Working Principle of BGP

www.huawei.com

Foreword

BGP is a kind of dynamic routing protocol mainly interconnecting AS`, to provide loop-free exchange of routing information. This section will introduce how BGP exchanges loop-free routing information.

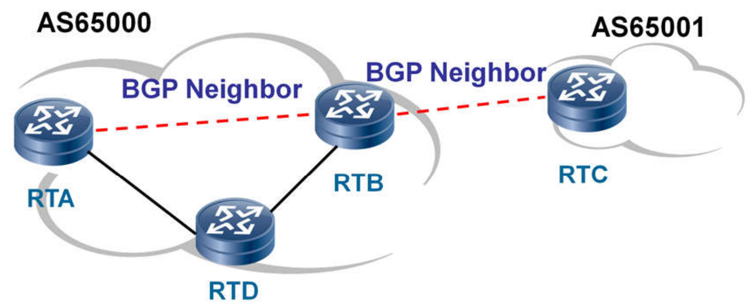
Objectives

Upon completion of this section, you will be able to :

- ⇨ Identify the two kinds of neighbor relationship in BGP protocol
- ⇨ Understand the BGP advertisement principle.
- ⇨ Understand how BGP advertises routing information.

Contents

Neighbor Relationship Based on the TCP Connection



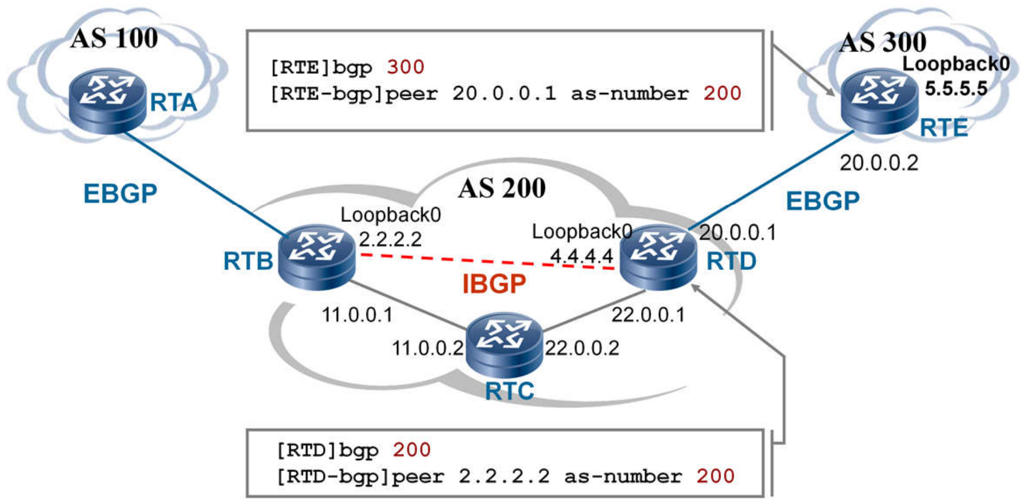
- BGP uses a TCP connection to build the neighbor relationship
- Either IGP or static route can be used to ensure IP reachability between the two routers that form the TCP connection

Similar to OSPF and IS-IS, BGP learns the route through neighbor. The route learning process is slightly different between them. In OSPF and IS-IS, the neighbor relationship is established automatically. However, in BGP, we have to manually specify the IP address of the neighbor. Therefore, in BGP, 2 routers that are connected directly via physical link might not necessarily form the neighbor relationship. In contrast, 2 routers that are not connected directly via physical link might form the neighbor relationship.

To establish the neighbor relationship, BGP rely on the TCP connection. The IP connectivity between 2 BGP routers that form the TCP connection must be assured. We can use the protocol other than BGP to realize the IP connectivity. In other words, either IGP or static route can be used for this purpose. For

convenience, we call the use of either IGP or static route to implement the IP connectivity as IGP connectivity or IGP reachability.

2 Types of BGP Neighbors—IBGP and EBGP



BGP runs in the following two modes: IBGP (Internal BGP), EBGP (External BGP)

•If two peers that exchange BGP messages belong to the same AS, they are Internal BGP (IBGP), such as RTB and RTD.

•If two peers that exchange BGP messages belong to different AS, they are External BGP (EBGP), such as RTD and RTE.

Although BGP runs between ASs, it is necessary to establish BGP connection between different border routers of an AS. Only in this way, the routing information can be transmitted to the whole network. To establish the communication between AS100 and AS300, we need to establish the IBGP connection between RTB and RTD.

The basic configuration of BGP is as follow:

Enable BGP (specify the local AS number), enter BGP view

[Router A] bgp as-number

Using the bgp command, you can enable BGP and enter BGP view.

