不一致，不能正常建立邻居
（4）能力属性:（单播，组播，IPV4，IPV6，VPNV4，是否支持路由刷新以及 4 字节的 as 号）
（5）认证： 在 bgp 进程下指 peer 时配置认证，认证密钥要一致。
注意：holdtime 可以不一致（默认 180s，不一致采用小的），能力属性可以不一致，但是支持协议需有交集。

BGP 邻居建立不成功的原因

1.AS 号或 peer 邻居地址出错；
2.BGP 的 router ID 是否有冲突；
3.BGP 对等体两端是否均采用环回口创建邻居；
4.物理上非直连的 EBGP 邻居是否配置多跳；
5.用于创建底层 TCP 的路由是否可达；
6.创建 BGP 对等体两端认证配置是否一致；
7.BGP 对等体是否配置了 peer x.x.x.x ignore；
8.是否配置了禁止 TCP 端口 179 的 ACL。

BGP 路由信息处理：

当从对等体接收到更新数据包时，路由器会把这些更新数据包存储到路由选择信息库(Routing Information Base, RIB)中，并指明是来自哪个对等体的(Adj-RIB-In)。这些更新数据包被输入策略引擎过滤后，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 RIB (Loc-RIB)中，然后被提交给本地 IP 路由选择表(IP-RIB)，以用作安装考虑。
除了从对等体接收来的最佳路径外，Loc-RIB 也会包含当前路由器注入的(被称为本地发起的路由)，并被选择为最佳路径的 BGP 前缀。Loc-RIB 中的内容在被通告给其他对等体之前，必须通过输出策略引擎。只有那些成功通过输出策略引擎的路由，才会被安装到输出 RIB (Adj-RIB-Out)中。

BGP 对等体交互路由原则：

从 IBGP 对等体获得的路由，只发布给 EBGP 对等体
从 EBGP 对等体获得的路由，发布给所有 EBGP 和 IBGP 对等体

只将 BGP 的最优路由发布给对等体
只发送更新的 BGP 路由
IBGP 与 IGP 的同步：从 IBGP 邻居学到的路由，只有当 IGP 中也存在相同的路由时才会宣告给 EBGP 对等体。VRP 平台缺省情况下 BGP 与 IGP 是取消同步机制的，并不可改变。

BGP 路由属性

BGP 路由属性是一套参数，它是对路由的进一步的描述

公认必遵

所有 BGP 路由器都必须识别，且必须存在于 Update 消息中
如果缺少这种属性，路由信息就会出错

公认任意

所有 BGP 路由器都可以识别，但不要求必须存在于 Update 消息中
即就算缺少这类属性，路由信息也不会出错

可选过渡

在 BGP 对等体之间具有可传递性的属性
BGP 路由器可以不支持此属性，但它仍然会接收这类属性，并传递给其他对等体

可选非过渡

如果 BGP 路由器不支持此属性，则相应的这类属性会被忽略，且不会传递给其他对等体

Origin 为公认必遵属性
AS_Path 为公认必遵属性
Next_Hop 为公认必遵属性

Local_Pref 为公认任意属性
Atomic-Aggregate 为公认任意属性

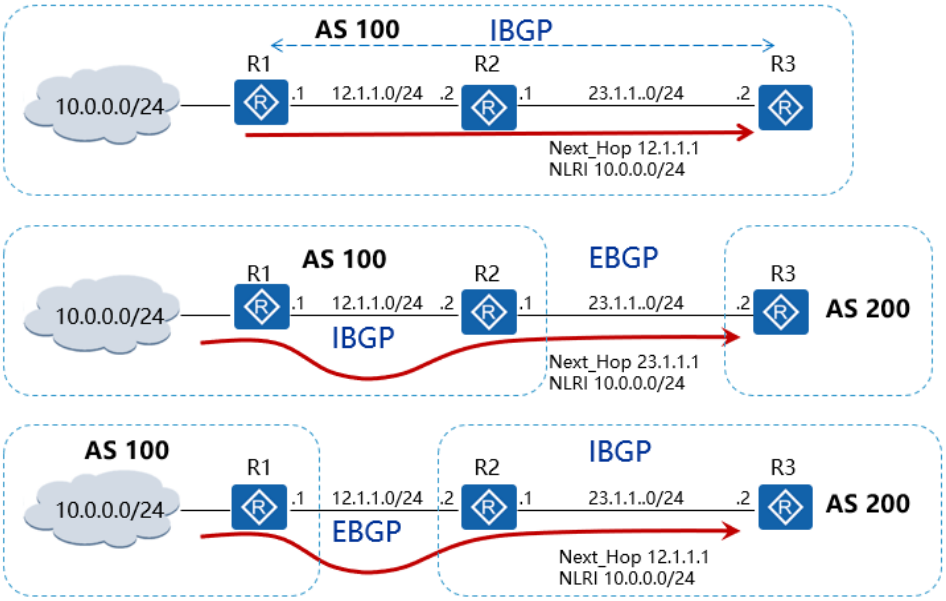
Aggregator 为可选过渡属性
community 为可选过渡属性

MED 为可选非过渡属性
Originator_ID 为可选非过渡属性
Cluster_List 为可选非过渡属性

BGP 选路规则：
1 优选协议首选值（PrefVal）最高的路由 ，华为私有，默认为 0 ，本地有效，越大越优
2 优选本地优先级（Local_Pref）最高的路由，默认为 100 ，IBGP 有效，越大越优
3 优选手动聚合、自动聚合、network 的路由、import-route 引入路由、从对等体学习路由
4 优选 AS 路径（AS_Path）最短的路由
5 比较 Origin 属性，依次优选 Origin 类型为 IGP、EGP、Incomplete 的路由
6 优选 MED 值最低的路由 ，默认为 0 ，EBGP 有效，越小越优
7 优选从 EBGP 邻居学来的路由（EBGP 路由优先级高于 IBGP 路由）
8 优选到下一跳 IGP Metric 较小的路由
前 8 条一样时，可以通过命令做负载分担
bgp 100
maximum load-balancing 2
9 优选 Cluster_List 最短的路由
10 优选 Router ID 最小的路由器发布的路由
11 比较对等体的 IP Address，优选从具有较小 IP Address 的对等体学来的路由

Next_Hop

Next_Hop 属性记录了路由的下一跳信息。BGP 的下一跳属性和 IGP 的有所不同，不一定就是邻居设备的 IP 地址。通常情况下，Next_Hop 属性遵循下面的规则：



1 BGP Speaker 将本地始发路由发布给 IBGP 对等体时，会把该路由信息的下一跳属性设置为本地与对端建立 BGP 邻居关系的接口地址。

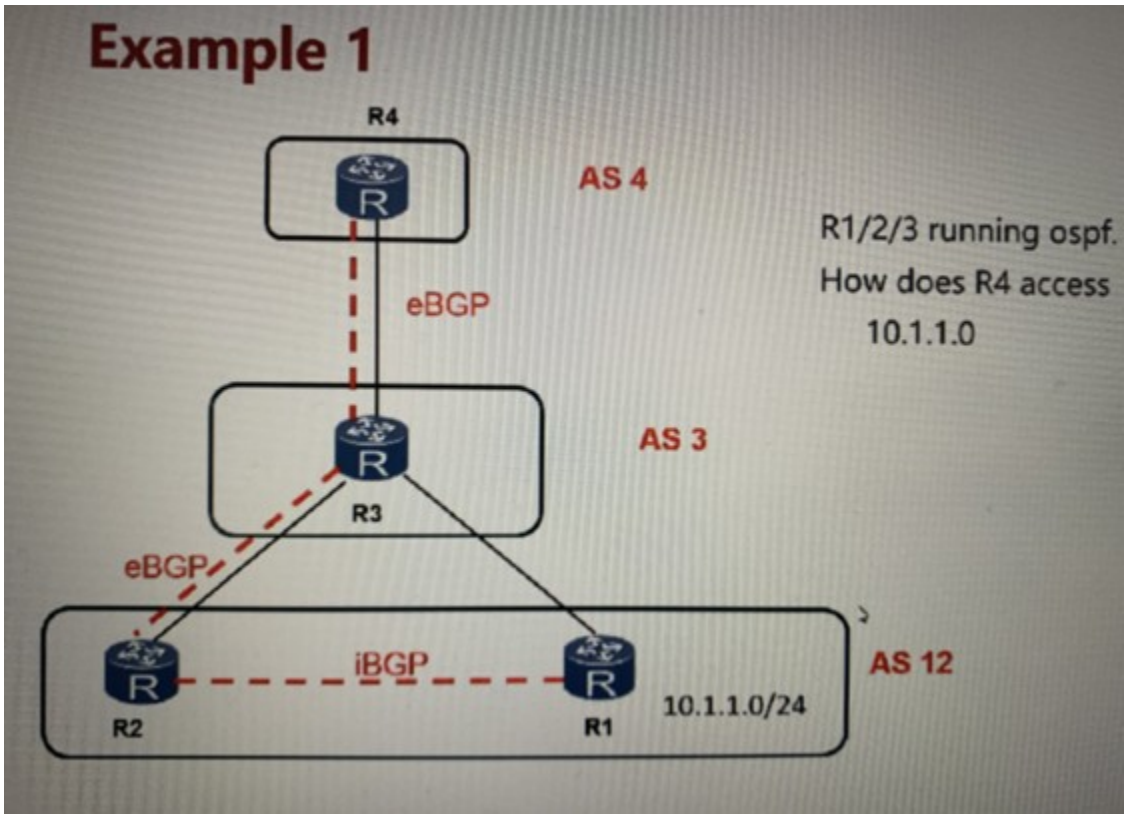
2 BGP Speaker 在向 EBGP 对等体发布某条路由时，会把该路由信息的下一跳属性设置为本地与对端建立 BGP 邻居关系的接口地址。

从 EBGP 邻居收到的路由，传给 EBGP 邻居时一定会改变下一跳吗？

默认情况下是会改变的，但是可以通过配置命令使下一跳不改变，或者在一些特殊场景下，发出路由给 EBGP 邻居时下一跳是不改变的，主要是为了防止出现次优路径的问题。

R2：

```
bgp 12
peer 192.168.23.3 next-hop-invariable
```



peer next-hop-invariable　命令配置不同 AS 域的 PE 向 EBGP 对等体发布路由时不改变下一跳；向 IBGP 对等体发布引入的 IGP 路由时使用 IGP 路由的下一跳地址。保证对端 PE 可以在流量传输时迭代到通往本端 PE 的 BGP LSP。

