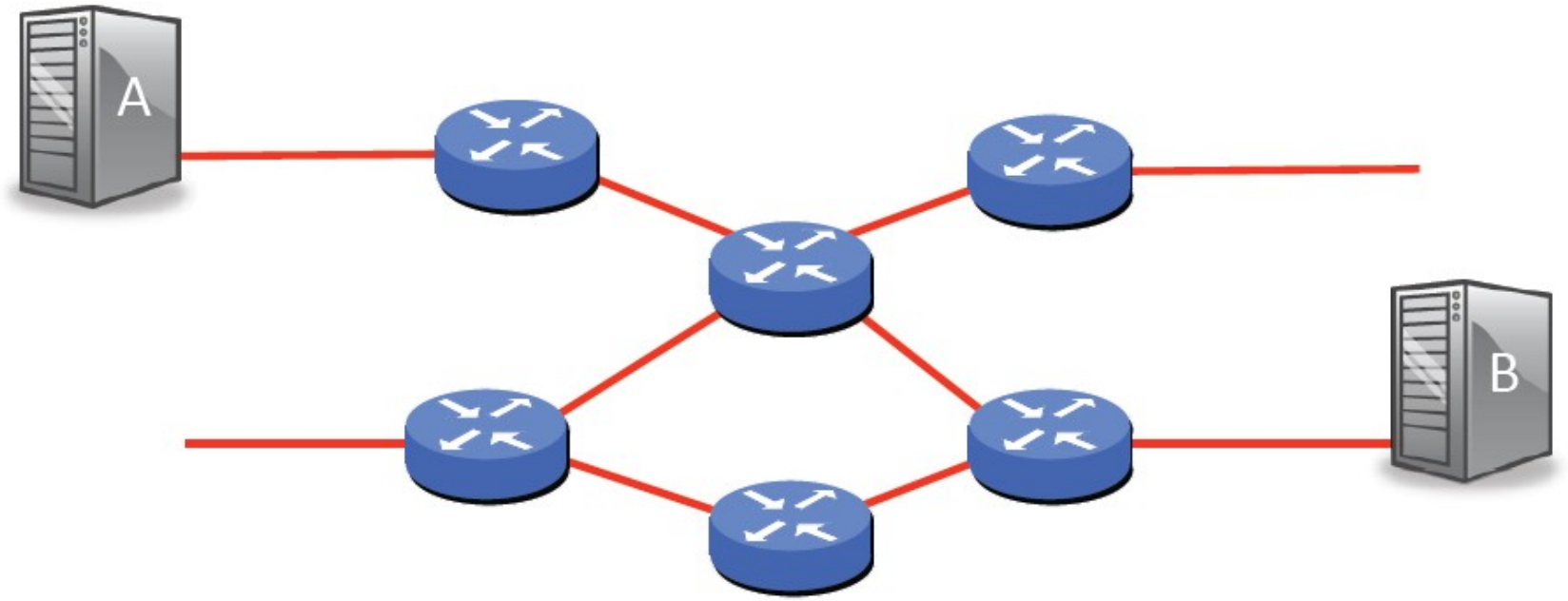
Routing

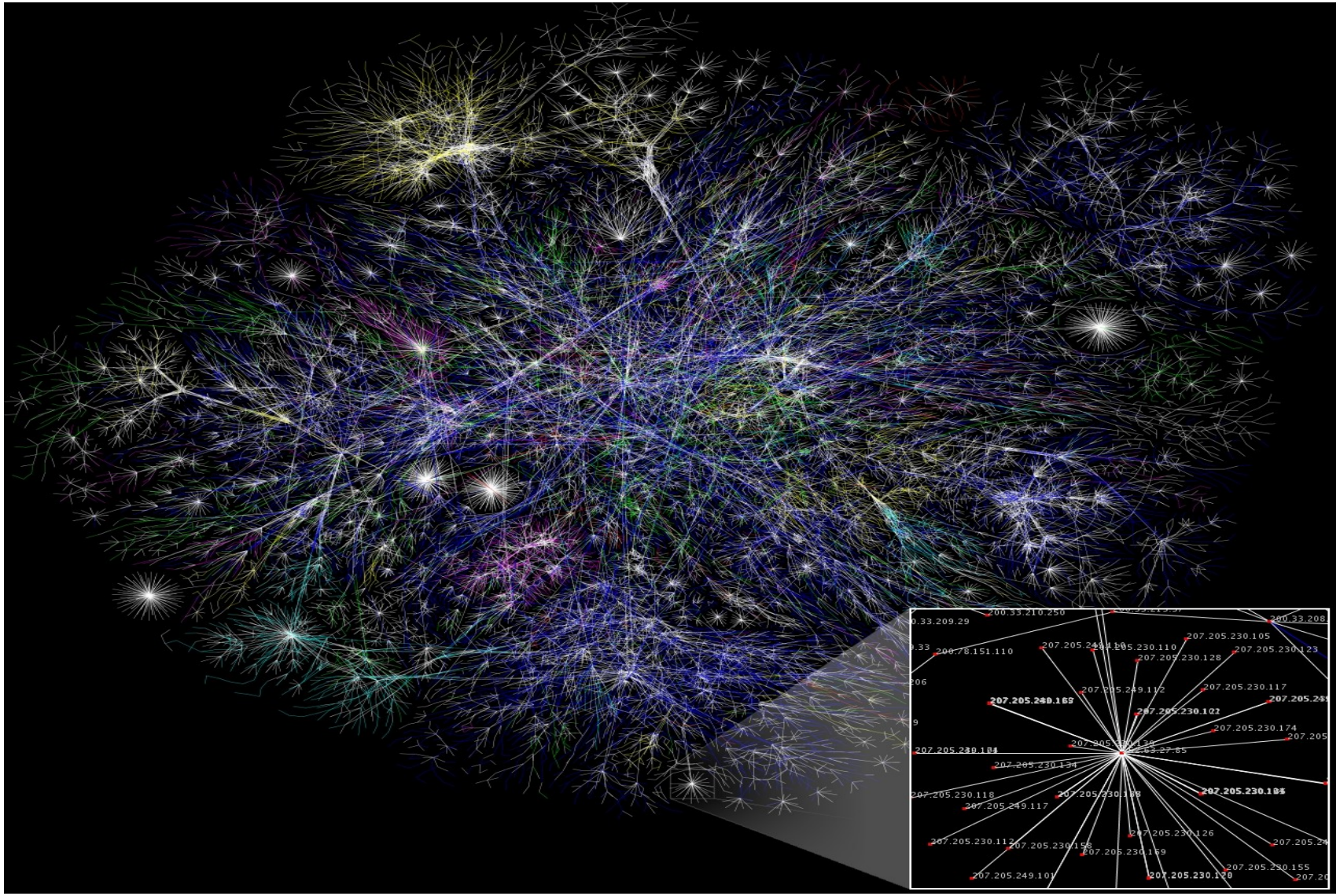
Problem



Who should determine how packets from A can reache B are ?

