BGP

1. 概述

BGP 为一种路径矢量协议，传递信息为路由条目，其为应用层协议，TCP 179 端口，更新报文均为单播报文。

AS，自制系统，唯一的标记一个园区网，其范围为 0-65535，其中 0-64511 为公有 AS 号，64512-65535 为私有号。

三张表：邻居表、BGP 表、路由表。

管理距离：IBGP 路由 AD 为 200；EBGP 路由 AD 为 20

IGBP 防环机制：收到 IBGP 对等体的路由不会再传给其他 IBGP 对等体；

EBGP 防环机制：EBPG 会通过 AS path 属性，不会将路由传给已包含的 AS 内的路由器

2. 报文类型

1，Open 报文：用于建立邻接关系。交互 BGP 版本、AS 号、Holdtime（默认 180s）和 RID 信息。RID 可以手动配置，也可以自动选举。自动选举的规则为：（1）选择 BGP 路由器中，在线环回口最大的 IP 地址作为 RID；（2）选择物理口最大的 IP 地址作为 RID。

2，Keeplive 报文：用于维护邻接关系，每 60s 发送一次。

3，Update 报文：交互路由掩码信息、路由属性以及撤销的路由。

用于传递更新的路由条目的前缀掩码，下一跳以及 BGP 属性等信息。

4，Notification 报文：当 BGP 发生错误时，会发送该报文。

3. 邻居状态

BGP 有 6 中邻居状态：Idle，Connect，Open Sent，Open Confirm，Active 和 Established

Idle：路由器通过路由表查找邻居的过程；

Connect：路由器找到邻居，并且完成了 TCP 三次握手；

Open Sent：路由器将本地 BGP 进程参数以 Open 报文发送给对端；

Open Confirm：路由器收到了对端的 Open 报文，并且参数正确；

Active：如果路由器没有收到对端发送的 Open 报文，会进入该状态，此时会重新 TCP 三次握手；

Established：邻居建立，开始传递路由

4. 属性

BGP 属性有 4 类：公认强制属性（origin，AS-path，下一跳），公认自选属性（local preference，atomic aggregate），可选传递属性（aggregator，community），可选非传递属性（MED，originator ID，cluster list）以及其他属性。

4.1. Weight

权重属性，思科私有属性，不可传递。

缺省值：若其下一跳为 0.0.0.0，则其缺省值为 32768（包括本地 network 进入的非 IGP 路由以及重分发进入的路由）；若其下一跳不为 0.0.0.0（包括本地 network 进入的 IGP 路由，以及邻居传递来的路由），则缺省值为 0。其取值范围为 0-65535，越大越优。

从该邻居收到的所有路由修改 weight 属性：neighbor 3.3.3.3 weight 1

精确修改 weight route map 调用

```
ip prefix-list wei_plist seq 5 permit 11.11.11.0/24
route-map wei_map permit 10
 match ip address prefix-list wei_plist
 set weight 1
route-map wei_map permit 20
router bgp 234
 nei 3.3.3.3 route-map wei_map in
```

4.2. Local Preference

公认自选属性，传递范围为一个 AS，缺省值为 100，越大越优。

AS 内对于同一条路由，通过该属性区分那条路由最优，用于通告给 IBGP 邻居，该路由是如何离开 AS 的。

修改命令：bgp default local-preference 101

或者使用 route-map：

```
ip prefix-list 10 seq 10 permit 111.111.111.0/24
route-map local permit 10
 match ip address prefix-list 10
 set local-preference 101
route-map local permit 20
router bgp 234
 neighbor 12.1.1.1 route-map local in
```

4.3. AS-Path

公认强制属性，传递范围是整个 Internet，越短越优。

用一串 AS 号描述目标路由经过哪些 AS。

```
access-list 10 permit 11.11.11.0
route-map ap1 permit 10
 match ip address 10
 set as-path prepend 5 6 7 8 #可以添加相同的 as
route-map ap1 permit 20
router bgp 234
 neighbor 12.1.1.1 route-map ap1 in
```

Sh ip bgp 中，AS-Path 显式的为数据层面的，分析控制层面的 as path 和数据层面相反。

```
neighbor 4.4.4.4 allowas-in #允许向已有的 AS-Path 传递路由
bgp maxas-limit 10 #允许最大传输的 AS-Path 数为 10
bgp bestpath as-path ignore #忽略 AS-path 属性
```

4.4. Origin

起源属性，公认强制属性，传递范围是整个 Internet。

描述路由以何种方式进入 BGP 中的，i 为 IGP 宣告进入 BGP 的，？为重分发进入 BGP 的，e 为通过 EGP 进入 BGP 的，可以通过 route-map 进行修改。i 优于 e 优于？。

配置举例：

```
ip prefix-list 10 per 11.11.11.0/24
```

route-map o per 10
 match ip add prefix-list 10
 set origin incomplete
route-map o per 20
router bgp 234
 neighbor 4.4.4.4 route-map o out