By default, the BGP is disabled.

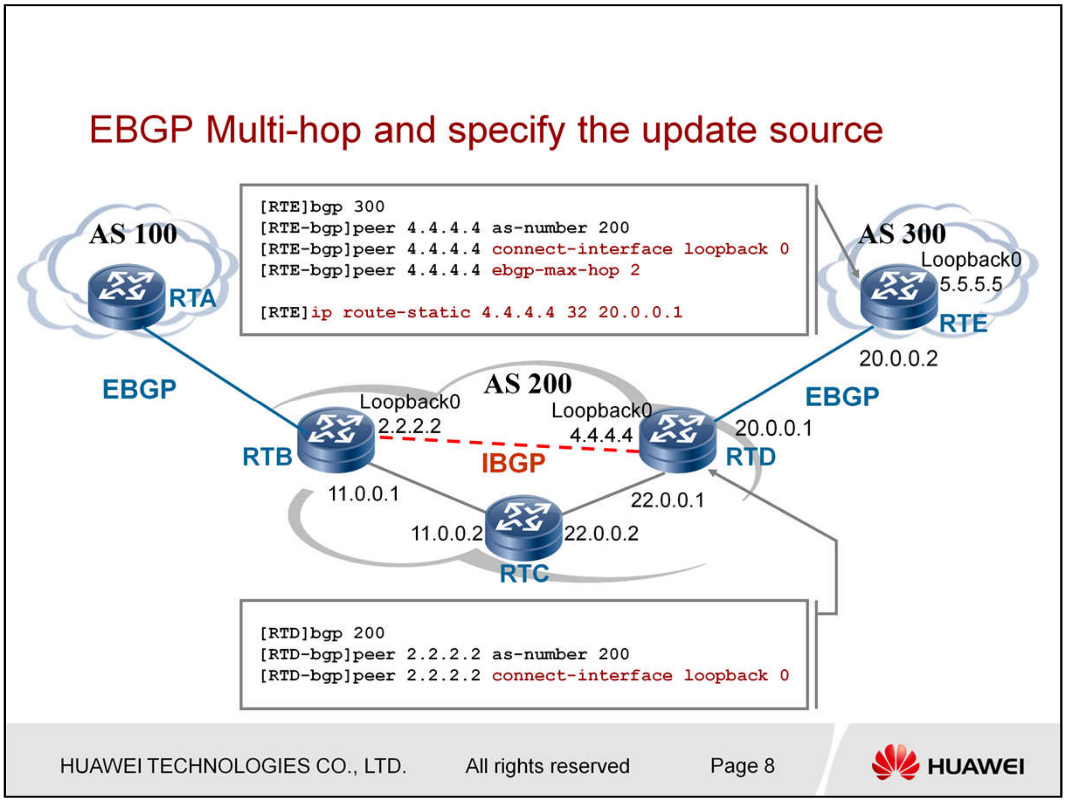
One router can run in only one AS. This means that one router can be specified with only one AS-number.

Specified the peer IP address and its AS number

[Router A-bgp] peer { group-name | ipv4-address | ipv6-address} asnumber as-number

By using peer as-number command, you can configure the AS number of a peer or peer group. Using the undo peer as-number command, you can delete the AS number of a peer or peer group

By default, the opposite end of the peer group has no AS number.



The direct physical connection is not necessarily exists between the IBGP peers.

However, the TCP connection must be established between them. We usually specify the IBGP neighbor by using the loopback interface to ensure the reliability of the route advertisement of IBGP peers. In the case where loopback interface is used, we must specify the source interface of route update packet.

peer { group-name | peer-address } connect-interface interface-name

In most of the cases, there is a direct physical link between the EBGP peers.

However, it is hard to realize this in some cases. To establish the EBGP connection with the peer on the indirectly connected network, we need to modify the maximum hop of EBGP packet by using the following command:

peer { group-name | peer-address } ebgp-max-hop [ttl]

“ttl” is the maximum hop value. It is in the range of 1 to 255. If you specify the maximum hop as 1, you can not establish the EBGP connection with the peer on the indirectly connected network.

Besides the “ebgp-max-hop” command, we need to ensure that the loopback interface of the BGP peer is reachable.