Route selection criteria?

What we can do with such a network ?



Autonomous system – AS

AS — *it is a system of IP networks and routers managed by one or more operators with a single routing policy.*

ASN — Autonomous system number

Globally, this is done by the IANA (Internet Assigned Numbers Authority), which delegates its tasks to the RIR (Regional Internet Registry) – i.e. regional organizations, each of which is responsible for a certain part of the planet (for Europe and Russia – it is RIPE NCC), which in turn delegates its tasks to the LIR (Local Internet Registry).

Autonomous system

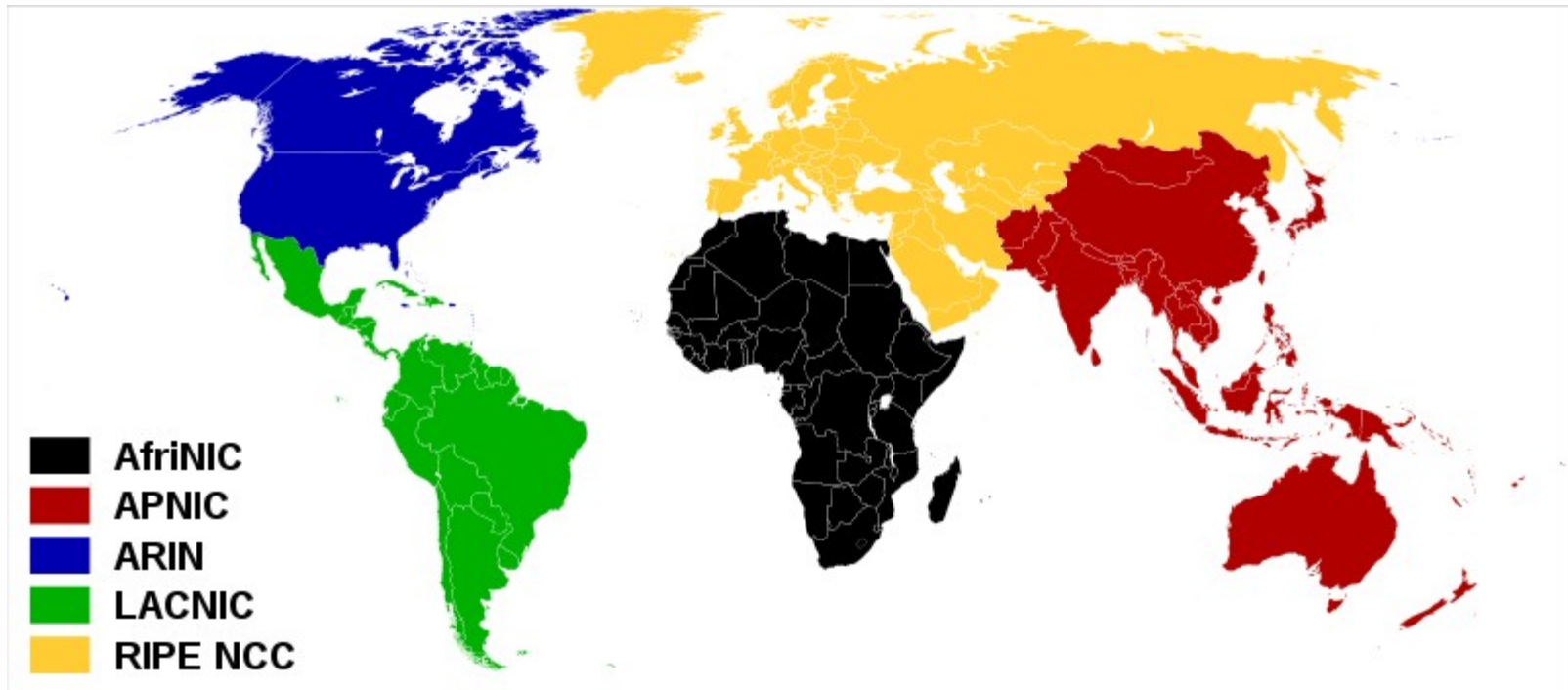
Unit of hierarchy in the Internet

Inside the AS its owner decides how to route data flows

A Protocol BGP-4 (Border Gateway Protocol v4 RFC 1771) must be used between the speakers

How to find AS number?

http://whatismyipaddress.com/ip-lookup


AfriNIC
APNIC
ARIN
LACNIC
RIPE NCC

RIR & LIR

Almost any organization can become a LIR if it has the necessary documents.

At LIR'a we would take ASN (as number) - for example 64500. And he himself could have ASN 64501.

Until 2007, only 16-bit AS numbers were possible, that is, a total of 65536 numbers were available. 0 and 65535 are reserved.
Numbers 64512 through 65534 are for private AS that are not routed globally
Rooms 64496-64511 – for use in examples and documentation.

Now it is possible to use 32-bit AS numbers.

It is impossible to speak about Autonomous systems without binding to blocks of IP addresses. In practice, each AS must have some block of addresses associated with it.

Protocol OSPF

Routing algorithms

Routing involves two parallel process:
- ***preparation of routing table***
- ***forwarding of the datagrams (with using this table).***

Formation of the route table produces:
- ***through the use routing protocol (dynamic routing)***
- ***under the influence of instructions network administrator – (static routing.)***

Routing algorithms

Routing algorithms solve the problem of path selection.