4.5. MED

Multi-Exit Discriminators，多出口鉴别器，在邻居的一跳 AS 传递，缺省值：IETF 最大值，Cisco 定义为 0，越小越优。

MED 会影响入站流量，用于同一路由器告诉邻居 AS，如何从邻居 AS 到达本地 AS 的路由最近。

举例：

ip prefix-list 10 seq 5 permit 11.11.11.0/24 #将本地路由通告给邻居 AS
route-map m permit 10
 match ip address prefix-list 10
 set metric 100
route-map m permit 20
router bgp 1
 neighbor 13.1.1.3 route-map m out

将思科路由器缺省值改为最大值：bgp bestpath med missing-as-worst

允许不同路由器发送来的同一条路由条目来比较其 MED 值：

bgp always-compare-med

4.6. 下一跳

1， 若将本地路由（直连路由和静态路由）通告进 BGP 进程，该路由器的本地 BGP 表关于它们的下一跳为 0.0.0.0；
2， 若将 IGP 获悉的路由通告进 BGP 进程，该路由器本地 BGP 表关于它们的下一跳为 IGP 路由的下一跳地址；
3， 若路由器通过 BGP 对等体收到一条路由，则该路由的下一跳为邻居的更新源地址；
4， 若路由器通过 EBGP 对等体学到一条路由，该路由器在传给其 IBGP 对等体时，默认情况下一跳不会改变（除非做 next-hop-self，或者其 IBGP 对等体有本地关于 EBGP 的更新源地址路由）；
5， 若路由器通过 BGP 对等体学到一条路由，该路由器在传递给 EBGP 对等体时，下一跳会改变为本地对于 EBGP 对等体的更新源地址。

4.7. Atomic aggregate 和 aggregator

4.7.1. 路由聚合

两种方式聚合路由：

1，可以手动写一条精确聚合路由，指向 null0，然后将其宣告进入 BGP；

2，使用 network 命令先宣告一条精确路由，然后使用 aggregate-address 192.168.4.0 255.255.252.0 聚合路由。此时会将聚合路由和明细路由同时传递，加 summary-only 可以抑制明细路由，也可以使用 suppress-map 来精确抑制。

4.7.2. Atomic aggregate

在传递聚合路由时，使用 summary-only 参数，会导致 AS-PATH 属性丢失的情况，所以在传递聚合路由时，可以加入 atomic aggregate 属性来标识该路由为聚合路由。传递范围是整个 Internet。

可以添加 as-set 参数来显示原来所在的 as-path。

4.7.3. aggregator

聚合路由会将 aggregator 一并传递给邻居，标识路由被聚合的路由器 ID。

4.8. Community

团体属性为公认自由属性，不可传递属性，只能传递一跳。

4.8.1. 标准团体属性

标准团体属性有三个值 No-advertise，No-export，Local-AS。

需要在传递路径上都配置团体属性，该属性才可以传递下去。

1， No-Advertise：收到携带该属性的 BGP 路由时，路由无法传递给其他 BGP 对等体；
2， No-Export：收到携带该属性的 BGP 路由时，路由无法传递给其他 EBGP 对等体。但若在联邦中，该属性可以在子 AS 之间进行传递；
3， Local-AS：收到携带该属性的 BGP 路由时，路由只能在本地 AS 内传递（包括联邦的子 AS 内传递）。

4.8.2. 扩展团体属性

XX：YY tag，可以使用该 tag 来过滤路由。

配置命令：

```
ip community-list standard DENY permit 50:50
route-map COM deny 10
 match community DENY
route-map COM per 20
router bgp 65001
 nei 1.1.1.1 route-map COM in
```

4.9. Originator ID 和 cluster list

在 RR 传递 RRC 的路由给其他 RRC 时，会带有这两种属性。Originator ID 表示通告者 RRC，cluster list 表示 RR。

可选属性，传递范围是一个 RR 域。

5. 路由选路原则

1，较高的权重；
2，较高的本地优先级；
3，本地通告的路由优于邻居传递来的路由（可能产生路由环路）；
4，最短的 AS-Path
5，起源属性：i>e>?
6，较小的 MED 值
7，EBGP 路由优于联邦 EBGP 路由，优于 IBGP 路由

