**BGP**

# 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

# 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。

# 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

# 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。

## 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

## 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

## 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

## 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

## 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

## 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。

## Atomic aggregate和aggregator

### 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。

### 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。

### 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。

## Community

团体属性为公认自由属性，不可传递属性，只能传递一跳。

### 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）。

### 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

## 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。

# 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

# Route Reflector

## 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。

## 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。

## 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

# 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

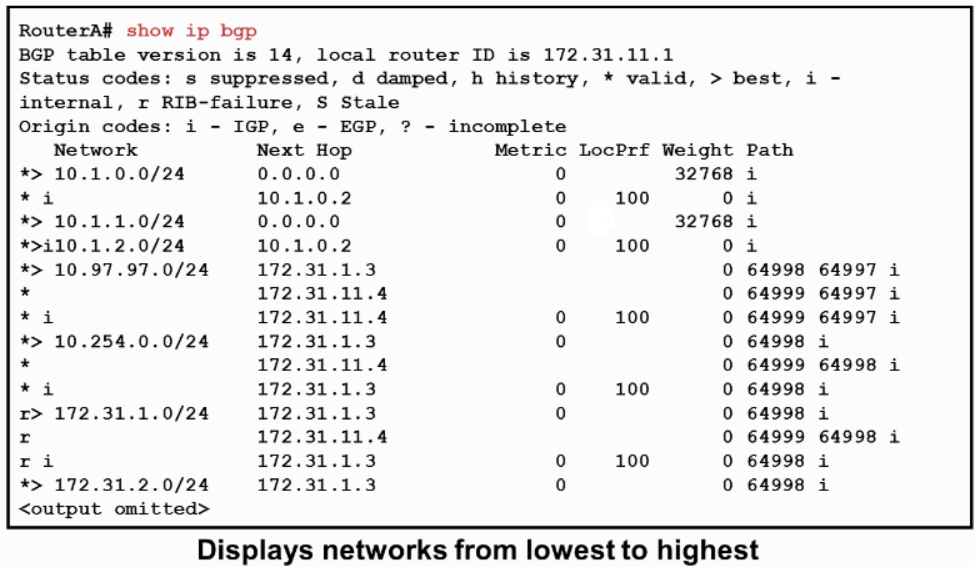
bgp confederation identifier 2

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。

# show ip bgp命令



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。

# 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