- ***unicast***
- ***multicast.***
- ***source routing***
- ***«hot potato».***
- ***……..***

Types of routing protocols

- *IGP (internal to your Autonomous system) - ISIS/OSPF/RIP/EIGRP.*
- *EGP (внешние) - BGP – Border Gateway Protocol.*

According to the types of algorithms used are divided into:

- *DV (Distance Vector)*
- *LS (Link State).*

Metrics

- *Typically, each segment that makes up a route, some value is assigned-an estimate of this segment's.*
- *Each routing Protocol uses its own route evaluation system.*

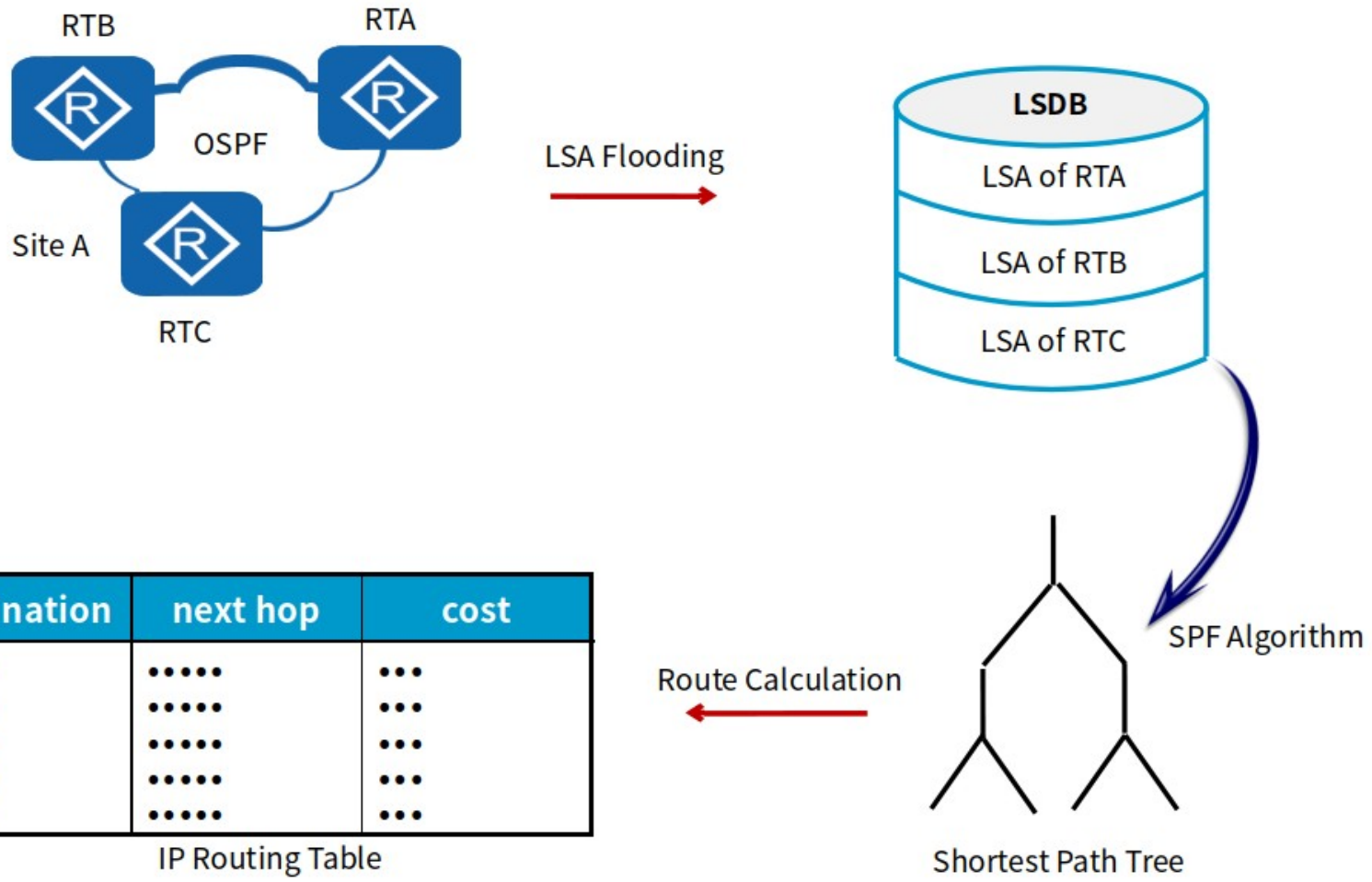
Examples:

- *RIP – hop count*
- *OSPF – based on the link bandwidth*

OSPF

- *OSPF (RFC 2328)*
- *line state changes are sent as needed*
- *each router uses Dijkstra's algorithm*
- *Can use authentication*
- *AS (AS in OSPF differs from AS in the Internet) can be divided in to areas*
- *analogue - IS-IS (RFC 1142)*

OSPF


RTB
RTA
OSPF
Site A
RTC
LSA Flooding
LSDB
LSA of RTA
LSA of RTB
LSA of RTC
SPF Algorithm
Route Calculation
Shortest Path Tree
destination
next hop
cost
IP Routing Table

Stages of the protocol OSPF

1) Routers exchange hello packets across all interfaces on which OSPF is activated. Routers that share a common data link become neighbors when they agree on certain parameters specified in their hello packets.

2) In the next phase of the Protocol, the routers will attempt to move into an adjacency state with their neighbors (for the shared bus networks like Ethernet).

3) Each router sends the so called link state advertisements (LSA) to the routers with which it is in the adjacency state

Stages of the protocol OSPF (continuation)

4)Each router that receives an advertisement from an adjacent router writes the information it transmits to the router's link state database and sends a copy of the advertisement to all other adjacent routers.

5)By sending advertisements within the same OSPF zone, all routers build an identical router link state database.