3 BGP Speaker 在向 IBGP 对等体发布从 EBGP 对等体学来的路由时，并不改变该路由信息的下一跳属性。 要设置 peer 1.1.1.1 next-hop-local ,改变下一跳

==========

BGP 4+

BGP4+不支持自动聚合

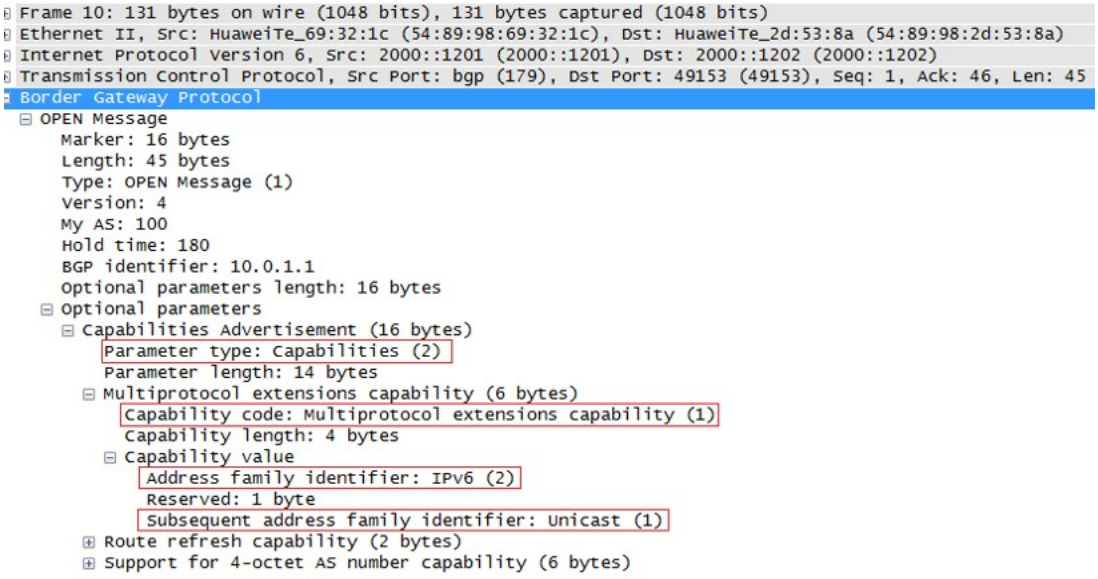
BGP 4 的扩展版本

扩展能力自协商机制
支持传递多种地址族地址（IPv6、VPNv4、VPNv6 等）
新增属性用以支持多地址族地址传递

当两台 BGP 对等体之间需要传输 IPv6 地址族的地址时，需要在 OPEN Message 中进行扩展能力的协商

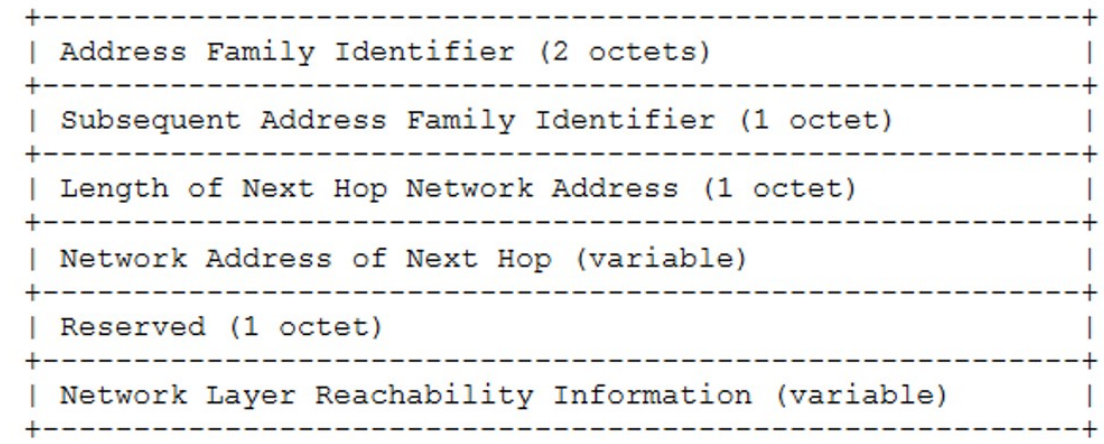
为实现对多种网络层协议的支持，BGP 需要将网络层协议的信息反映到 NLRI 及 Next_Hop。因此 MP-BGP 引入了两个新的可选非过渡路径属性：
MP_REACH_NLRI：Multiprotocol Reachable NLRI，多协议可达 NLRI。type code=14
用于发布可达路由及下一跳信息。
MP_UNREACH_NLRI：Multiprotocol Unreachable NLRI，多协议不可达 NLRI。type code=15
用于撤销不可达路由。
NLRI 网络层可达信息（Network Layer Reachability Information）

OPEN 消息中的 Capabilities Advertisement 字段用于扩展能力的协商。



MP_REACH_NLRI 属性（Type Code=14） BGP4+使用此属性来通告 IPv6 路由。

MP_UNREACH_NLRI 属性(Type Code= 15) BGP 4+用该属性撤销路由

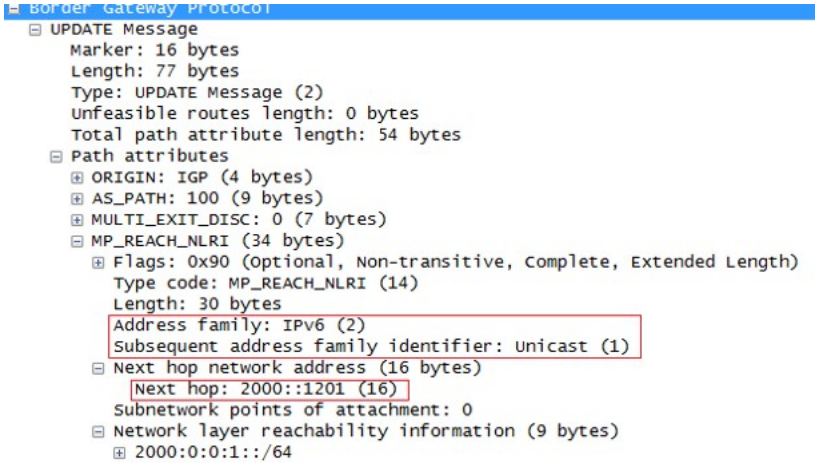


地址族信息（Address Family Information）域：由 2 字节的地址族标识 AFI（Address Family Identifier）和 1 字节的子地址族标识 SAFI（Subsequent Address Family Identifier）组成。

下一跳长度（Length of Next Hop Network Address）域：1 字节长度，表示下一跳地址的长度，通常情况下为 16

下一跳地址（Network Address of Next Hop）域：长度由上

一个字段决定，一般情况下为全球单播地址。
保留字段（Reserved）域：一字节，必须为 0
网络层可达信息（Network Layer Reachability Information）域：表示含有匹配相同属性的路由信息。当此字段为 0 时，表示为缺省路由。



当传递 IPv6 路由时
AFI=2 ，SAFI=1(Unicast) ,SAFI=2(Multicast)
下一跳地址长度字段决定了下一跳地址的个数
长度字段=16，下一跳地址为下一跳路由器的全球单播地址
长度字段=32，下一跳地址为下一跳路由器的全球单播地址和链路本地地址

前言

- 为了满足大规模路由的需求，需要通过BGP（Border Gateway Protocol）在AS之间传递数量庞大的IPv4、v6路由，并且通过各种策略进行路径的选取，控制等；此外，为了满足MPLS VPN的业务，在AS内及AS间需要MP-BGP。针对上述需求，本章节将介绍BGP原理以及BGP对IPv6的扩展特性—BGP4+。

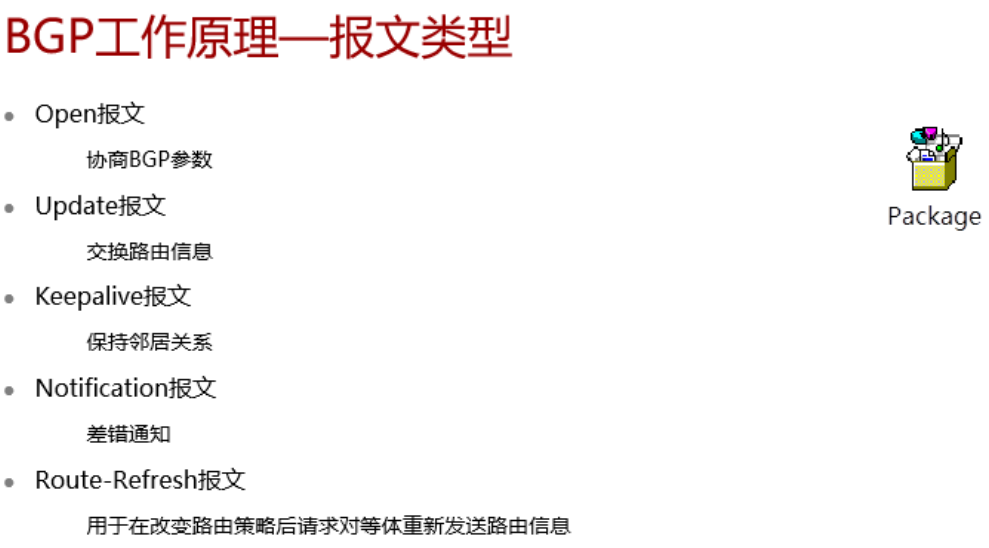
BGP概述

- BGP概述
  - 外部网关协议
  - 使用TCP作为其传输层协议
  - 支持CIDR
  - 增量更新
  - 路径矢量路由协议
  - 无环路
  - 路由策略丰富
  - 可防止路由振荡
  - 易于扩展

- BGP 是一种用于自治系统（Autonomous System）之间的动态路由协议。早期发布的三个版本分别是 BGP-1（RFC1105）、BGP-2（RFC1163）和 BGP-3（RFC1267），主要用于交换 AS 之间的可达路由信息，构建 AS 域间的传播路径，防止路由环路的产生，并在 AS 级别应用一些路由策略。当前使用的版本是 BGP-4（RFC4271）。
- BGP 作为事实上的 Internet 外部路由协议标准，被广泛应用于 ISP 之间。

- BGP 概述
- BGP 是一种外部网关协议（EGP），与 OSPF、RIP 等内部网关协议（IGP）不同，其着眼点不在于自动发现网络拓扑，而在于在 AS 之间选择最佳路由和控制路由的传播。
- BGP 使用 TCP 作为其传输层协议（监听端口号为 179），提高了协议的可靠性，且不需要专门的机制来确保连接的可控性。

- BGP 进行域间的路由选择，对协议的稳定性要求非常高。因此用 TCP 协议的高可靠性来保证 BGP 协议的稳定性。
- BGP 的对等体之间必须在逻辑上连通，并进行 TCP 连接。目的端口号为 179，本地端口号任意。
- 路由更新时，BGP 只发送更新的路由，大大减少了 BGP 传播路由所占用的带宽，适用于在 Internet 上传播大量的路由信息。
- BGP 从设计上避免了环路的发生。
- AS 之间：BGP 通过携带 AS 路径信息来标记途经的 AS，带有本地 AS 号的路由将被丢弃，从而避免了域间产生环路。
- AS 内部：BGP 在 AS 内学到的路由不再通告给 AS 内的 BGP 邻居，避免了 AS 内产生环路。
- BGP 提供了丰富的路由策略，能够对路由实现灵活的过滤和选择。
- BGP 提供了防止路由振荡的机制，有效提高了 Internet 网络的稳定性。
- BGP 易于扩展，能够适应网络新的发展。主要是通过 TLV 进行扩展。



- BGP 的运行是通过消息驱动的，共有 Open、Update、

Notification、Keepalive 和 Route-Refresh 等 5 种消息类型。

- Open 消息：是 TCP 连接建立后发送的第一个消息，用于建立 BGP 对等体之间的连接关系。对等体在接收到 Open 消息并协商成功后，将发送 Keepalive 消息确认并保持连接的有效性。确认后，对等体间可以进行 Update、Notification、Keepalive 和 Route-Refresh 消息的交换。
- Update 消息：用于在对等体之间交换路由信息。一条 Update 消息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 消息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 消息里的路由属性适用于该 Update 消息中的 NLRI（Network Layer Reachability Information）字段里的所有目的地（用 IP 前缀表示）。
- 一条 Update 消息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），清楚的定义了 BGP Speaker 之间先前通告过的路由。
- 一条 Update 消息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。
- Keepalive 消息：BGP 会周期性的向对等体发出 Keepalive 消息，用来保持连接的有效性。
- Notification 消息：当 BGP 检测到错误状态时，就向对等体发出 Notification 消息，之后 BGP 连接会立即中断。
- Route-Refresh 消息：通过 OPEN 消息告知 BGP peer 本地支持路由刷新能力（Route-Refresh capability）。在所有 BGP 路由器使能 Route-Refresh 能力的情况下，如果 BGP 的入口路由策略发生了变化，本地 BGP 路由器会向对等体发布 Route-Refresh 消息，收到此消息的对等体会将其路由信息重新发给本地 BGP 路由器。这样，可以在不中断 BGP 连接的

情况下，对 BGP 路由表进行动态刷新，并应用新的路由策略。

- BGP 报文应用：
- BGP 使用 TCP 建立连接，本地监听端口为 179。和 TCP 连接建立相同，BGP 连接的建立也要经过一系列的对话和握手。TCP 通过握手协商通告其端口等参数，BGP 的握手协商的参数有：BGP 版本、BGP 连接保持时间、本地的路由器标识（Router ID）、授权信息等。这些信息都在 Open 消息中携带。
- BGP 连接建立后，如果有路由需要发送则发送 Update 消息通告对端。Update 消息发布路由时，还要携带此路由的路由属性，用以帮助对端 BGP 协议选择最优路由。在本地 BGP 路由变化时，要通过 Update 消息来通知 BGP 对等体。
- 经过一段时间的路由信息交换后，本地 BGP 和对端 BGP 都无新路由通告，趋于稳定状态。此时要定时发送 KEEPALIVE 消息以保持 BGP 连接的有效性。对于本地 BGP，如果在保持时间内，未收到任何对端发来的 BGP 消息，就认为此 BGP 连接已经中断，将断开此 BGP 连接，并删除所有从该对等体学来的 BGP 路由。
- 当本地 BGP 在运行中发现错误时（如对端 BGP 版本本地不支持、本地 BGP 收到了结构非法的 Update 消息等），要发送 Notification 消息通告 BGP 对等体。本地 BGP 退出 BGP 连接时，也需发送 Notification 报文。