Contents

BGP Route Advertisement Principles (1)

- After the BGP connection is established, information from the BGP tables is exchanged between the BGP speakers..
- When multiple paths exist, a BGP speaker only selects the best route from the BGP for its own use.

```
[RTA]display bgp routing-table

 Total Number of Routes: 2

 BGP Local router ID is 1.1.1.1
 Status codes: * - valid, > - best, d - damped,
               h - history,  i - internal, s - suppressed, S - Stale
               Origin : i - IGP, e - EGP, ? - incomplete
      Network            NextHop        MED        LocPrf     PrefVal Path/Ogn

 *>i  192.168.3.0        10.1.1.2                              0       200i
 * i                     10.2.2.2                              0       200i
```


Under the normal circumstance, when there is more than one alternatives route to the same IP subnet, the BGP speaker will select the best route for its own use.

The best route is the candidate for installation in the IP routing table. However, before a route can be installed, the router will check if there is there is any other routing protocol that has information about the same subnet. If the subnet is known via different sources, the router uses the route preference to determine which source is more trustworthy. The router will install the route with smaller route preference value. That is to say the router will select also the best route for its own use and the best route of BGP speaker might not the best route for the

router. For example, there are 2 routes to the same IP subnet. One is the best route selected by the BGP speaker and another is the static route. In this case, router will install the static route into the IP routing table because it has lower route preference value compare to BGP. Therefore, the best route of BGP will not

be installed into the IP routing table.

As shown in the slide above, there are 2 routes towards network 192.168.3.0 on RTA. The next hops are 10.1.1.2 and 10.2.2.2 respectively.

Base on the route selection criteria (will learn later), BGP will select a best route which is indicated with “>” sign. After that, the router will check if any better routes of other routing protocols exist. If exists, the better route will be installed in the IP routing table. Else, the best route of BGP will be installed into the IP routing table.

For example, the static route to 192.168.3.0 is better compare to BGP route to the same IP subnet (the route preference of static route is 60 while the route preference for BGP is 255, lower route preference value is more trustworthy).

BGP Route Advertisement Principles (2)

- BGP speaker advertises only the routes used by itself to its peers.

```
[RTA]display ip routing-table
Routing Tables: Public
         Destinations : 12       Routes : 12

Destination/Mask    Proto  Pre  Cost        NextHop         Interface

        1.1.1.1/32  Direct 0    0           127.0.0.1       InLoopBack0
       10.1.1.0/30  Direct 0    0           10.1.1.1        Serial0
       10.2.2.0/30  Direct 0    0           10.2.2.1        Serial2
      127.0.0.0/8   Direct 0    0           127.0.0.1       InLoopBack0
    192.168.3.0/24  BGP    255  0           10.1.1.2        Serial0
```


BGP speaker advertises only the best routes used by itself to its peers. This means that it only advertises the BGP routes which are installed in the IP routing table to its peers.

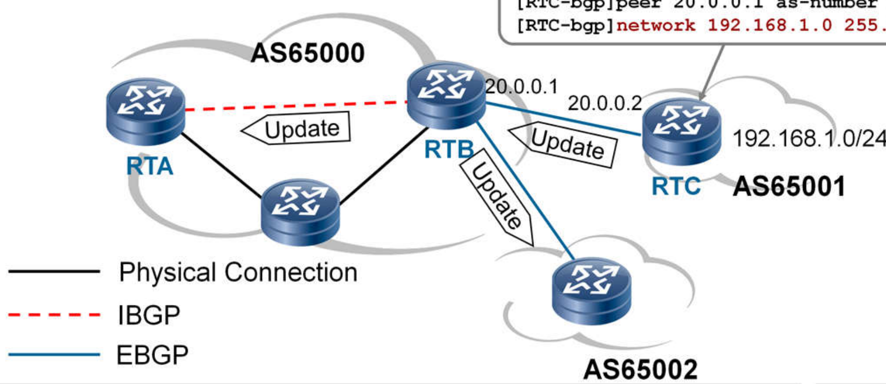
Once the best route (“>”) that has been selected by BGP is installed in the IP routing table, BGP will send Update message which consists of that best route entry to other BGP peer. However, BGP will never sends the updates back on the

same BGP session upon which it was received)