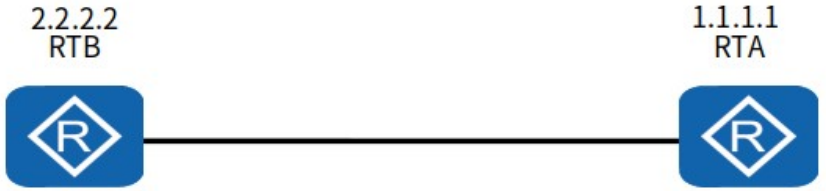
6)Each router independently runs Dijkstra's shortest path algorithm

1)Each router builds a routing table from its shortest path tree

OSPF roughter types

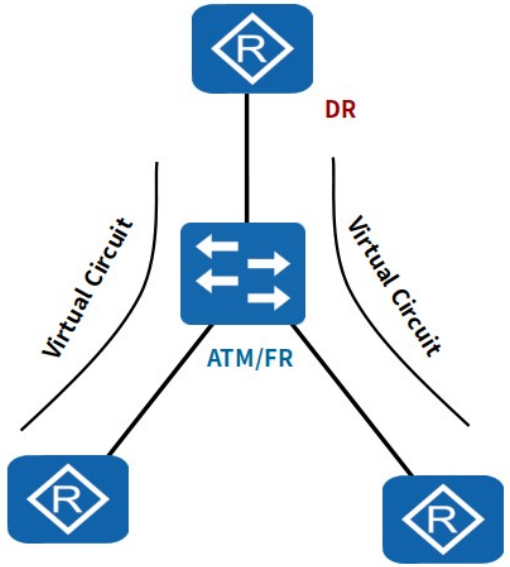
- ***Internal router*** *— a router whose interfaces all belong to the same zone. These routers have only one link-state database.*
- ***Area border router, ABR*** *— connects one or more zones to the backbone zone and acts as a gateway for cross-zone traffic.*
- ***AS boundary router, ASBR*** *— a router that has one port in the OSPF domain and another port in the domain of any of the internal gateway protocols (for example, RIP or EIGRP).*

OSPF Supported Network Types

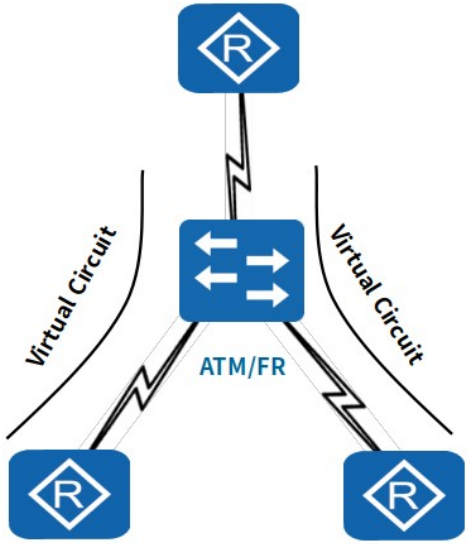


Broadcast

Point-to-Point



Non-Broadcast Multi-Access (NBMA)

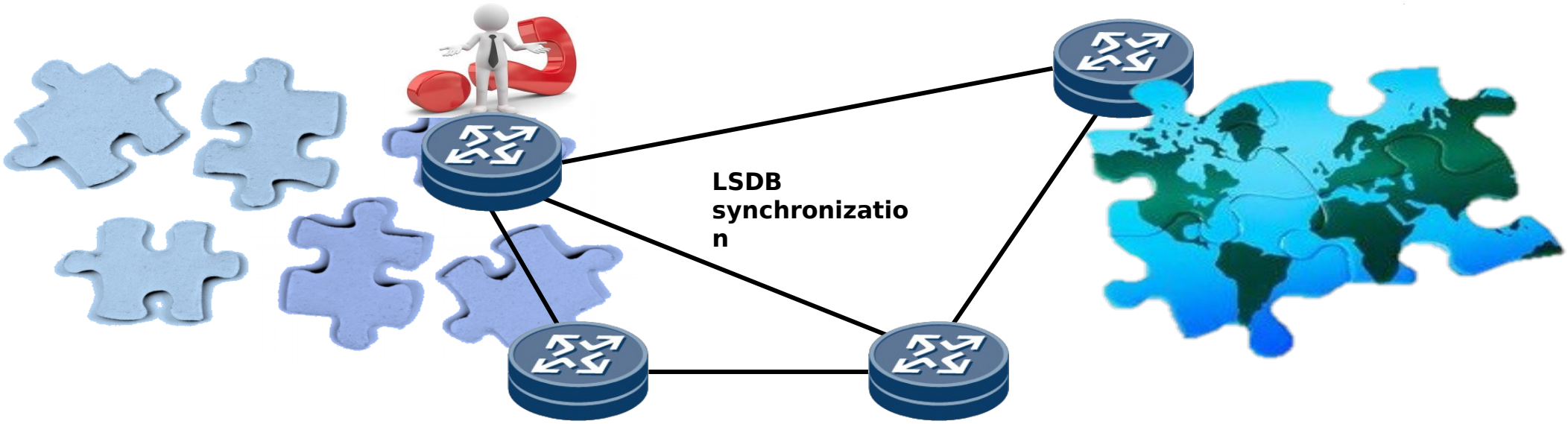


Point to Multi-Point

Roughter ID

Every OSPF router must have unique router ID.

Who am I and who are others?
Each router must be identified
in an LSDB.



OSPF pakets types

Type	Packet Type	Packet Function
1	Hello	To discover and maintain OSPF neighbor relationships.
2	Database Description (DD)	To exchange LSDB summary.
3	Link State Request (LSR)	To request specific link state information.
4	Link State Update (LSU)	To send requested link state information.
5	Link State Ack (LSAck)	To acknowledge receipt of an LSA.

OSPF packet header format

All five types of OSPF packets are encapsulated directly in an IP packet. The OSPF Protocol number in the IP header is 89.

All OSPF packets have the same header

OSPF packet header fields:

***Version number** — the version of the OSPF Protocol. The current version is for IPv4-2.*

***Packet type**-specifies which type of OSPF packet is sent:*

1 — Hello

2 — Database Description

3 — Link State Request

4 — Link State Update

5 — Link State Acknowledgment

***Packet length** — the length of the OSPF packet in bytes. The length includes the header.*

***Router ID**-determines which router sent the packet.*

***Area ID**-determines in which zone the packet is generated.*

***Checksum**-used to check the integrity of the OSPF packet, to detect errors during transmission.*

OSPF packet header format (c)

Authentication type *— the type of authentication that is used between routers:*

0 - authentication is not used,

1- clear text authentication,

2- MD5 authentication.