- BGP 报头
- Marker（标记）：16 字节，固定为 1。
- Length（长度）：两字节无符号整数。指定了消息的全长，包括头部。
- Type（类型）：1 字节，指示报文类型：
- Open

- Update
- Keepalive
- Notification
- Route-Refresh

- Open 报文结构
- Version：BGP 的版本号。对于 BGPv4 来说，其值为 4。
- My Autonomous System：本地 AS 编号。通过比较两端的 AS 编号可以确定是 EBGP 连接还是 IBGP 连接。
- Hold Time：在建立对等体关系时两端要协商 Hold time，并保持一致。如果两端所配置的 Hold time 时间不同，则 BGP 会选择较小的值作为协商的结果。如果在这个时间内未收到对端发来的 Keepalive 消息，则认为 BGP 连接中断。如果保持时间为 0，则标识不发送 Keepalive 报文。
- BGP Identifier：BGP 路由器的 Router ID，以 IP 地址的形式表示，用来识别 BGP 路由器。
- Opt Parm Len（Optional Parameters Length）：可选参数的长度。如果为 0 则没有可选参数。
- Optional Parameters：是一个可选参数用于 BGP 验证或多协议扩展（Multiprotocol Extensions）等功能。每一个参数为一个（Parameter Type-Parameter Length-Parameter Value）三元组。

- Update 报文结构
- Withdrawn Routes Length ：（2 字节无符号整数）不可达路由长度，表示 Withdrawn Routes 字段的数据长度。如果 Withdrawn Routes Length 字段数值为 0，则表示 Withdrawn Routes 字段没有任何数据，在 UPDATE 消息中不会被显示。
- Withdrawn Routes ：（变长）撤销路由。该字段包括一

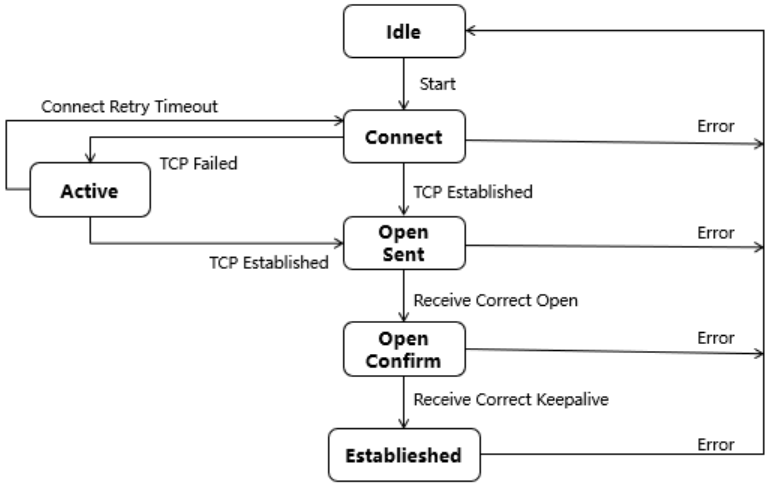
系列的 IP 地址前缀信息，以<length, prefix>的格式来表示，比如<19,198.18.160.0>表示一个 198.18.160.0 255.255.224.0 的网络。

- Path Attribute Length ：（2 字节无符号整数）路由属性长度，表示 Path Attribute 字段的数据长度。如果 Path Attribute Length 数值为 0，则表示 Path Attribute 字段没有任何数据，在 UPDATE 消息中不会被显示。
- Network Layer Reachability Information ：（变长）网络可达信息。包括一系列的 IP 地址前缀。格式与撤消路由字段一样<length, prefix>。

- Keepalive 报文结构
- KeepAlive 报文的组成只包括一个 BGP 数据报头。
- 缺省情况下，发送 KeepAlive 的时间间隔为 60 秒，Hold Time 是 180 秒。每次从邻居处接收到 KeepAlive 报文将重置 Hold Time 定时器，如果 Hold Time 定时器超时，就认为对等体 Down 掉。

- Notification 报文结构
- Errorcode：错误码。1 字节长的字段。每个不同的错误都使用唯一的代码表示，而每一个错误码都可以拥有一个或多个错误子码，但如果某些错误码并不存在错误子码的话，则该错误子码字段以全 0 表示。
- Errsubcode：错误子码。

BGP工作原理—状态机



- BGP 有限状态机共有六种状态，分别是 Idle、Connect、Active、OpenSent、OpenConfirm 和 Established。
- 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 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 报文或 TCP 拆除链路通知等 Error 事件后，BGP 都会转至 Idle 状态。
- 在 Connect 状态下，BGP 启动连接重传定时器（Connect Retry，缺省为 32 秒），等待 TCP 完成连接。
- 此阶段主动发起 TCP 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 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。

- 如果发生其他事件（由系统或者操作人员启动的），则退回到 Idle 状态。
- 在 Active 状态下，BGP 总是在试图建立 TCP 连接。
- 此阶段等待对方发起 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。
- 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。
- 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。
- 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 连接断开消息，那么 BGP 断开连接，转至 Idle 状态。

BGP工作原理—数据库

- IP路由表 (IP-RIB)

  全局路由信息库，包括所有IP路由信息

- BGP路由表 (Loc-RIB)

  BGP路由信息库，包括本地BGP Speaker选择的路由信息

- 邻居表

  对等体邻居清单列表

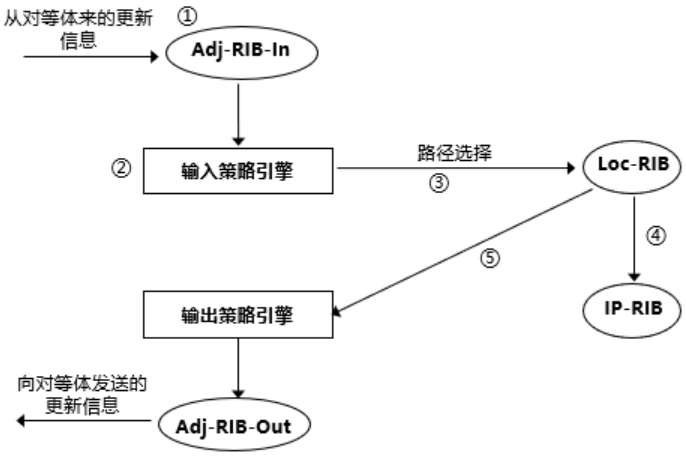
- Adj-RIB-In

  对等体宣告给本地BGP Speaker的未处理的路由信息库

- Adj-RIB-Out

  本地BGP Speaker宣告给指定对等体的路由信息库

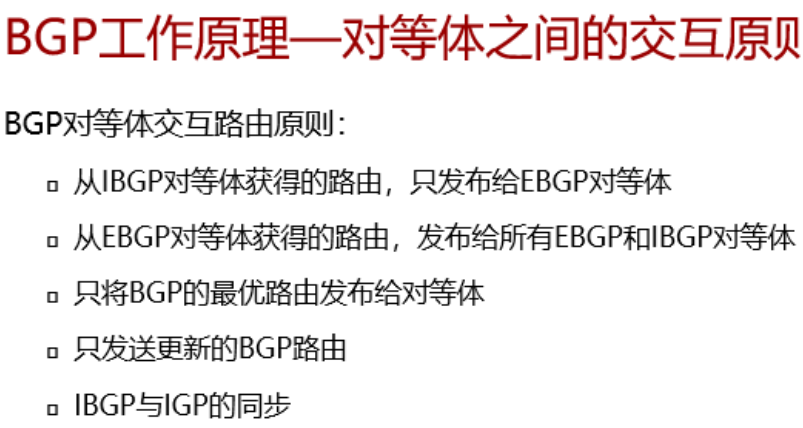
BGP工作原理—BGP路由信息处理



- BGP 路由信息处理：
- 当从对等体接收到更新数据包时，路由器会把这些更新数据包存储到路由选择信息库(Routing Information Base, RIB)中，并指明是来自哪个对等体的(Adj-RIB-In)。这些更新数据包被输入策略引擎过滤后，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 RIB (Loc-RIB)中，

然后被提交给本地 IP 路由选择表(IP-RIB)，以用作安装考虑。

- 除了从对等体接收来的最佳路径外，Loc-RIB 也会包含当前路由器注入的(被称为本地发起的路由)，并被选择为最佳路径的 BGP 前缀。Loc-RIB 中的内容在被通告给其他对等体之前，必须通过输出策略引擎。只有那些成功通过输出策略引擎的路由，才会被安装到输出 RIB (Adj-RIB-Out)中。

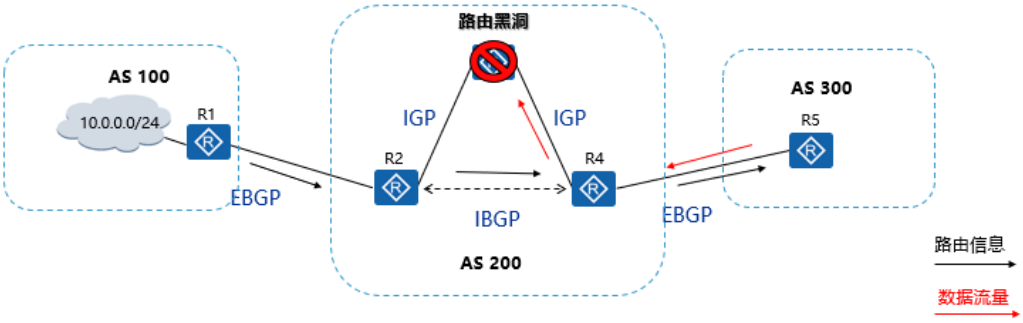


- BGP 设备将最优路由加入 BGP 路由表，形成 BGP 路由。
- 从 IBGP 对等体获得的 BGP 路由，BGP 设备只发布给它的 EBGP 对等体
- 从 EBGP 对等体获得的 BGP 路由，BGP 设备发布给它所有 EBGP 和 IBGP 对等体
- 当存在多条到达同一目的地址的有效路由时，BGP 设备只将最优路由发布给对等体
- 路由更新时，BGP 设备只发送更新的 BGP 路由
- 从 IBGP 邻居学到的路由，只有当 IGP 中也存在相同的路由时才会宣告给 EBGP 对等体

BGP工作原理—IBGP与IGP同步

- 同步（Synchronization）

在IBGP路由加入路由表并发布给EBGP对等体之前，会先检查IGP路由表。只有在IGP也知道这条IBGP路由时，它才会被加入到路由表，并发布给EBGP对等体。



- 同步是指 IBGP 和 IGP 之间的同步，其目的是避免误导外部 AS 的路由器。

- 拓扑说明（在同步开启情况下）
- R4 通过 BGP 学习到 R1 宣告的 10.0.0.0/24 网络。R4 在将该网络通告给 R5 之前，会首先检查自己的 IGP 路由表是否已经存在 10.0.0.0/24 网络。如果 R4 本地 IGP 路由表项存在 10.0.0.0/24 网络，则将该网络通告给 R5；如果 R4 本地 IGP 路由表项不存在 10.0.0.0/24 网络，则不能将该网络通告给 R5。

- 注意事项：
- VRP 平台缺省情况下 BGP 与 IGP 是取消同步机制的，并不可改变。但取消同步是有条件的，在以下两种情况下可以取消同步：
- 本 AS 不是过渡 AS。
- 本 AS 内所有路由器建立 IBGP 全连接。

BGP属性特点—概述

- BGP路由属性是一套参数，它是对路由的进一步的描述
- 公认必遵
  - 所有BGP路由器都必须识别，且必须存在于Update消息中
  - 如果缺少这种属性，路由信息就会出错
- 公认任意
  - 所有BGP路由器都可以识别，但不要求必须存在于Update消息中
  - 即就算缺少这类属性，路由信息也不会出错
- 可选过渡
  - 在BGP对等体之间具有可传递性的属性
  - BGP路由器可以不支持此属性，但它仍然会接收这类属性，并传递给其他对等体
- 可选非过渡
  - 如果BGP路由器不支持此属性，则相应的这类属性会被忽略，且不会传递给其他对等体

- BGP 路由属性是一套参数，它对特定的路由进一步的描述，使得 BGP 能够对路由进行过滤和选择。

- 常用的属性类别如下所示：
- Origin 为公认必遵属性
- AS_Path 为公认必遵属性
- Next_Hop 为公认必遵属性
- Local_Pref 为公认任意属性
- community 为可选过渡属性
- MED 为可选非过渡属性
- Originator_ID 为可选非过渡属性
- Cluster_List 为可选非过渡属性

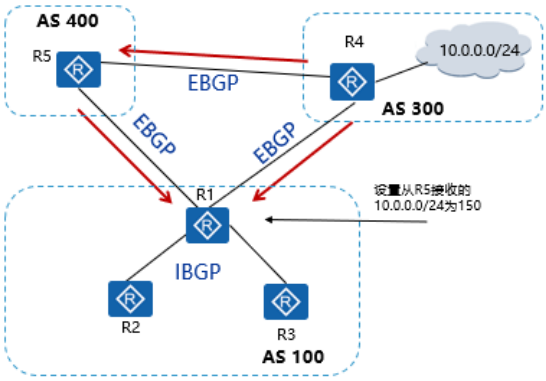
BGP属性特点—Origin

- Origin属性用来定义路径信息的来源，该属性为公认必遵
  - IGP
    - 通过路由始发AS的IGP得到的路由信息，如通过**network**命令注入BGP的路由
    - 标识符为“i”
  - EGP
    - 通过EGP得到的路由信息
    - 标识符为“e”
  - Incomplete
    - 通过其他方式学习到的路由信息，如通过**import-route**命令注入BGP的路由
    - 标识符为“?”