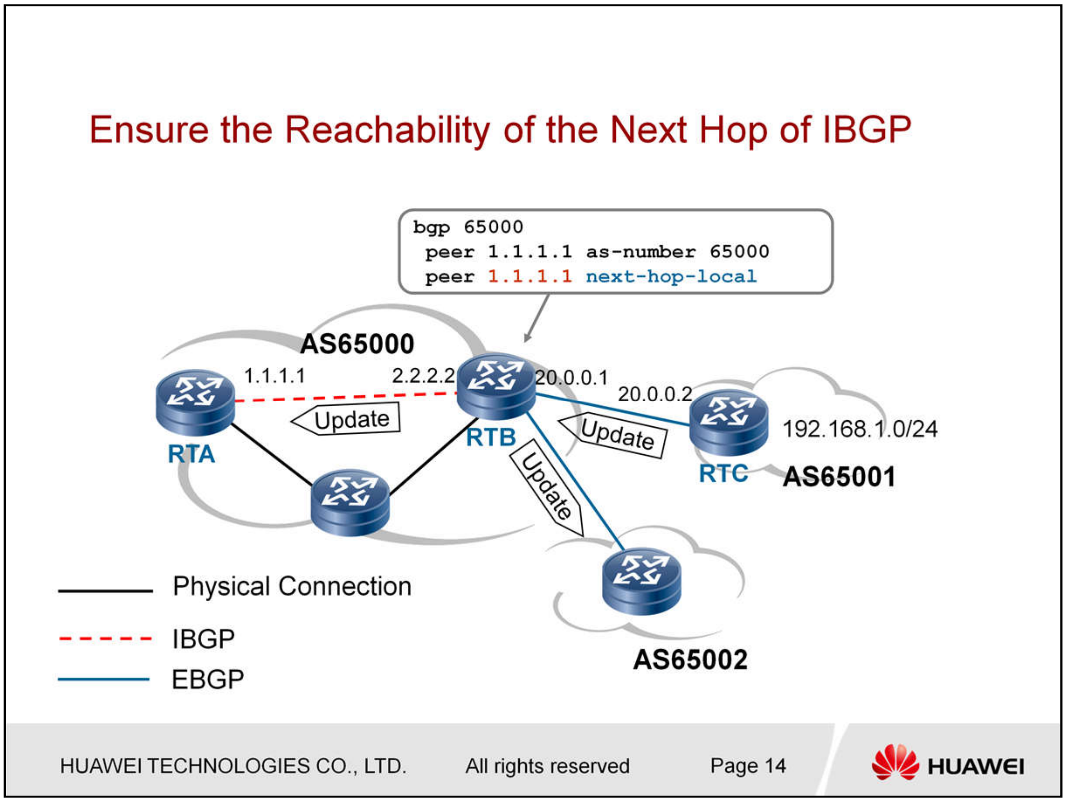
BGP Route Advertisement Principles (3)

- For the routes obtained from EBGP, BGP speaker will advertise them to all its neighbors (including EBGP and IBGP)

```
[RTC]bgp 65001
[RTC-bgp]peer 20.0.0.1 as-number 65000
[RTC-bgp]network 192.168.1.0 255.255.255.0
```



For the routes obtained from EBGP, BGP speaker will advertise them to all its neighbors (including EBGP and IBGP).



IGPs are used to exchange the routing information between the routers.

Therefore, the next hop of any router is the interface IP address of the router that announces the route. This can be easily understood. However, the BGP is used to exchange the loop-free routing information between the autonomous systems.

The autonomous system can be regarded as an “abstract router” in BGP. So, RTB will directly send the update (which has been received from RTC) to RTA without making any changes to the information in the route update packet. As a result, the next hop for RTA to reach the network 192.168.1.0/24 is therefore

20.0.0.2. It is possible that RTA does not know the route 20.0.0.2. This causes unreachability of RTA to route 192.168.1.0/24.

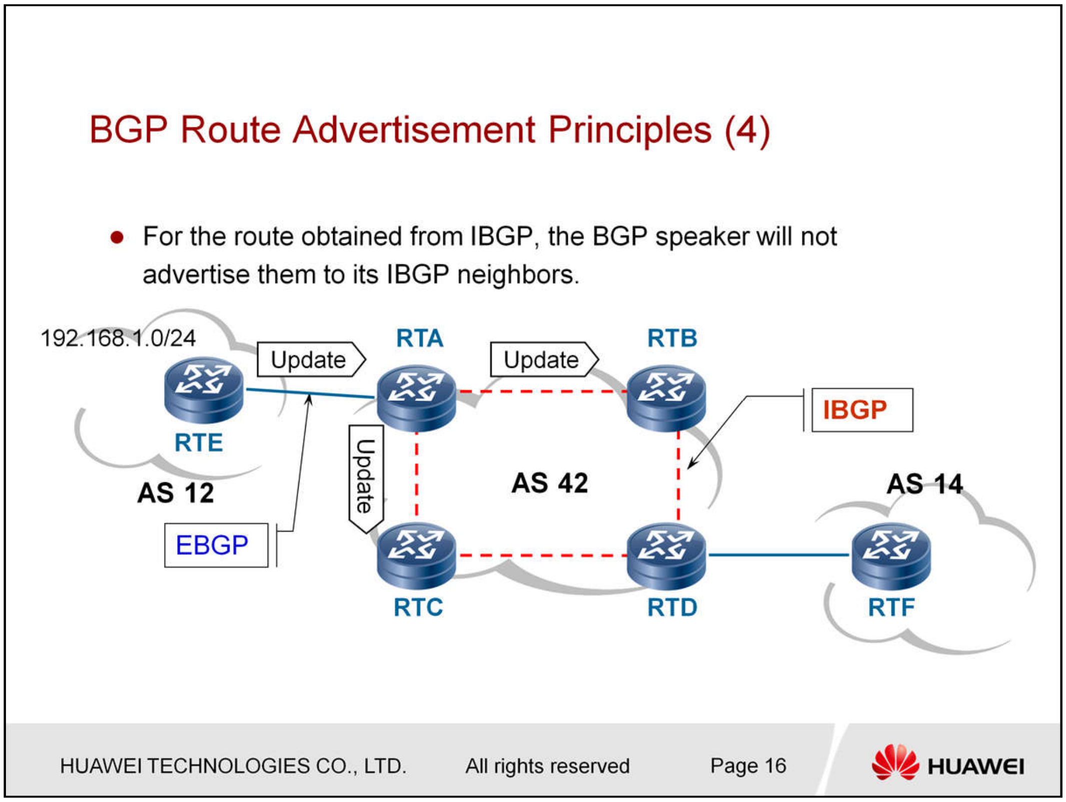
BGP provides a command that can be used to force the next hop to be the IP address of the border IBGP neighbor. This is to ensure that the IBGP neighbor can find the correct next hop information.