Authentication data *- used when authenticating routers.*

The ***Data*** *field is different for different types of OSPF packets:*

Hello*-list of famous neighbors*

DD *- contains the summary information of the channel state database, which includes all known router IDs and their latest sequence number and other information.*

LSR *- contains the type of LSU required and the ID of the router that has this LSU.*

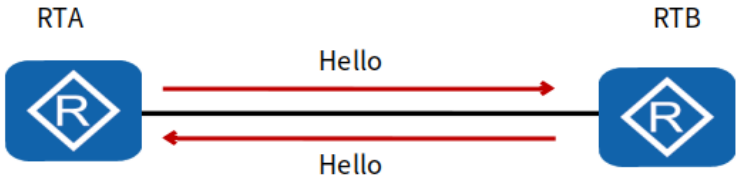
LSU *- contains a complete record of announcements about the state of the channel. Multiple LSAs can be transferred in a single service pack.*

LSAck *- field empty*

Hello package header format

With the Hello packet, the router detects its neighbors;

- *It sends parameters that routers need to negotiate before they become neighbors;*
- *Hello packets act as keepalive packets between neighbors;*
- *Responsible for establishing two-way communications between neighboring routers (two-way communication is established when the router sees itself in the list of neighbors hello-packet received from a neighboring router);*
- *It is used when selecting DR and BDR in broadcast and non-broadcast multiple access networks.*



Hello Interval	Options	Router Priority
Router Dead Interval		
Designated Router		
Backup Designated Router		
Neighbor		

OSPF types of LSA

Type 1 LSA *— Router LSA — router channel status announcement.*
Type 2 LSA *— Network LSA — the announcement about the state of the network. DR is distributed in multiple-access networks.*
Type 3 LSA *— Network Summary LSA — summary announcement of the status of the network channels. The ad is distributed by edge routers.*
Type 4 LSA *— ASBR Summary LSA — summary announcement of the state of the Autonomous system edge router channels.*
Type 5 LSA *— AS External LSA — announcement of the status of the external channels of the Autonomous system.*
Type 6 LSA *— Multicast OSPF LSA — specialized LSA that use multicast OSPF applications.*
Type 7 LSA *— AS External LSA for NSSA — announcements about the status of the external channels of the Autonomous system in the NSSA area.*

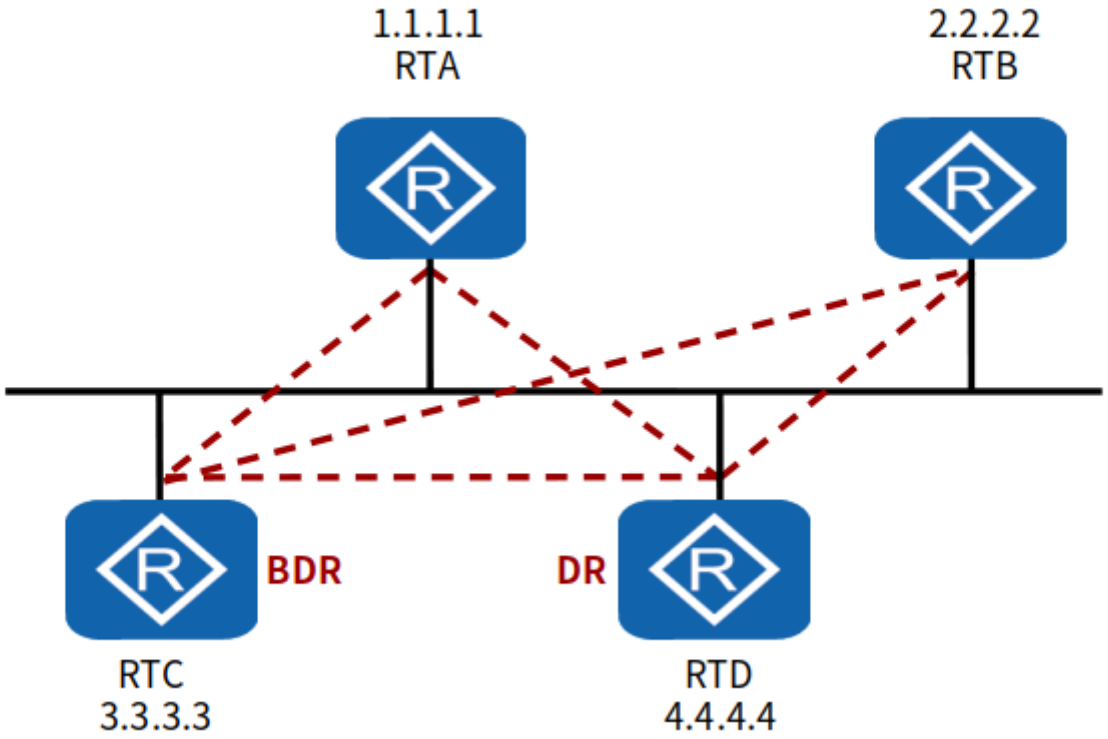
OSPF types of LSA

Type 8 LSA *— Link LSA — announces link-local address and prefix(s) router to all routers sharing a channel (link).*
Sent only if more than one channel is present router. Distributed only within the channel (link).
Type 9 LSA *— Intra-Area-Prefix LSA puts in compliance: list IPv6 prefixes the router pointing to the Router LSA, list IPv6 prefixes and the transit network pointing to the Network LSA.*
Distributed only within one zone.