- Origin 属性用来定义路径信息的来源，标记一条路由是怎么成为 BGP 路由的。它有以下 3 种类型：
- IGP：具有最高的优先级。通过路由始发 AS 的 IGP 得到的路由信息，比如通过 network 命令注入到 BGP 路由表的路由，其 Origin 属性为 IGP。
- EGP：优先级次之。通过 EGP 得到的路由信息，其 Origin 属性为 EGP。
- Incomplete：优先级最低。通过其他方式学习到的路由信息。比如 BGP 通过 import-route 命令引入的路由，其 Origin 属性为 Incomplete。

BGP属性特点—PrefVal

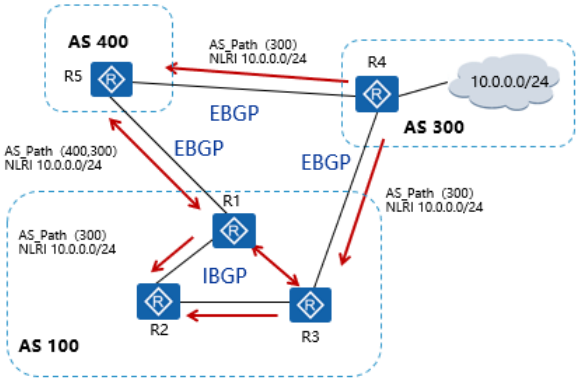
- 协议首选值（PrefVal）是华为设备的特有属性，该属性仅在本地有效，不会传递给BGP邻居。因为协议首选值是人为主动设置的，代表本地用户的意愿，因而在BGP进行选路时会优先比较协议首选值。



- 在 BGP 进行选路时会优先比较协议首选值。默认情况下均为 0，该值越大越优先。

BGP属性特点—AS_Path

- AS_Path属性按矢量顺序记录某条路由从本地到目的地址所要经过的所有AS编号。该属性为公认必遵。



- AS_Path 属性可以当做 BGP 选路的参考属性之一，AS_Path 长度越短越优先。此外，当 BGP 路由器从 EBGP 对等体接收路由时，如果发现 AS_Path 列表中有本 AS 号，则不接收该路由，从而避免了 AS 间的路由环路。

- 当 BGP Speaker 本地通告一条路由时：

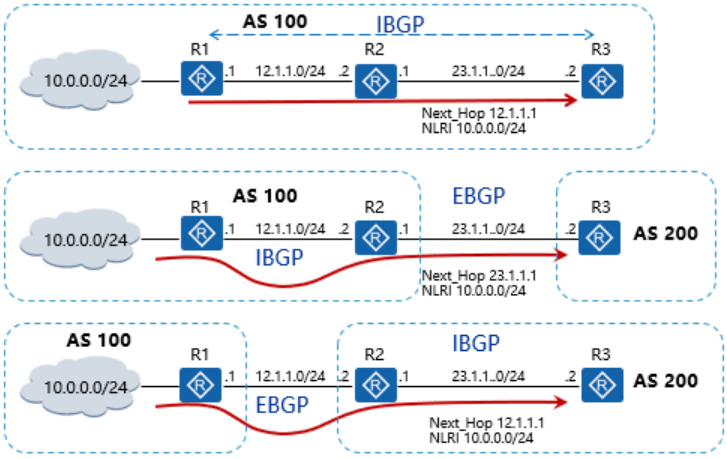
- 当 BGP Speaker 将这条路由通告到其他 AS 时，便会将本地 AS 号添加在 AS_Path 列表中，并通过 Update 消息通告给邻居路由器。
- 当 BGP Speaker 将这条路由通告到本地 AS 时，便会在 Update 消息中创建一个空的 AS_Path 列表。

- 当 BGP Speaker 传播从其他 BGP Speaker 的 Update 消息中学习到的路由时：
- 当 BGP Speaker 将这条路由通告到其他 AS 时，便会把本地 AS 编号添加在 AS_Path 列表的最前面（最左面）。收到此路由的 BGP 路由器根据 AS_Path 属性就可以知道去目的地址所要经过的 AS。离本地 AS 最近的相邻 AS 号排在前面，其他 AS 号按顺序依次排列。
- 当 BGP Speaker 将这条路由通告到本地 AS 时，不会改变这条路由相关的 AS_Path 属性。

- 拓扑描述：
- 当 R4 将网段 10.0.0.0/24 通告给 AS400 和 AS100 时，会在 AS_PATH 中添加自己的 AS 号。当 R5 将网段 10.0.0.0/24 通告给 AS100 时，也会添加添加自己的 AS 号。当 AS100 内的 R1、R3 和 R2 之间将网段 10.0.0.0/24 相互通告时，AS_PATH 属性不会改变，在其他 BGP 选路条件相同的前提下，BGP 会选择 AS_PATH 路径最短的，即选择通过 R3 直达 R4 的路由。

BGP属性特点—Next_Hop

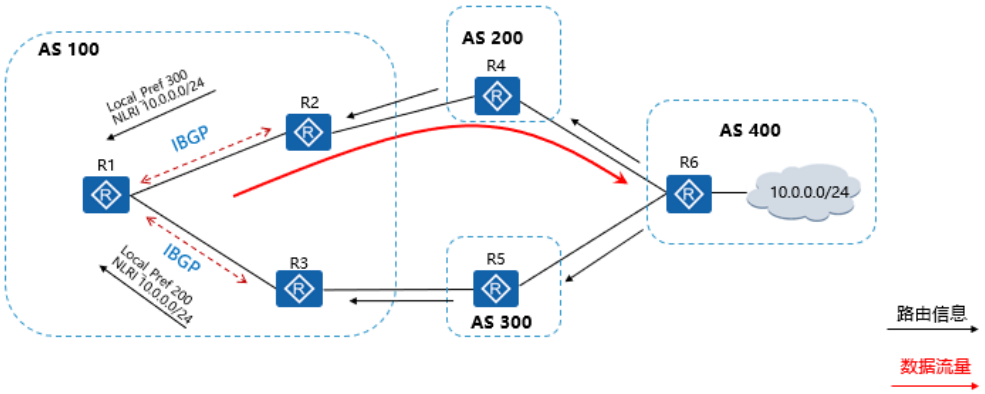
- Next_Hop属性记录了路由的下一跳信息，该属性为公认必遵



- Next_Hop 属性记录了路由的下一跳信息。BGP 的下一跳属性和 IGP 的有所不同，不一定就是邻居设备的 IP 地址。通常情况下，Next_Hop 属性遵循下面的规则：
- BGP Speaker 将本地始发路由发布给 IBGP 对等体时，会把该路由信息的下一跳属性设置为本地与对端建立 BGP 邻居关系的接口地址。
- BGP Speaker 在向 EBGP 对等体发布某条路由时，会把该路由信息的下一跳属性设置为本地与对端建立 BGP 邻居关系的接口地址。
- BGP Speaker 在向 IBGP 对等体发布从 EBGP 对等体学来的路由时，并不改变该路由信息的下一跳属性。

BGP属性特点—Local_Pref

- Local_Pref属性表明BGP路由器的优先级，该值越大越优先。该属性为公认任意。

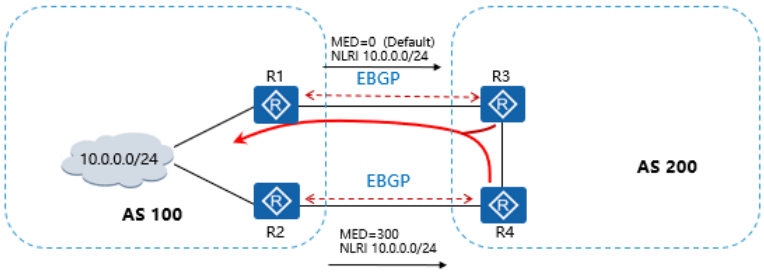


- Local_Pref 属性
- 该属性仅在 IBGP 对等体之间有效，不通告给其他 AS。它表明路由器的 BGP 优先级。
- 当 BGP 路由器通过不同的 IBGP 对等体得到目的地址相同但下一跳不同的多条路由时，将优先选择 Local_Pref 属性值较高的路由，缺省情况下该值为 100。

- 拓扑描述
- AS100 内，R1，R2 和 R3 之间分别两两建立 IBGP 对等体关系，而 R2 和 R3 分别和位于 AS200 和 AS300 的路由器建立 EBGP 对等体关系。这样路由器 R2 和 R3 都会从自己的 EBGP 对等体收到 10.0.0.0/24 这条路由，为了让 AS100 内的三台路由器优选 R2 作为 10.0.0.0/24 这条路由在本 AS 的出口，我们只需要在 R2 和 R3 上适当的对该路由的 Local Pref 属性进行修改，就可以达到目的。

BGP属性特点—MED

- MED属性类似于IGP的代价值，用于AS间的路由选路。该属性为可选非过渡



路由信息

数据流量

- 当一个运行 BGP 的设备通过不同的 EBGP 对等体（EBGP 对等体需属于同一 AS）得到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。

- MED 属性仅在相邻两个 AS 之间传递，收到此属性的 AS 一方不会再将其通告给任何其他第三方 AS。MED 属性可以手动配置，如果路由没有配置 MED 属性，BGP 选路时将该路由的 MED 值按缺省值 0 来处理。

- 拓扑描述
- R1 和 R2 将网段 10.0.0.0/24 传递给各自的 EBGP 邻居 R3 和 R4，R3 和 R4 在其他条件相同的情况下，优先选择 MED 值较低的路径，即均选择经由 R1 访问网络 10.0.0.0/24。

BGP路由计算—选路规则

当到达同一目的地存在多条路由时，BGP依照如下策略顺序进行路由选择：

- 如果此路由的下一跳不可达，忽略此路由
- 优选协议首选值（PrefVal）最高的路由
- 优选本地优先级（Local_Pref）最高的路由
- 优选手动聚合路由、自动聚合路由、network命令引入的路由、import-route命令引入的路由、从对等体学习的路由
- 优选AS路径（AS_Path）最短的路由
- 比较Origin属性，依次优选Origin类型为IGP、EGP、Incomplete的路由
- 优选MED值最低的路由
- 优选从EBGP邻居学来的路由（EBGP路由优先级高于IBGP路由）
- 优选到下一跳IGP Metric较小的路由
- 优选Cluster_List最短的路由
- 优选Router ID最小的路由器发布的路由
- 比较对等体的IP Address，优选从具有较小IP Address的对等体学来的路由