Configuration Command:

peer { group-name | ipv4-address } next-hop-local

By using the undo peer next-hop-local, you can restore to the default configuration.

By default, when the BGP notifies the EBGP peer of the route, it change the nexthop attribute to its local IP address. When the BGP notify the IBGP peers of the routes obtained from EBGP, it does not change the next-hop attribute of the route.



RTC learns the route update from its IBGP RTA. Assume that this rule doesn't enforced. RTC will send the route update learnt from IBGP RTA to its IBGP RTD.

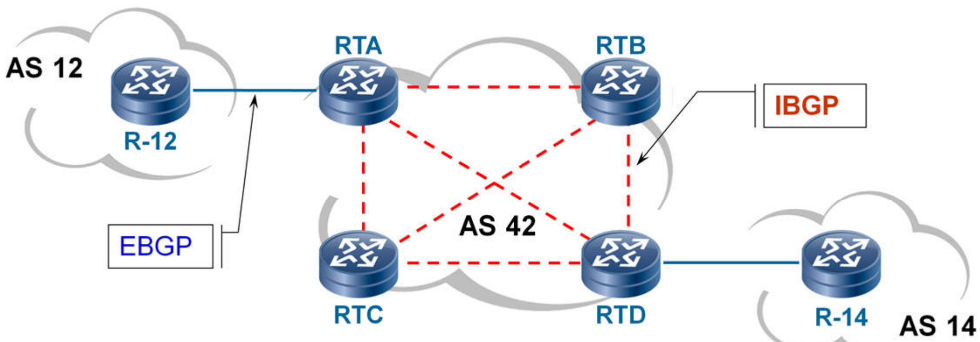
RTD will send the route update learnt from IBGP RTC to its IBGP RTB. Similarly, RTB will send the route update learnt from IBGP RTD to its IBGP RTA. As a result, routing loop is generated.

Therefore, this rule is used to prevent routing loop inside an autonomous system.

However, the enforcement of this rule introduces a new problem to the network:

RTD will not able to receive the BGP route from AS12. To resolve this problem, we often establish the IBP logical full mesh connection. This means that we need to create 2 more IBGP connections for the network above. One of the IBGP connection is created between RTA-RTD and another one is created between RTB-RTC.

IBGP Full Mesh



- IBGP logical full mesh, the routers inside the AS need to maintain many IBGP sessions.
  - ⇨Route Reflector
  - ⇨Confederation

IBG Full Mesh is one of the solution for route advertisement problem caused by the IBGP split horizon rule as described in the previous slide. The disadvantage of this method is that the router need to maintain a large number of IBGP sessions and this consume a lot of resources.