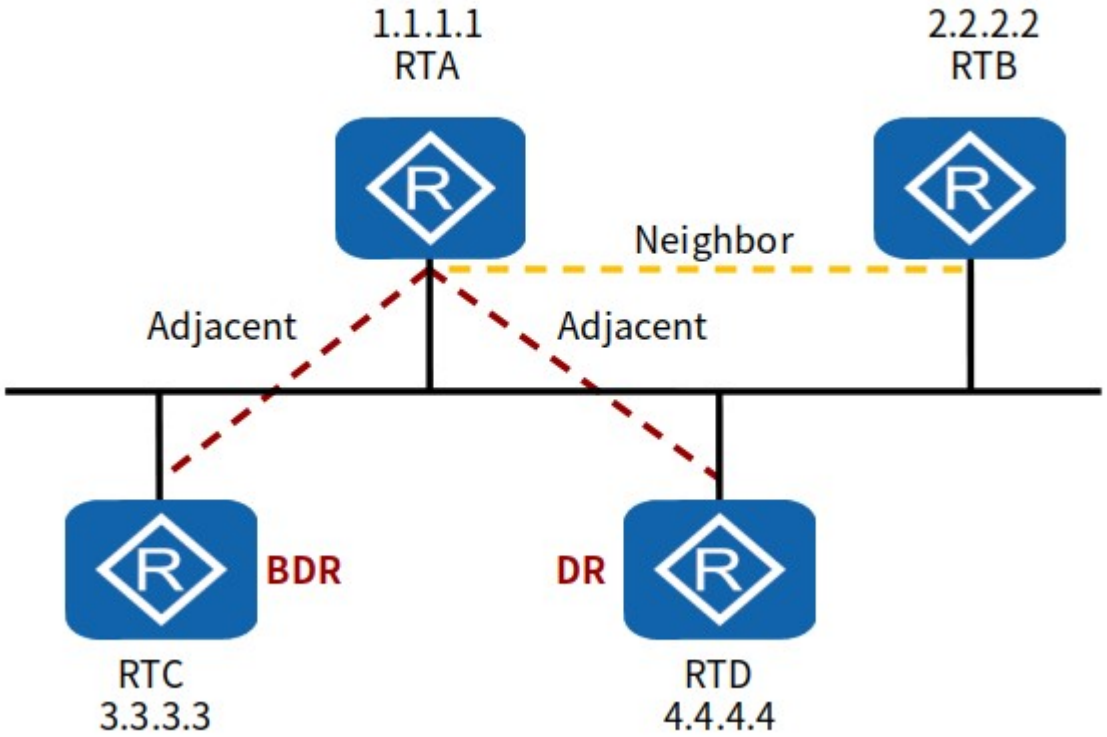
OSPF areas types

- *backbone area*
- *standard area*
- *stub area*
- *Totally stubby area*
- *Not-so-stubby area (NSSA)*

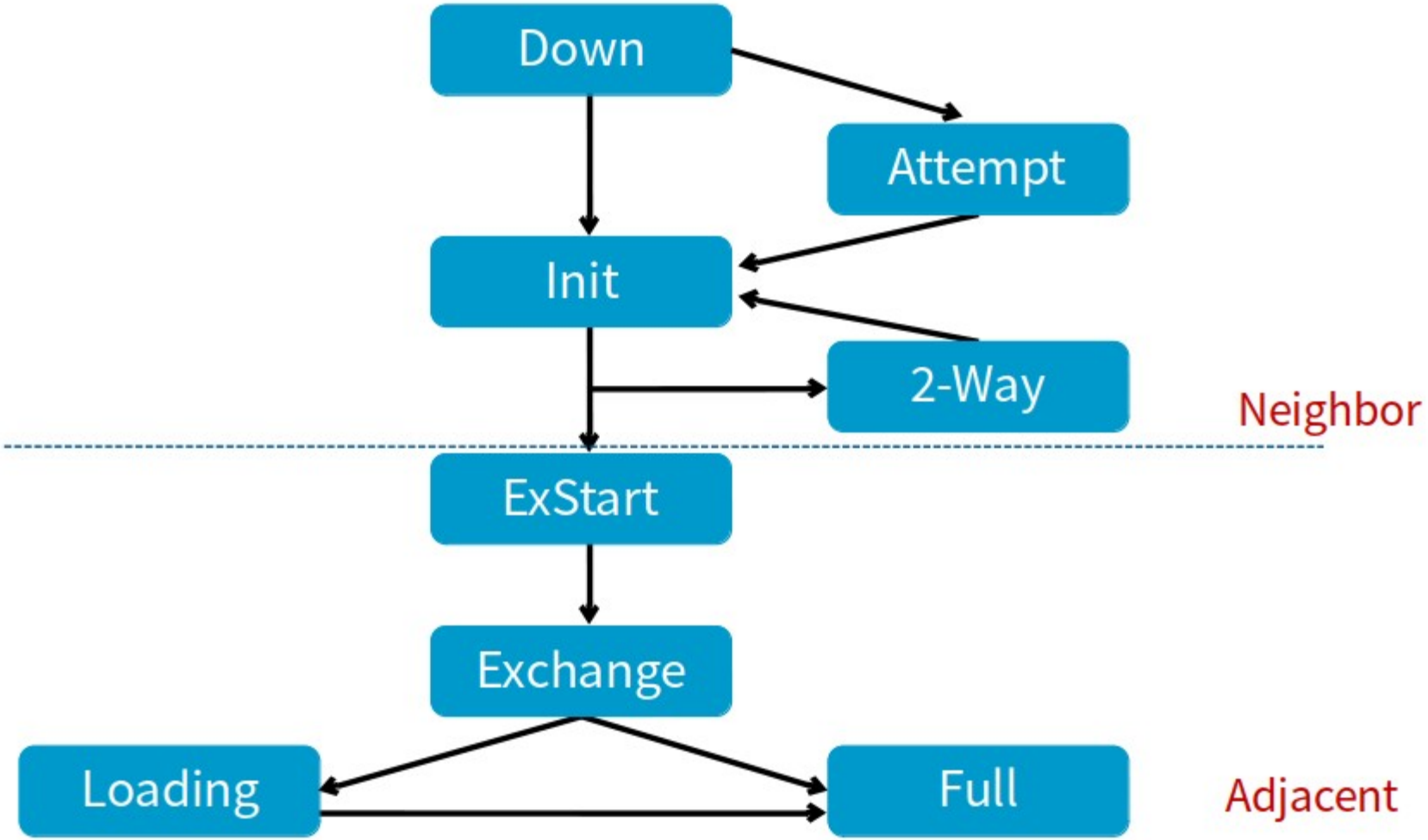
Designated Router & Backup Designated Router