8，如果为内部路由，选择到下一跳最近的路由，也就是 IGP 度量值最小的路由；

9，如果外部路由，选择 multipath

10，较老的 EBGP 路由（一般不作为参考对象）

11，如果均来自一个 AS 的路由，并且启用了 BGP 多路功能（命令为 maximum-path），在路由表中安装等价路由；

12，如果没有 BGP 多路功能，选择 RID 最小的路由，

13，最小的 Cluster List 长度

14，较低的邻居 IP 地址的路由

6. Route Reflector

6.1. 定义

路由反射器，简称 RR；

Cluster，在同一个 AS 之内，RR 所能涉及到的范围；

RRC，路由反射器客户端。

RR 和 RRC 之间有 IBGP 邻接关系，而 RRC 之间没有邻接关系。

6.2. 工作机制

RR 收到一条 EBGP 路由，会将其传递给其它 EBGP 对等体、IGBP 对等体（包括 RRC 和 non-RRC）；

RR 收到一条 RRC 传递的 IBGP 路由，会将其发送给其他 EBGP 对等体、IGBP 对等体（包括 RRC 和 non-RRC）；

RR 收到一条 non-RRC 传递的 IBGP 路由，会将其传递给其他 EBGP 对等体和 RRC，不会传递给 non-RRC。

被 RR 反射的路由，不会修改任何 BGP 属性。

6.3. 配置

在 RR 上 BGP 进程中配置：neighbor 23.1.1.3 route-reflector-client，宣告 23.1.1.3 为本地的 RRC

7. Confederation

考虑到在 AS 内部没有防环机制，iBGP 之间传递路由只能有一跳。

联邦，在一个 AS 之内，划分出多个子 AS 域，建立 EBGP 邻接关系，可以将路由母 AS 之内进行多跳的传递。

举个例子：

R1-R2-R3-R4

R1 在 AS1，R2、R3、R4 在 AS2，R2,、R3 在 65002 子 AS，R4 在 65004 子 AS。

R1:

```
router bgp 1
 bgp log-neighbor-changes
 network 1.1.1.1 mask 255.255.255.255
 neighbor 12.1.1.2 remote-as 2
```

R2:

```
router bgp 65002
 bgp router-id 2.2.2.2
```

bgp log-neighbor-changes
bgp confederation identifier 2
neighbor 12.1.1.1 remote-as 1
neighbor 23.1.1.3 remote-as 65002

R3:

router bgp 65002
bgp router-id 3.3.3.3
bgp log-neighbor-changes
bgp confederation identifier 2
bgp confederation peers 65004
neighbor 23.1.1.2 remote-as 65002
neighbor 34.1.1.4 remote-as 65004

R4:

router bgp 65004
bgp router-id 4.4.4.4
bgp log-neighbor-changes
bgp confederation identifier 2
bgp confederation peers 65002
neighbor 34.1.1.3 remote-as 65002

可以将路由反射器和联邦联合使用，解决复杂问题。

8. show ip bgp 命令

```
RouterA# show ip bgp
BGP table version is 14, local router ID is 172.31.11.1
Status codes: s suppressed, d damped, h history, * valid, > best, i -
internal, r RIB-failure, S Stale
Origin codes: i - IGP, e - EGP, ? - incomplete
   Network          Next Hop            Metric LocPrf Weight Path
*> 10.1.0.0/24      0.0.0.0                  0         32768 i
* i                 10.1.0.2                 0    100      0 i
*> 10.1.1.0/24      0.0.0.0                  0         32768 i
*>i10.1.2.0/24      10.1.0.2                 0    100      0 i
*> 10.97.97.0/24    172.31.1.3                             0 64998 64997 i
*                   172.31.11.4                            0 64999 64997 i
* i                 172.31.11.4              0    100      0 64999 64997 i
*> 10.254.0.0/24    172.31.1.3               0             0 64998 i
*                   172.31.11.4                            0 64999 64998 i
* i                 172.31.1.3               0    100      0 64998 i
r> 172.31.1.0/24    172.31.1.3               0             0 64998 i
r                   172.31.11.4                            0 64999 64998 i
r i                 172.31.1.3               0    100      0 64998 i
*> 172.31.2.0/24    172.31.1.3               0             0 64998 i
<output omitted>
```

Displays networks from lowest to highest

BGP 表中，从左到右，*为合法路由，有资格加入路由表；r 为 RIB-failure 路由，也有资格加表，但由于管理距离，无法加表；s 为抑制路由；>为最优路由，实际加入路由表中的路由；i 为路由通过 ibgp 学到的；后面的 i 标识起源属性，意为通过 igp 进入 BGP 的。

同步概念：

如果路由器通过 IBGP 学到一条路由，该路由器必须再通过 IGP 学到该路由才可以加表。

9. 一些命令

BGP 进程下：

neighbor IP-ADD shutdown，用来将 BGP 邻居 down

neighbor IP-ADD update-source INTERFACE，修改更新源地址

neighbor IP-ADD ebgp-multihop TTL-VALUE，修改 EBGP 建立邻居的 TTL 值，默认为 1.

neighbor IP-ADD next-hop-self，BGP 对于 IBGP 邻居传递路由时，其下一条地址不变，配置该命令会将下一条指向自己。

neighbor IP-ADD password PASSWORD

neighbor IP-ADD soft-reconfiguration inbound 允许 sh ip bgp neighbors IP-ADD received-routes

clear ip bgp * soft in/out 软清除 BGP 邻接关系，重新发一次路由更新

clear ip bgp *硬重置 BGP 邻接关系，使 BGP 重新进行三次握手

show ip bgp summary 查看邻居状态等信息