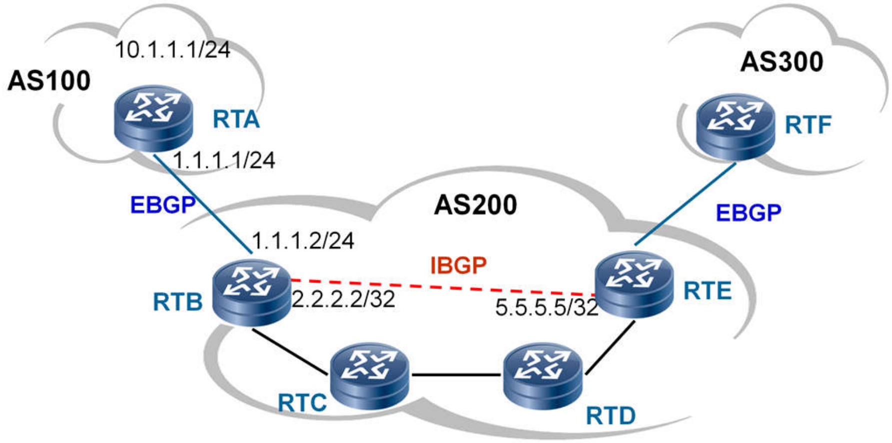
Besides that, BGP provides 2 other methods for the problem caused by the IBGP split horizon rule.

Route-Reflector -- RFC 2796

Confederation -- RFC 3065

BGP Route Advertisement Principles (5)

- For the routes obtained from IBGP, whether the BGP speaker will advertise them to its EBGP neighbors depends on the synchronization state of IGP and BGP.



HUAWEI TECHNOLOGIES CO., LTD.   HUAWEI

The concept of synchronization between BGP and IGP: BGP speaker will not advertise the routing information learnt from the IBGP peer to its EBGP peer unless all routers within the AS had learned about that route through the IGP.

This is known as synchronization. If a router knows about these destinations via an IGP, it assumes that the route has already been propagated inside the AS, and internal reachability is assured.

One of the main responsibilities of BGP is to transmit the network reachability information of its autonomous system to other autonomous system. AS shown in the diagram above, RTB will encapsulate the routing information towards network 10.1.1.0/24 into the BGP update message and advertise it to RTE via the TCP connection established by RTB and RTE. If RTE does not take the synchronization into account, it will directly accept this routing information and send it to RTF. If RTE or RTF has data packet to be sent to network 10.1.1.0/24, this data packet must pass RTD and RTC to reach the destination 10.1.1.0/24.

Since synchronization of IGP and BGP was not taken into account earlier, the routing tables of RTD and RTC do not contain the routing information to network 10.1.1.0/24 and result in the data packet to be discarded when it reaches RTD.

Therefore, BGP must synchronize with IGP (RIP, OSPF and so on). In other words, a router when receives route update information from its IBGP peer will attempt to verify the internal reachability for that route before advertising it to other EBGP peers (check whether that route contains in the routing table of IGP, whether non-BGP routers can deliver traffic to that route). If that route is known via an IGP, the router announces it to other EBGP peers.Otherwise, the router treats the route as not being synchronized with the IGP and does not advertise it.

There are many solutions for the synchronization problem. The simplest one is for RTB to redistribute the BGP routing information to the IGP routing table and then the IGP routing protocol will advertise it to RTE. As a result, the synchronization is achieved. However, this is not recommended because the BGP routing table is very huge. This results in the high consumption of memory and CPU on the IGP routers. The low and middle end equipments that cannot handle this burden will fail.

Another solution is to configure a static route to 10.1.1.0/24 on RTB, and then redistribute it into IGP routing table. The synchronization is therefore achieved.