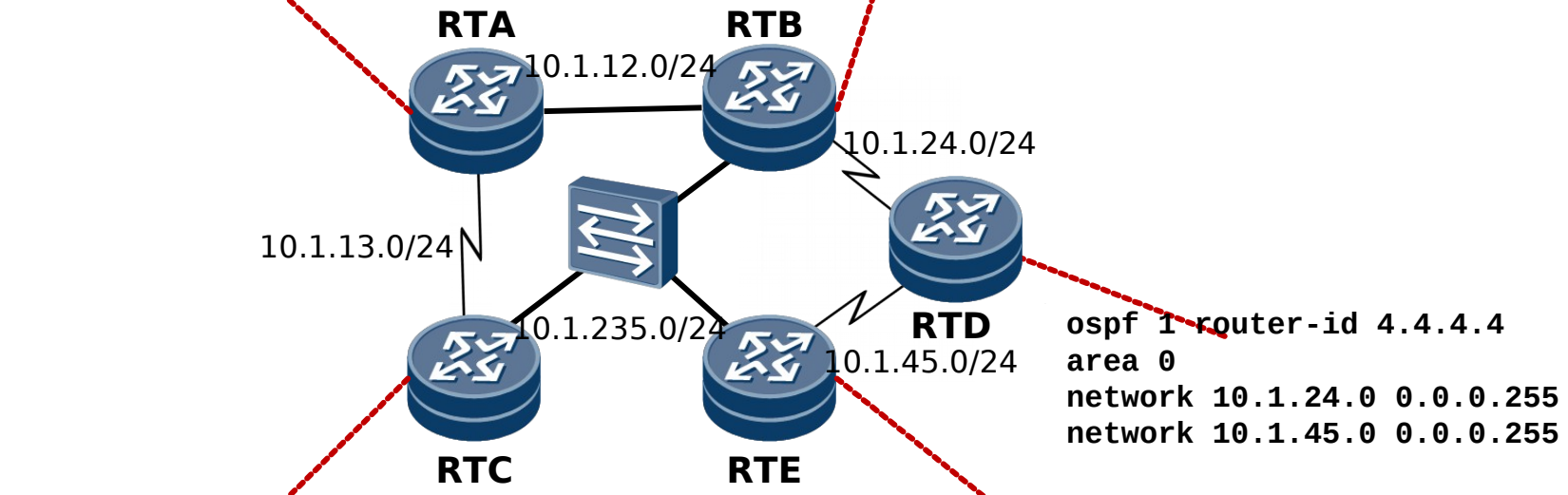
OSPF DR и BDR


1.1.1.1
RTA
2.2.2.2
RTB
Neighbor
Adjacent
Adjacent
BDR
DR
RTC
3.3.3.3
RTD
4.4.4.4

OSPF status


Down
Attempt
Init
2-Way
Neighbor
ExStart
Exchange
Loading
Full
Adjacent

OSPF Example



BGP

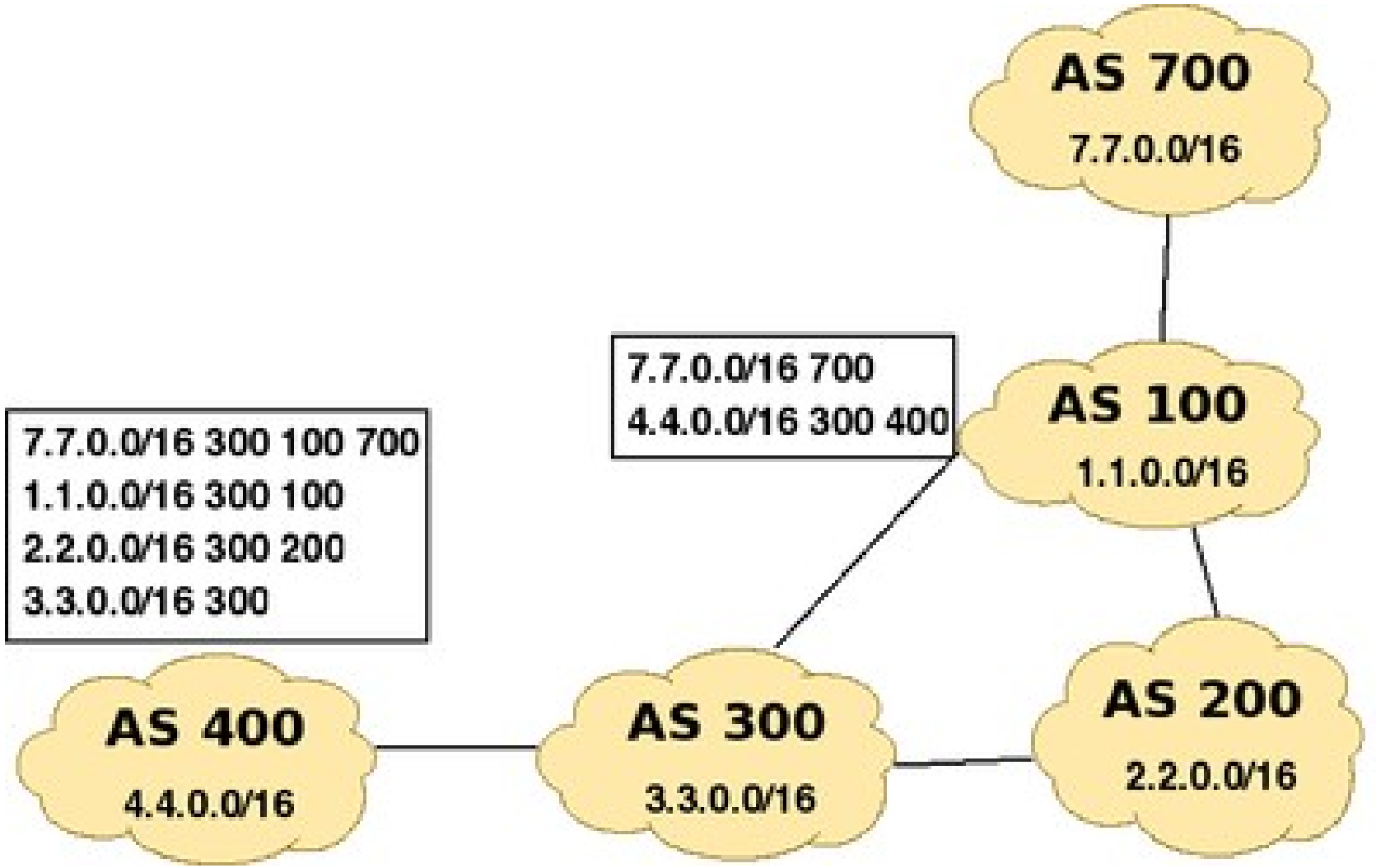
Border Gateway Protocol (BGP-4)
Main terms