- BGP 路径选择
- 下一跳地址必须可达。
- 协议首选值（PrefVal）是华为设备的特有属性，该属性仅在本地有效。
- 如果路由没有本地优先级，BGP 选路时将该路由按缺省的本地优先级 100 来处理。通过执行 default local-preference 命令可以修改 BGP 路由的缺省本地优先级。
- 本地生成的路由包括通过 network 命令或 import-route 命令引入的路由、手动聚合路由和自动聚合路由。
- 优选聚合路由（聚合路由优先级高于非聚合路由）。
- 通过 aggregate 命令生成的手动聚合路由的优先级高于通过 summary automatic 命令生成的自动聚合路由。
- 通过 network 命令引入的路由的优先级高于通过 import-route 命令引入的路由。
- 优选 AS 路径（AS_Path）最短的路由。
- AS_Path 的长度不包括 AS_CONFED_SEQUENCE 和 AS_CONFED_SET。
- AS_SET 的长度为 1，无论 AS_SET 中包括多少 AS 号。
- 执行 bestroute as-path-ignore 命令后，BGP 选路时，忽

略 AS_Path 的比较。

- 优选 MED 值最低的路由。
- BGP 只比较来自同一个 AS（不包括联盟的子 AS）的路由的 MED 值。即，只有两条路由的 AS_SEQUENCE（不包括 AS_CONFED_SEQUENCE）属性的第一个 AS 号相同时，BGP 才会比较二者的 MED 值。
- 如果路由没有 MED 属性，BGP 选路时将该路由的 MED 值按缺省值 0 来处理；执行 bestroute med-none-as-maximum 命令后，BGP 选路时将该路由的 MED 值按最大值 4294967295 来处理。
- 执行 compare-different-as-med 命令后，BGP 将强制比较来自不同自治系统中的邻居的路由的 MED 值。除非能够确认不同的自治系统采用了同样的 IGP 和路由选择方式，否则不要使用 compare-different-as-med 命令（可能产生环路）。
- 执行 bestroute med-confederation 命令后，只有当 AS_Path 中不包含外部 AS 号（不属于联盟的子 AS），且 AS_CONFED_SEQUENCE 的第一个 AS 号相同时，才能比较 MED 值的大小。
- 执行 deterministic-med 命令后，将消除路由接收顺序对选路结果的影响。

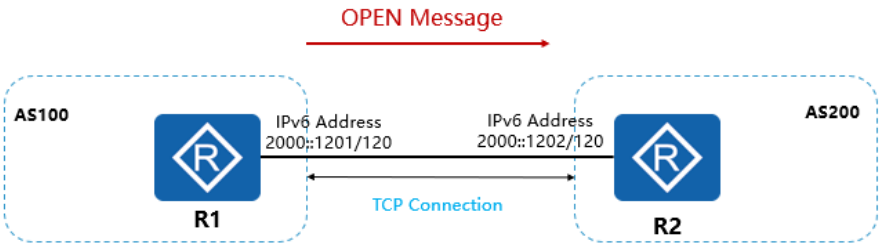
- 负载分担
- 当到达同一目的地址存在多条等价路由时，可以通过 BGP 等价负载分担实现均衡流量的目的。
- 形成 BGP 等价负载分担的条件是：BGP 选路规则中“到下一跳的 IGP metric”这条规则之前所有需要比较的属性完全相同。

BGP 4+ 概述

- BGP 4+ 概述
  - BGP 4 的扩展版本
  - 扩展能力自协商机制
  - 支持传递多种地址族地址（IPv6、VPNv4、VPNv6等）
  - 新增属性用以支持多地址族地址传递

BGP 4+ 扩展能力协商（1/2）

- 当两台BGP对等体之间需要传输IPv6地址族的地址时，需要在OPEN Message 中进行扩展能力的协商



BGP 4+ 扩展能力协商（2/2）

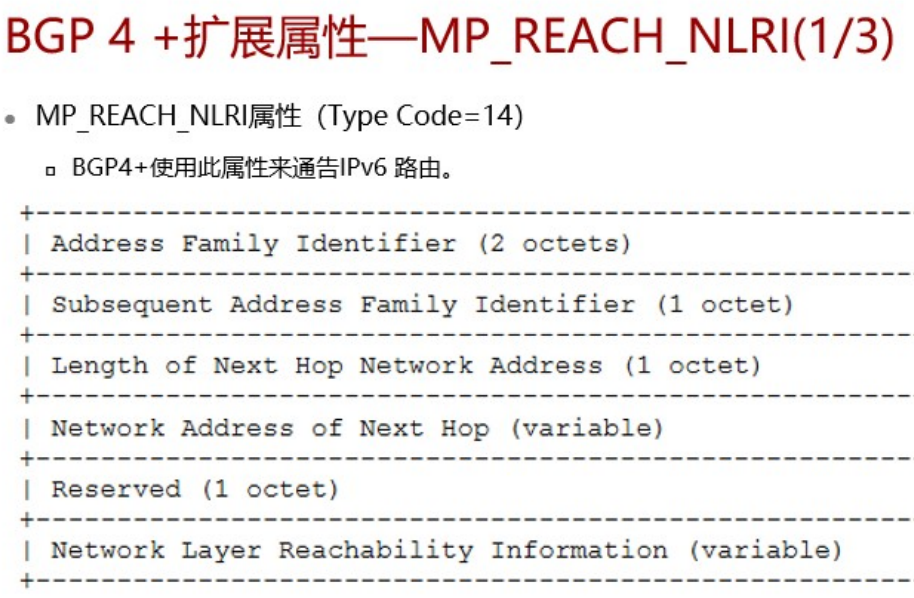
- OPEN消息中的Capabilities Advertisement字段用于扩展能力的协商。

```
Frame 10: 131 bytes on wire (1048 bits), 131 bytes captured (1048 bits)
Ethernet II, Src: HuaweiTe_69:32:1c (54:89:98:69:32:1c), Dst: HuaweiTe_2d:53:8a (54:89:98:2d:53:8a)
Internet Protocol Version 6, Src: 2000::1201 (2000::1201), Dst: 2000::1202 (2000::1202)
Transmission Control Protocol, Src Port: bgp (179), Dst Port: 49153 (49153), Seq: 1, Ack: 46, Len: 45
Border Gateway Protocol
  OPEN Message
    Marker: 16 bytes
    Length: 45 bytes
    Type: OPEN Message (1)
    Version: 4
    My AS: 100
    Hold time: 180
    BGP identifier: 10.0.1.1
    Optional parameters length: 16 bytes
    Optional parameters
      Capabilities Advertisement (16 bytes)
        Parameter type: Capabilities (2)
        Parameter length: 14 bytes
        Multiprotocol extensions capability (6 bytes)
          Capability code: Multiprotocol extensions capability (1)
          Capability length: 4 bytes
          Capability value
            Address family identifier: IPv6 (2)
            Reserved: 1 byte
            Subsequent address family identifier: Unicast (1)
        Route refresh capability (2 bytes)
        Support for 4-octet AS number capability (6 bytes)
```

- 除了多地址族的能力协商外，还有
- 4 字节 AS 号能力

- Route-Refresh 支持能力
- 多层标签能力

等能力都会在该字段列出来进行协商。

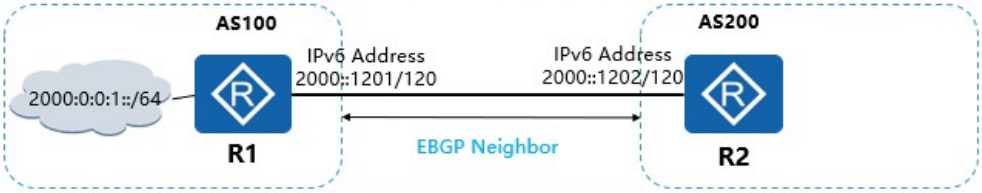


- 地址族信息（Address Family Information）域：由 2 字节的地址族标识 AFI（Address Family Identifier）和 1 字节的子地址族标识 SAFI（Subsequent Address Family Identifier）组成。
- 下一跳长度（Length of Next Hop Network Address）域：1 字节长度，表示下一跳地址的长度，通常情况下为 16
- 下一跳地址（Network Address of Next Hop）域：长度由上一个字段决定，一般情况下为全球单播地址。
- 保留字段（Reserved）域：一字节，必须为 0
- 网络层可达信息（Network Layer Reachability Information）域：表示含有匹配相同属性的路由信息。当此字段为 0 时，表示为缺省路由。

BGP 4 +扩展属性—MP_REACH_NLRI(2/3)

- 当传递IPv6路由时
  - AFI=2 ， SAFI=1(Unicast) ,SAFI=2(Multicast)
  - 下一跳地址长度字段决定了下一跳地址的个数
    - 长度字段=16，下一跳地址为下一跳路由器的全球单播地址
    - 长度字段=32，下一跳地址为下一跳路由器的全球单播地址和链路本地地址
  - 保留字段，恒等于0
  - NLRI字段，可变长字段，表示路由前缀和掩码信息。

BGP 4 +扩展属性—MP_REACH_NLRI(3/3)




```
Border Gateway Protocol
  UPDATE Message
    Marker: 16 bytes
    Length: 77 bytes
    Type: UPDATE Message (2)
    Unfeasible routes length: 0 bytes
    Total path attribute length: 54 bytes
  Path attributes
    ORIGIN: IGP (4 bytes)
    AS_PATH: 100 (9 bytes)
    MULTI_EXIT_DISC: 0 (7 bytes)
    MP_REACH_NLRI (34 bytes)
      Flags: 0x90 (Optional, Non-transitive, Complete, Extended Length)
      Type code: MP_REACH_NLRI (14)
      Length: 30 bytes
      Address family: IPv6 (2)
      Subsequent address family identifier: Unicast (1)
      Next hop network address (16 bytes)
        Next hop: 2000::1201 (16)
      Subnetwork points of attachment: 0
      Network layer reachability information (9 bytes)
        2000:0:0:1::/64
```

BGP 4 +扩展属性—MP_UNREACH_NLRI(1/2)

- MP_UNREACH_NLRI属性(Type Code= 15)
  - BGP 4+用该属性撤销路由

```
+-----------------------------------------------------+
| Address Family Identifier (2 octets)                |
+-----------------------------------------------------+
| Subsequent Address Family Identifier (1 octet)      |
+-----------------------------------------------------+
| Withdrawn Routes (variable)                         |
+-----------------------------------------------------+
```

- 地址族信息（Address Family Information）域：由 2 字节的地址族标识 AFI（Address Family Identifier）和 1 字节的子地址族标识 SAFI（Subsequent Address Family Identifier）组成。
- 撤销路由（Withdrawn Routes）域：表示撤销的路由条目。格式为<掩码长度，路由前缀> ，当此掩码长度为 0 时，表示为缺省路由。

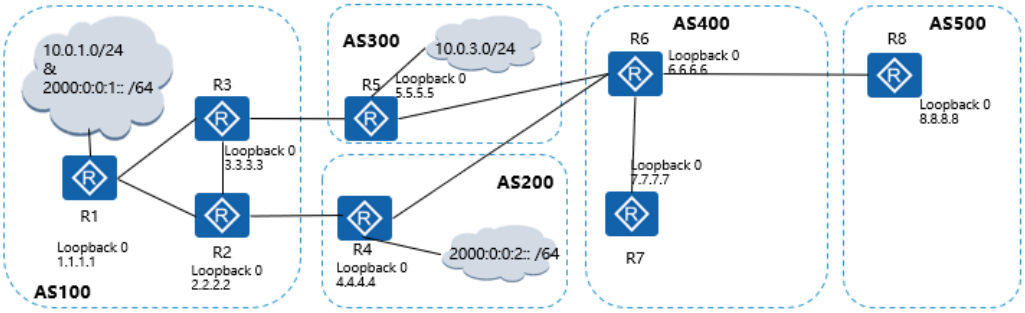
BGP 4 +扩展属性—MP_UNREACH_NLRI(2/2)

- 当撤销IPv6路由时
  - AFI=2， SAFI=1(Unicast) ,SAFI=2(Multicast)
  - Withdrawn Routes 字段代表需要撤回的路由前缀及掩码。

```
⊟ Border Gateway Protocol
  ⊟ UPDATE Message
      Marker: 16 bytes
      Length: 39 bytes
      Type: UPDATE Message (2)
      Unfeasible routes length: 0 bytes
      Total path attribute length: 16 bytes
    ⊟ Path attributes
      ⊟ MP_UNREACH_NLRI (16 bytes)
        ⊞ Flags: 0x90 (Optional, Non-transitive, Complete, Extended Length)
          Type code: MP_UNREACH_NLRI (15)
          Length: 12 bytes
          Address family: IPv6 (2)
          Subsequent address family identifier: Unicast (1)
        ⊟ Withdrawn routes (9 bytes)
          ⊞ 2000:0:0:1::/64
```