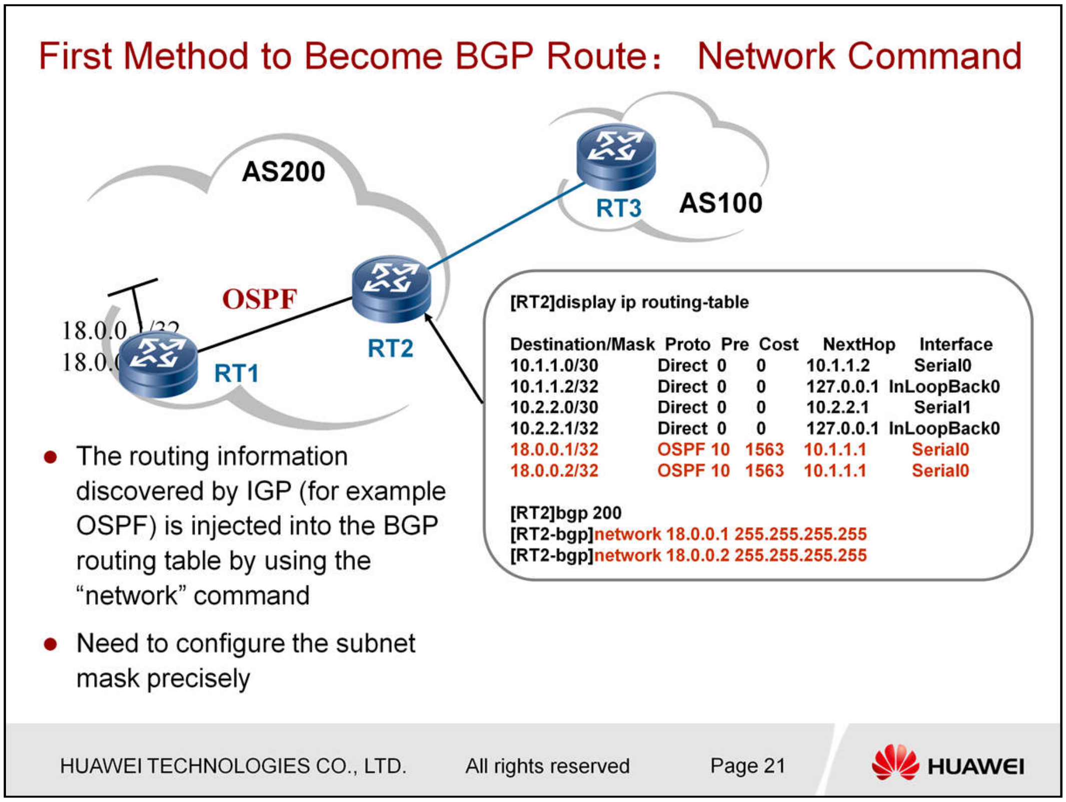
No matter what kind of method you are using, it is not suitable to be used in the large network. .By default, the VRP platform adopts "Undo synchronization" and it cannot be changed.

Certain conditions must be met in order to cancel the synchronization. We can cancel the synchronization when all the routers in the ASs establish an IBGP full connection. This means that the routers between RTB-RTC, RTB-RTD, RTB-RTE, RTC-RTD, RTC-RTE, and RTD-RTE has established the IBGP neighbor relationship via TCP connection. Since the IBGP neigbor relationship has been formed between RTB-RTD, RTD has learnt the BGP route towards 10.1.1.0/24 from RTB. when the data arrive at RTD, RTD will in turn transmit the data to RTC.

Similarly, RTC has learnt the BGP route towards network 10.1.1.0/24 from RTB because the IBGP neighbor relationship has been formed between RTB-RTC.

Therefore, RTC will in turn send the data to RTB. As a result, the data will not lost during the data transmission.

Contents



The main function of BGP is to transmit the loop-free routing information between the autonomous systems instead of discovering and calculating routing information. Therefore, the routing information of BGP needs to be redistributed into BGP using command.

To become BGP route, there are 2 methods : using network command or import command

In fact, we can use the aggregate command to redistribute the IGP aggregated route into the BGP routing table. The recondition to use this method is the BGP routing table must contain those detailed route first. Therefore, we do not

consider this method as one of the solutions.

By using the Network command: the router will inject the routing information of IP routing table into the BGThen, the injected routing information will be sent to other peer via BGP. The

precondition is that the routing information to be injected into the BGP routing table via network command must be existed in the IP routing table.

The related command:

network ipv4-address [mask | mask-length] [route-policy route-policy-name]

ipv4-address: Specified the IPv4 network address advertised by BGP in dotted decimal format.

mask/mask-length: Specifies the network mask or the length of the network mask.

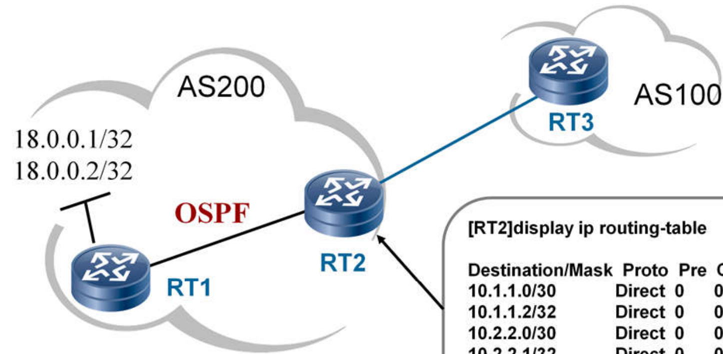
Default subnet mask will be used if it is not specified.

route-policy-name: Specifies the routing policy applied to the advertised routes.

By default, BGP do not advertise any of the local routes.

P routing table according to the Network specified.