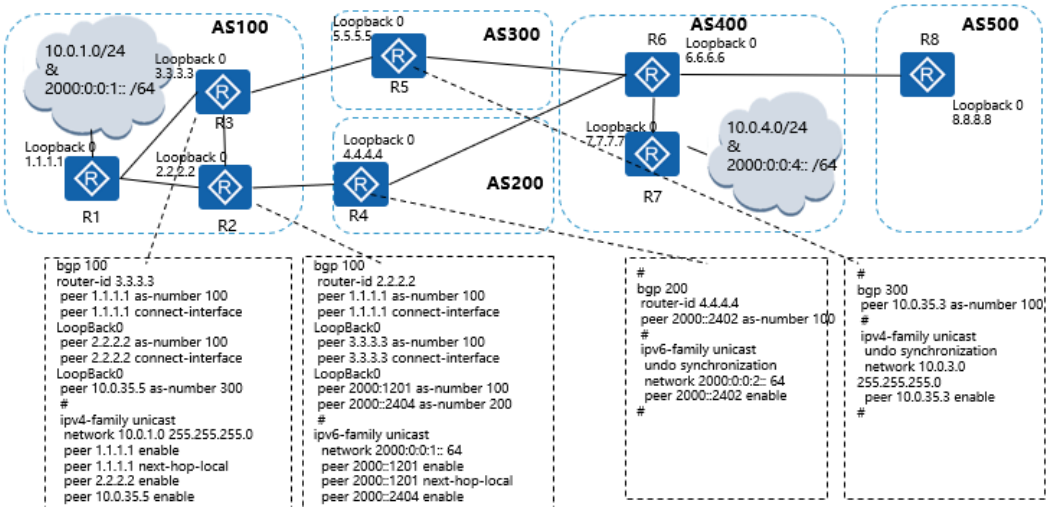
配置BGP基本功能

- 配置需求：①各AS之间的直连边界路由之间建立EBGP邻居关系
  ②AS内建立稳定的IBGP邻居关系
  ③所有的IPv4和IPv6网络之间能够互通
  ④AS内部通过IGP实现互通



- 地址分配规则：
- Rx 和 Ry（X<Y）直连接口的 IPv4 网段为：10.0.xy.0/24. Rx 相应接口的地址为 10.0.xy.x ，Ry 为 10.0.xy.y
- Rx 和 Ry（X<Y）直连接口的 IPv6 网段为：2000::xy00/120. Rx 相应接口的地址为 2000::xy0x ，Ry 为 2000::xy0y
- 各路由器的 LoopBack 0 接口的地址已给出,各 LoopBack 0 接口的 IPv6 地址为 2000::z(z 为相应路由器的编号)
- 提示：
- AS 内可运行 OSPF，ISIS 等协议来实现互通
- 稳定的 IBGP 关系可通过 loopback 接口来建立
- EBGP 邻居关系直接用物理接口建立即可

配置BGP基本功能（续）



- 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。

- 具体用法
- 上述命令均为 BGP 进程视图下的命令

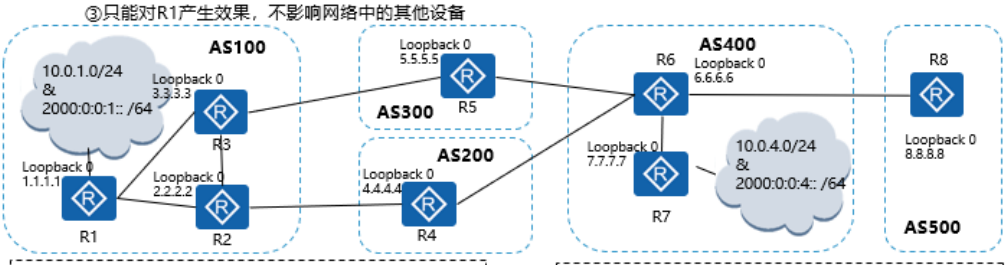
- 参数意义
- peer ipv4-address as-number as-number
- ip-address：对等体的 IPv4 地址。
- as-number：对等体的对端 AS 号。
- peer ipv4-address connect-interface interface-type interface-number [ipv4-source-address]
- ip-address：对等体的 IPv4 地址。
- interface-type interface-number：接口类型和接口号。
- ipv4-source-address：建立连接时的 IPv4 源地址。

- peer ipv4-address next-hop-local
- ip-address：对等体的 IPv4 地址。

- 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 PrefRcv 表示本端从对等体上收到路由前缀的数目。
- IPv6 的配置与 IPv4 基本一致，但是在指定完 peer 地址和 as-number 之后，需手工进入 ipv6-family unicast 视图，执行 peer peer-ip-address enable 命令来激活。
- 本页图中的互连接口 IPv6 地址，掩码位是 112。

配置BGP– PrefVal属性

- 配置需求：①从R1去往AS400中的10.0.4.0/24网段走路径R1-R2-R4-R6-R7
  ②从R1去往AS400中的2000:0:0:4::/64网段走路径R1-R3-R5-R6-R7
  ③只能对R1产生效果，不影响网络中的其他设备

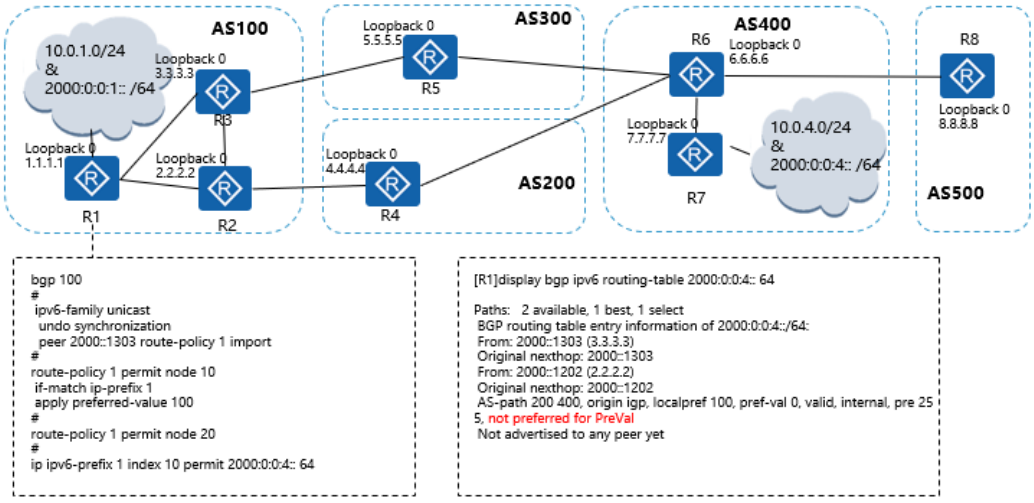



```
<R1>display bgp ipv6 routing-table 2000:0:0:4::64
BGP routing table entry information of 2000:0:0:4::/64:
 From: 2000::1202 (2.2.2.2)
  Original nexthop: 2000::1202
 AS-path 200 400, origin igp, localpref 100, pref-val 0, valid, internal, best,
select, active, pre 255
From: 2000::1303 (3.3.3.3)
Original nexthop: 2000::1303
 AS-path 300 400, origin igp, localpref 100, pref-val 0, valid, internal, pre 25
5, not preferred for router ID
```

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

*>i 10.0.4.0/24  2.2.2.2                    100       0       200 400i
* i             3.3.3.3                    100       0       300 400i
```

- 此拓扑与“基础配置”一致，已建立了 BGP 邻居关系

配置BGP– PrefVal属性（续）



- 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。
- **apply preferred-value** preferred-value　命令用来在路由策略中配置改变 BGP 路由的首选值的动作。

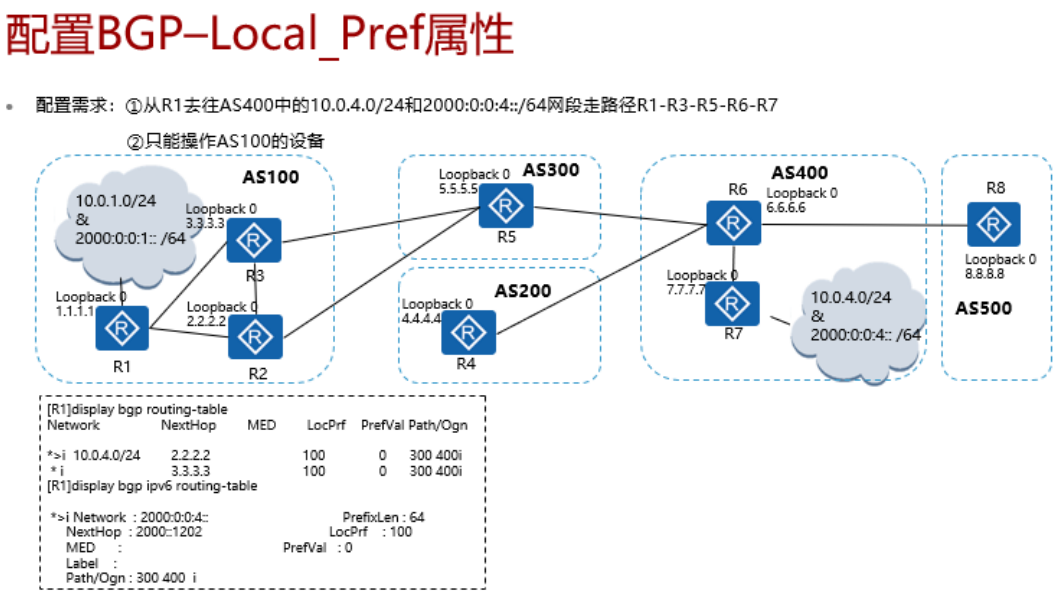
- 具体用法
- peer route-policy 命令为 BGP 视图命令

- 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。
- preferred-value：指定 BGP 的首选值。在选择路由时，协议优选首选值最高的 BGP 路由。整数形式，取值范围 0～6

5535，默认为 0

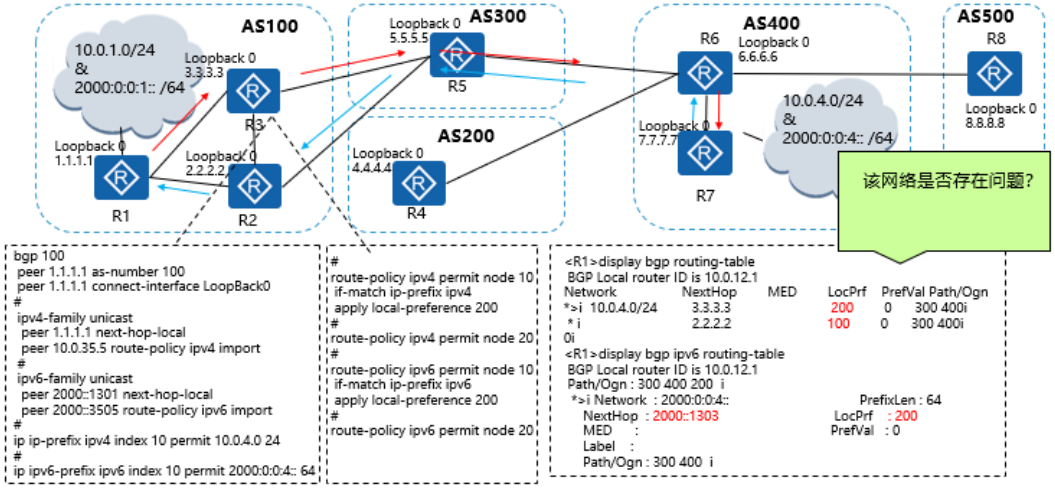
- 实验现象
- 我们使用 display bgp routing-table 和 display bgp ipv6 routing-table 命令查看 BGP 路由表。

- 注意事项
- Preferred-value 是 BGP 协议的私有属性，该命令只对 BGP 路由生效。Preferred-value 是 BGP 选路规则中的 weight 值，不是 RFC 规定的标准属性，所以该命令仅在本地生效，在 BGP 的出口策略中不生效。



- 沿用 BGP“基础配置”拓扑，仅建立了 BGP 邻居关系。

配置BGP-Local_Pref属性（续）



- 命令含义：
- **apply local-preference** preference　配置路由的本地优先级

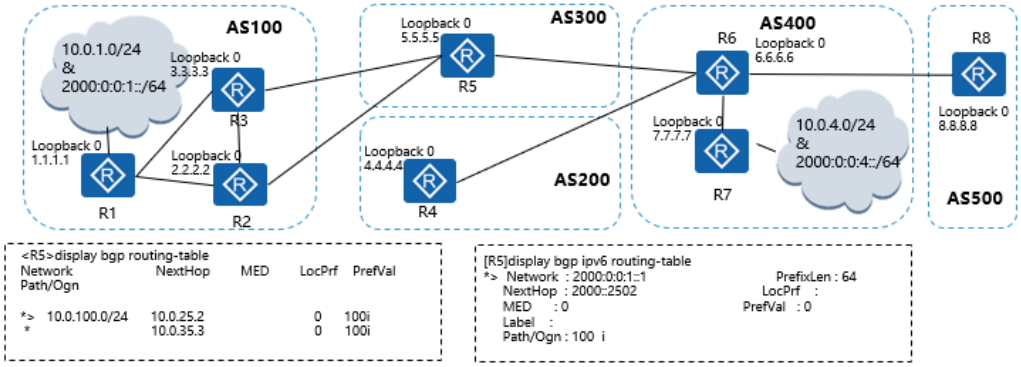
- 参数含义：
- Preference　指定的 BGP 路由的本地优先级，整数形式，取值范围是 0～4294967295，默认情况下为 100

- 注意事项：
- 策略生效后，将影响 BGP 路由选路。

- 本地优先级仅用于同一个 AS 域内的选路，不向域外发布这个属性，所以用于配置 EBGP 邻居的 export 方向的策略时，apply local-preference 命令的设置不生效。

配置BGP– MED属性

- 配置需求：①解决上一步网络出现的问题
  ②只能操作AS100的设备

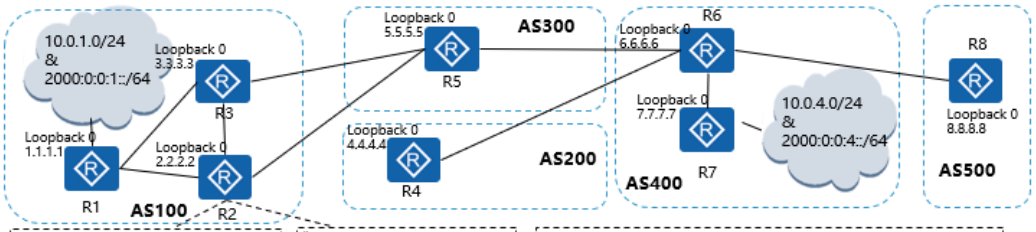



```
<R5>display bgp routing-table
Network                 NextHop        MED        LocPrf    PrefVal
Path/Ogn

*>  10.0.100.0/24       10.0.25.2                 0         100i
 *                      10.0.35.3                 0         100i
```

```
[R5]display bgp ipv6 routing-table
*>  Network  : 2000:0:0:1::1                    PrefixLen : 64
    NextHop  : 2000::2502                       LocPrf    :
    MED      : 0                                PrefVal   : 0
    Label    :
    Path/Ogn : 100  i
```

- 保留上一步 local-preference 的配置，其他不变。
- 需要解决上一步出现的来回路径不一致的问题，可通过 R2 宣告更高 MED 属性的路由，使得 R5 选择 R3 宣告的路由。

配置BGP– MED属性（续）




```
bgp 100
 router-id 2.2.2.2
#
 ipv4-family unicast
peer 10.0.35.5 route-policy ipv4 export
#
 ipv6-family unicast
peer 2000::3505 route-policy ipv6 export
#
ip ip-prefix ipv4 index 10 permit 10.0.1.0
24
ip ipv6-prefix ipv6 index 10 permit
2000:0:0:1::64
```

```
#
route-policy ipv4 permit node 10
 if-match ip-prefix ipv4
 apply cost 100
#
route-policy ipv4 permit node 20
#
route-policy ipv6 permit node 10
 if-match ip-prefix ipv6
 apply cost 100
#
route-policy ipv6 permit node 20
#
```

```
<R5>display bgp routing-table
Network                 NextHop        MED        LocPrf    PrefVal  Path/Ogn

*>  10.0.100.0/24       10.0.35.3      0                              100i
 *                      10.0.25.2      100                            100i
```

```
[R5]display bgp ipv6 routing-table
*>  Network  : 2000:0:0:1::1                    PrefixLen : 64
    NextHop  : 2000::3503                       LocPrf    :
    MED      : 0                                PrefVal   : 0
    Label    :
    Path/Ogn : 100  i
```

- 命令含义
- **apply cost [+ | -]**　cost 命 令用来在路由策略中配置改变路由的开销值的动作。

- 参数含义：

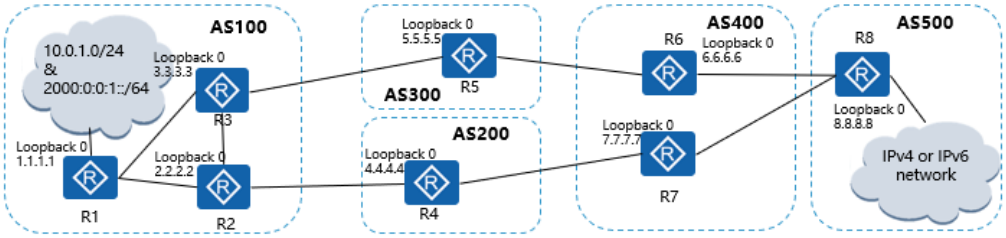
- + ：表示增加开销值。
- - ： 表示减少开销值。
- cost ：指定路由的开销值。对路由的选路进行控制，需要将路由的开销设置为固定值时，可以通过调整开销值避免路由环路的产生。

- 注意事项：
- 在缺省情况下，BGP 只比较来自同一 AS 的路由的 MED 值。这里的 AS 不包括联盟的子 AS。为了使 BGP 在联盟内选择最优路由时能够比较 MED 值，可以配置 bestroute med-confederation 命令。
- 配置 bestroute med-confederation 命令后，只有当 AS_Path 中不包含外部自治系统（不在联盟内的自治系统）号时才比较 MED 值的大小。如果 AS_Path 中包含外部自治系统号，则不进行比较。
- 例如：自治系统 65000、65001、65002 和 65004 属于同一联盟。四条到达同一目的地址的待选路由如下所示：

- path1：AS_Path=65000 65004，med=2
- path2：AS_Path=65001 65004，med=3
- path3：AS_Path=65002 65004，med=4
- path4：AS_Path=65003 65004，med=1
- 在配置 bestroute med-confederation 命令后，因为 path1、path2 和 path3 的 AS_Path 中不包含同一联盟外的自治系统，所以当 BGP 需要通过比较 MED 值来选择路由时，将只比较 path1、path2 和 path3 的 MED 值。而 path4 的 AS_Path 中包含同一联盟外的自治系统，因此不比较 path4 的 MED 值。

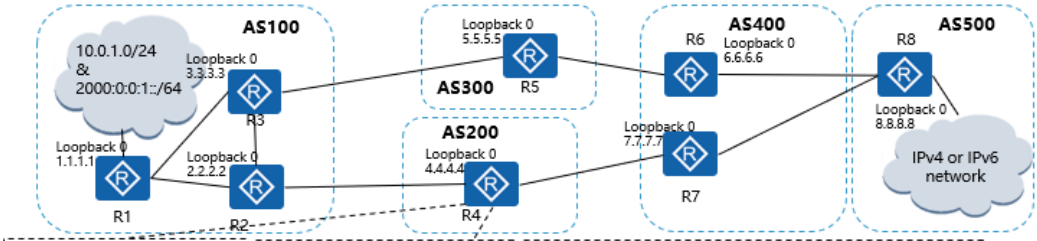
配置BGP– BGP AS_Path属性1

- 配置需求1：①AS500内的用户需要优选经过AS200到达AS100内的两个网段
  ②仅能操作AS200内的设备



- 拓扑与配置采用“基础配置”的内容。仅有基本的 BGP 邻居配置

配置BGP– BGP AS_Path属性1（续）




```
bgp 200
 router-id 4.4.4.4
#
 ipv4-family unicast
 peer 10.0.47.7 route-policy ipv4
export
#
 ipv6-family unicast
 peer  2000::4707 route-policy ipv6
export
#
ip ip-prefix ipv4 index 10 permit
10.0.1.0 24
ip ipv6-prefix ipv6 index 10 permit
2000:0:0:1::64
```

```
route-policy ipv4 permit node 10
 if-match ip-prefix ipv4
 apply as-path none overwrite
#
route-policy ipv4 permit node 20
#
route-policy ipv6 permit node 10
 if-match ip-prefix ipv6
 apply as-path none overwrite
#
route-policy ipv6 permit node 20
```

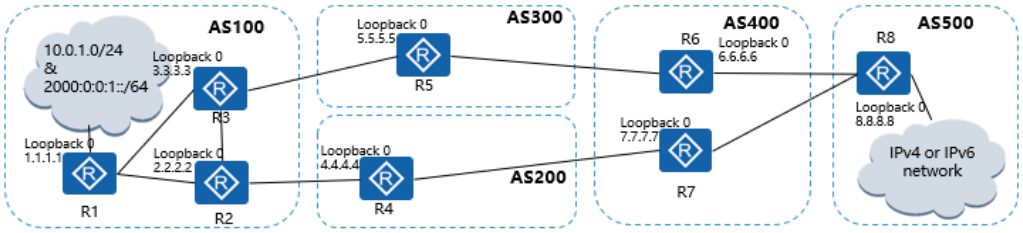
```
[R8]display bgp routing-table
     Network         NextHop        MED        LocPrf    PrefVal Path/Ogn