Second Method to Become BGP Route： Import Command




```
[RT2]display ip routing-table

Destination/Mask  Proto  Pre  Cost   NextHop    Interface
10.1.1.0/30       Direct 0    0      10.1.1.2   Serial0
10.1.1.2/32       Direct 0    0      127.0.0.1  InLoopBack0
10.2.2.0/30       Direct 0    0      10.2.2.1   Serial1
10.2.2.1/32       Direct 0    0      127.0.0.1  InLoopBack0
18.0.0.1/32       OSPF   10   1563   10.1.1.1   Serial0
18.0.0.2/32       OSPF   10   1563   10.1.1.1   Serial0

[RT2]bgp 200
[RT2-bgp]import-route ospf
```

- The IGP route or static route is injected into the BGP routing table of RT2 by using the import-route command

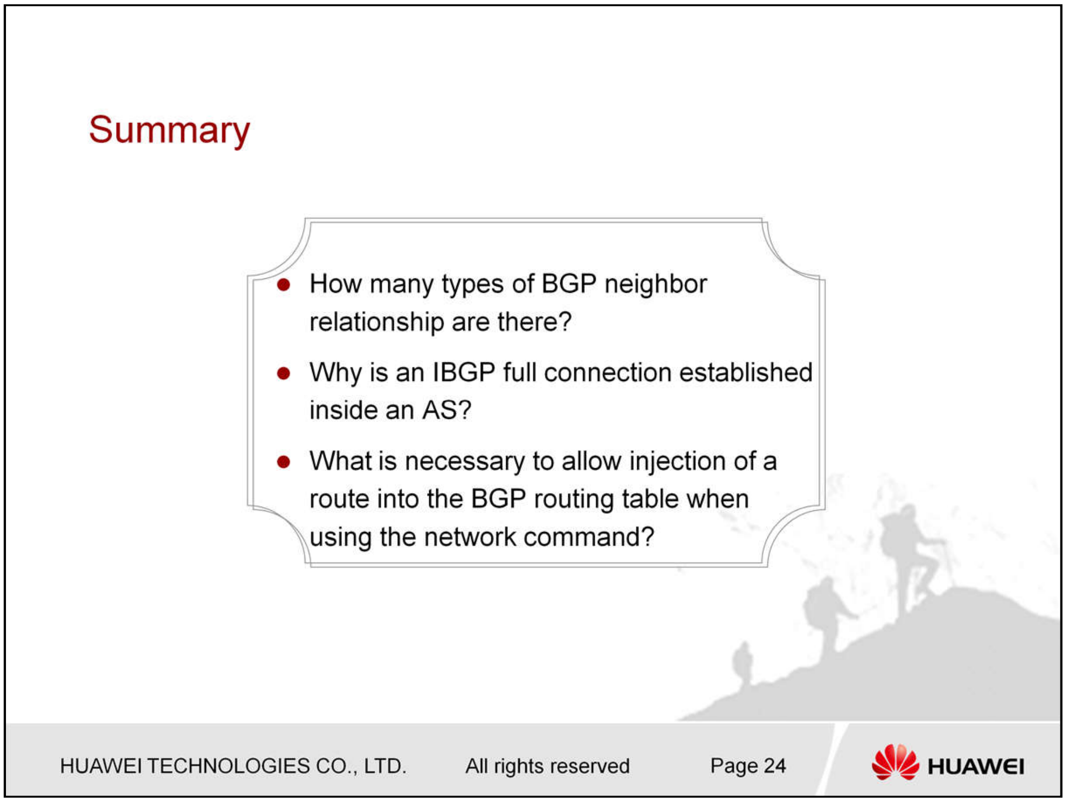
The second method is redistributing other routing protocols into BGP routing table by using the import command. Besides, we can specify the routing policy for the imported route.

import-route protocol [process-id] [med med | route-policy route-policy-name]

protocol: specifies the importable external routing protocols. At present, the importable external routing protocols include isis, ospf, static, direct and rip.

process-id: specifies the process id if the imported routing protocol are isis, ospf or rip.

Med: specifies the med metric of the imported route. The value is ranging from 0 to 65535. route-policy-name: When the routes are imported from other routing protocol, will filter those routes by using the specified routing policy.



1. How many types BGP relationship?

A：IBGP and EBGP. If two peers that exchange BGP messages belong to the same AS, they are Internal BGP (IBGP). If two peers that exchange BGP messages belong to different AS, they are External BGP (EBGP).

2. Why the IBGP full connection is established inside an AS?

A：For the route obtained from IBGP, the BGP speaker will not advertise them to its IBGP neighbors. This is to prevent the routing loop inside an AS. So, IBGP full connection must be established between the BGP speakers inside an AS in order to transmit the routing information to all other IBGP peers

successfully.

3. What is the requirement to inject the route into the BGP routing table by using network command.

A：The routes to be injected into the BGP routing table via network command must be available inside the IP routing table. In addition, we must specify the length of the subnet mask precisely.

Thank You

www.huawei.com