*>   10.0.1.0/24     10.0.78.7                             400 300 100i
*                    10.0.68.6                             400 i
[R8]display bgp ipv6 routing-table
Total Number of Routes: 3
*>   Network  : 2000:0:0:1::1                PrefixLen : 64
     NextHop  : 2000::7807                   LocPrf    :
     MED      : 0                            PrefVal   : 0
     Label    :
     Path/Ogn : 400 i
```

- 命令含义
- **apply as-path** { { as-number-plain | as-number-dot } &<1-10> { additive } | none overwrite }
- 参数含义
- *as-number-plain*：指定要替换或增加的整数形式的 AS 号。在同一个命令行中最多可以同时指定 10 个 AS 号。

- *as-number-dot ：指*定要替换或增加的点分形式的 AS 号。在同一个命令行中最多可以同时指定 10 个 AS 号。
- **additive** ：在原有的 AS_Path 列表中追加指定的 AS 号。
- **overwrite** ：用指定的 AS 号覆盖原有的 AS_Path 列表。
- **None** ：清空原来的 AS_Path 列表。
- 注意事项：
- 策略生效后，将会影响 BGP 路由选路。
- 配置该命令会直接影响网络流量所经过的途径，另外也可能造成环路和选路错误，**请谨慎使用该命令**。

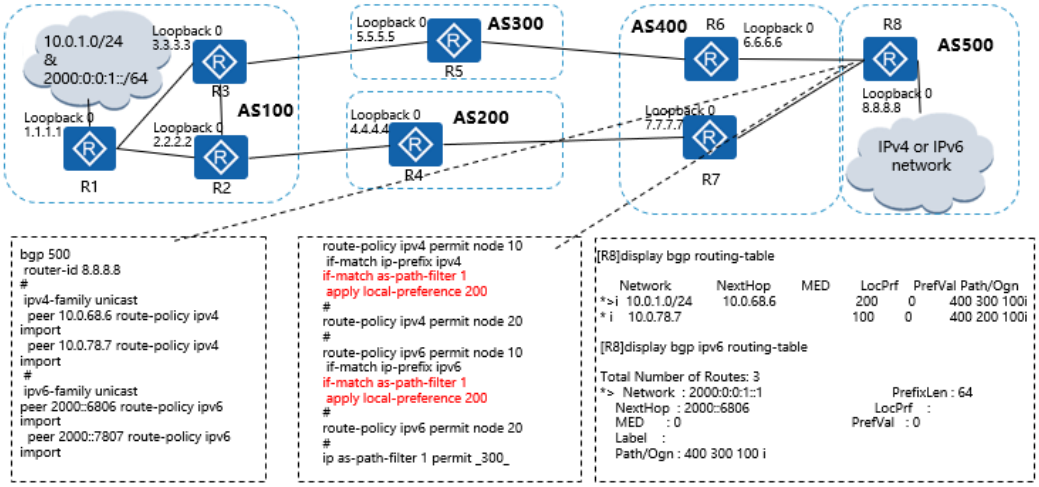
配置BGP– BGP AS_Path属性2

- 配置需求：① AS500内的用户需要优选经过AS300到达AS100的两个网段
  ②仅能在AS500内的设备上进行操作
  ③AS500不知道全网的拓扑信息



- 拓扑与配置采用“基础配置”的内容。仅有基本的 BGP 邻居配置

配置BGP– BGP AS_Path属性2（续）



- 命令含义：
- **if-match as-path-filter** { as-path-filter-number &<1-16> | as-path-filter-name }
- 参数含义：
- as-path-filter-number ：指定 AS 路径过滤器号。在一个命令行中可以配置多个此参数，但最大不能超过 16。 整数形式，取值范围 1～256。
- *as-path-filter-name* 指定 AS 路径过滤器名称。字符串形式，区分大小写，不支持空格，长度范围是 1～51，且不能都是数字。
- 注意事项：
- 在一个命令行中可以配置多个 AS-Path-filter 值，但最多不能超过 16 个。它们之间是“或”的关系，即通过其中某一个 AS 路径过滤器的过滤就可以通过该命令的过滤。

配置BGP– BGP AS_Path属性2之as-path-filter

- AS路径过滤器（AS_Path-Filter）
  - AS路径过滤器是一组针对BGP路由的AS_Path属性进行过滤的规则
  - AS路径过滤器的默认行为是deny，即路由如果没有在某一次过滤中被permit则最终不能通过该过滤器的过滤。
  - AS路径过滤器的匹配条件使用正则表达式指定
    - 举例：① ^10_ 匹配AS_Path属性以AS10开头的路由
      ② _20_ 匹配AS_Path属性中包含AS20的路由
      ③.* 匹配所有AS_Path属性

- **命令含义：**
- **ip as-path-filter** { *as-path-filter-number* | *as-path-filter-name* } { **deny** | **permit** } *regular-expression* 命令用来创建 AS 路径过滤器。
- 参数含义：
- as-path-filter-number 指定的 AS 路径过滤器号。 整数形式，取值范围 1～256。
- as-path-filter-name 指定的 AS 路径过滤器名称。 字符串形式，区分大小写，不支持空格，长度范围是 1～51，且不能都是数字。当输入的字符串两端使用双引号时，可在字符串中输入空格。
- deny 指定 AS 路径过滤器的匹配模式为拒绝。
- permit 指定 AS 路径过滤器的匹配模式为允许。

配置BGP– BGP AS_Path属性2之正则表达式（1/2）

特殊字符	功能	举例
\	转义字符。将下一个字符（特殊字符或者普通字符）标记为普通字符。	*匹配*
^	匹配行首的位置。	^10可以匹配 “10,20,30”，“100,200,300” ……
$	匹配行尾的位置。	1$可以匹配为101， “11” ……
*	匹配前面的子正则表达式零次或多次。	10*可以匹配1、10、100、1000、…… (10)*可以匹配空、10、1010、101010、……
+	匹配前面的子正则表达式一次或多次。	10+可以匹配10、100、1000、…… (10)+可以匹配10、1010、10101、……
?	匹配前面的子正则表达式零次或一次。	10?可以匹配1或者10 (10)?可以匹配空或者10
.	匹配任意单个字符。	.0可以匹配20、30……

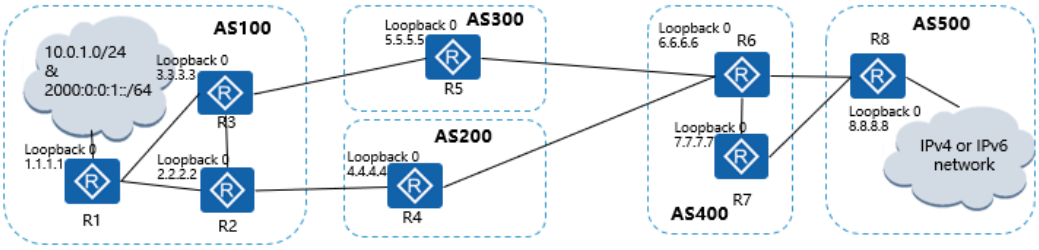
配置BGP– BGP AS_Path属性2之正则表达式（2/2）

特殊字符	功能	举例
()	一对圆括号内的正则表达式作为一个子正则表达式，匹配子表达式并获取这一匹配。圆括号内也可以为空。	100(2)+可以匹配1002、10022
_	匹配一个符号，包括逗号、左大括号、右大括号、左括号、右括号和空格，在表达式的开头或结尾时还可作起始符、结束符（同^，$）。	_65001_可以匹配65001、20 65001、65001 30、……
x|y	匹配x或y。	100|200匹配100或者200； 1(2|3)4匹配124或者134
[xyz]	匹配正则表达式中包含的任意一个字符。	[123]匹配255中的2
[^xyz]	匹配正则表达式中未包含的字符。	[^123]匹配除123之外的任何字符
[a-z]	匹配正则表达式指定范围内的任意字符。	[0-9]匹配0到9之间的所有数字
[^a-z]	匹配正则表达式指定范围外的任意字符。	[^0-9]匹配所有非数字字符

配置BGP–负载分担

- 配置需求：①AS500中的设备充分利用两条链路访问AS100

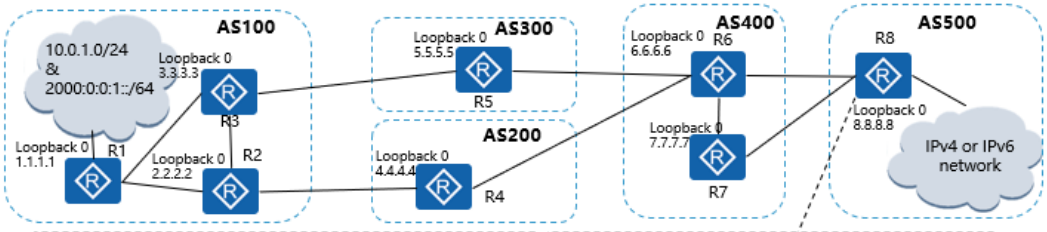
②只能操作AS500中的设备




```
[R8]display ip routing-table protocol bgp
Destination/Mask   Proto   Pre  Cost      Flags NextHop         Interface
      10.0.1.0/24  EBGP    255  0           RD  10.0.68.6       GigabitEthernet0/0/0

[R8]display ipv6 routing-table protocol bgp
 Destination  : 2000:0:0:1::                PrefixLength : 64
 NextHop      : 2000::6806                  Preference   : 255
 Cost         : 0                           Protocol     : EBGP
 RelayNextHop : ::                          TunnelID     : 0x0
 Interface    : GigabitEthernet0/0/0        Flags        : D
```

- 拓扑与配置采用“基础配置”的内容。仅有基本的 BGP 邻居配置

配置BGP–负载分担




```
[R1]display ip routing-table protocol bgp
Destination/Mask   Proto   Pre  Cost      Flags NextHop         Interface
      10.0.1.0/24  EBGP    255  0           RD  10.0.68.6       GigabitEthernet0/0/0
                   EBGP    255  0           RD  10.0.78.7       GigabitEthernet0/0/1
[R8]display ipv6 routing-table
Destination  : 2000:0:0:1::                PrefixLength : 64
NextHop      : 2000::6806                  Preference   : 60
Cost         : 0                           Protocol     : EBGP
RelayNextHop : ::                          TunnelID     : 0x0
Interface    : GigabitEthernet0/0/1        Flags        : RD

Destination  : 2000:0:0:1::                PrefixLength : 64
NextHop      : 2000::7807                  Preference   : 60
Cost         : 0                           Protocol     : EBGP
RelayNextHop : ::                          TunnelID     : 0x0
Interface    : GigabitEthernet0/0/0        Flags        : RD
```

```
bgp 100
 ipv4-family unicast
  maximum load-balancing ebgp 2
 ipv6-family unicast
  maximum load-balancing ebgp 2

[R8]display bgp routing-table
   Network         NextHop        MED        LocPrf    PrefVal Path/Ogn
*> 10.0.1.0/24     10.0.68.6                            400 300 100i
*                  10.0.78.7                            400 300 100i
[R8]display bgp ipv6 routing-table
Total Number of Routes: 3
*> Network  : 2000:0:0:1::1                PrefixLen : 64
   NextHop  : 2000::6806                   LocPrf    :
   MED      : 0                            PrefVal   : 0
   Label    :
   Path/Ogn : 400 300 100 i
```

配置了负载分担之后
依旧会选出一条最优路由

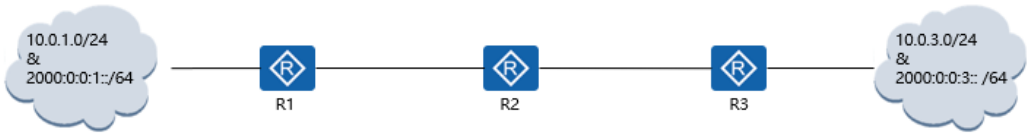
- 命令含义：
- **maximum load-balancing** 命令用来设置等价路由的最大条数。
- 具体用法
- 命令 maximum load-balancing 为 BGP 视图命令。

- 参数意义
- ebgp：仅 EBGP 路由参与负载分担。
- ibgp：仅 IBGP 路由参与负载分担。
- number：BGP 路由表中最大等价路由条数。

- 注意事项
- 如果配置了 maximum load-balancing number 命令，那么再配置 maximum load-balancing ebgp number 或 maximum load-balancing ibgpnumber 命令都不会生效；如果配置了 maximum load-balancing ebgp number 或 maximum load-balancing ibgp number 命令，那么再配置 maximum load-balancing number 命令也不会生效。
- AS_Path 不仅需要长度相等，内容也必须一致才能形成负载分担。可在 BGP 进程视图下使用 load-balancing as-path-ignore 设置路由在形成负载分担时不比较路由的 AS-Path 属性。
- 实验结果
- 使用命令 display ip routing-table protocol bgp 可以查看到通过 BGP 学到的等价路由。

BGP故障诊断

- 三台路由中的直连路由器之间建立BGP邻居关系
- 配置完成之后发现R1和R3所连接的网络无法互访
- 分析并解决此问题



BGP故障排除流程

- 由于本章主讲BGP，假设网络中非BGP部分没有问题
  - BGP邻居状态无法到达Established状态：
    - IGP不通
    - ACL过滤了TCP的179端口
    - 邻居的Router ID冲突
    - 配置的邻居的AS号错误
    - 用Loopback口建立邻居时没有配置peer connect-interface
    - 用建立EBGP邻居时未配置peer ebgp-max-hop
    - peer valid-ttl-hops配置错误。
    - 对端发送的路由数量是否超过peer route-limit命令设定的值。
    - 对端配置了peer ignore
    - 两端的地址族不匹配

BGP故障排除流程（续）

- BGP邻居关系正常的情况下，但是BGP路由表没有该表项
  - 下一跳地址是否可达
  - 入口是否进行了策略限制
  - 接收前缀的条目是否进行了限制
  - 对端出口是否进行了策略限制
  - 该前缀在对端BGP路由表中是否最优
  - 对端是否配置了active-route-advertise

BGP故障排除流程（续）

- BGP邻居关系正常的情况下，BGP路由表存在某些表项不是最优
  - 根据选路原则，某些表项不是最优
  - 某些前缀是否为抑制状态

思考题

1. BGP 在OPEN报文中协商地址族支持情况（ ）
   A. T
   B. F

2. AS_Path长度相同，但是顺序不同，不会影响负载分担的建立（ ）
   A. T
   B. F

- 参考答案：
- T，正确
- F，AS_Path 需要完全一样才能形成负载分担。

本章总结

- BGP 4 & BGP4+原理
  - BGP基本原理
  - BGP属性及其特点
  - BGP 4 + 原理及其新增属性
- BGP 4 &BGP 4 + 配置
  - 华为设备上的配置命令熟悉
  - 不同场景下的配置选择
- BGP 故障排查
  - BGP排查思路