- *autonomous system, AS*
- *transit AS*
- *path*
- *path attributes, PA*
- *BGP speaker*
- *neighbor, peer*
- *Network Layer Reachability Information, NLRI*

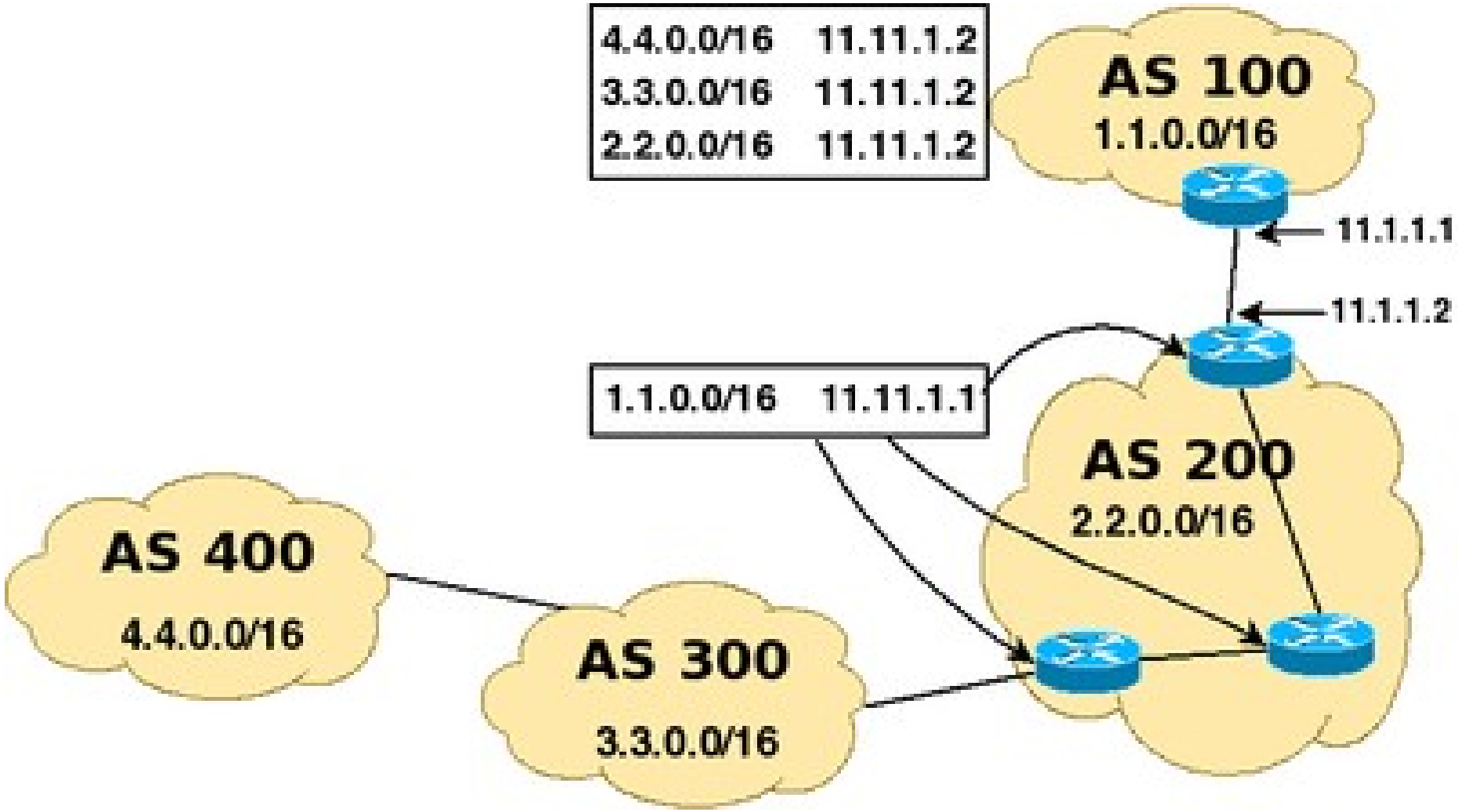
Border Gateway Protocol (BGP-4)

- BGP use «path vector»
- Each BGP router sends out a list of paths (path-AS list)
  - AS_PATH
  - The 171.64/16 network can be accessed by following the route {AS7,AS52,AS13}
- The presence of a loop in a route is determined locally and such routes are ignored
- From the variety of available routes, the one that most corresponds to the AS policy is selected
- If the router/lines fail, the route is removed from the list

AS list


AS 700
7.7.0.0/16
7.7.0.0/16 700
4.4.0.0/16 300 400
AS 100
1.1.0.0/16
7.7.0.0/16 300 100 700
1.1.0.0/16 300 100
2.2.0.0/16 300 200
3.3.0.0/16 300
AS 400
4.4.0.0/16
AS 300
3.3.0.0/16
AS 200
2.2.0.0/16

Next hop


4.4.0.0/16 11.11.1.2
3.3.0.0/16 11.11.1.2
2.2.0.0/16 11.11.1.2
AS 100
1.1.0.0/16
11.1.1.1
11.1.1.2
1.1.0.0/16 11.11.1.1
AS 200
2.2.0.0/16
AS 400
4.4.0.0/16
AS 300
3.3.0.0/16

BGP packets

- ***Open*** **-** BGP session setup (TCP port 179)
- ***Keep Alive -*** Health check at regular intervals
- ***Notification*** **-** Closing peering session
- ***Update*** **-** Advertising a new route or withdrawing a previously advertised route

BGP advertisement = prefix + path attributes

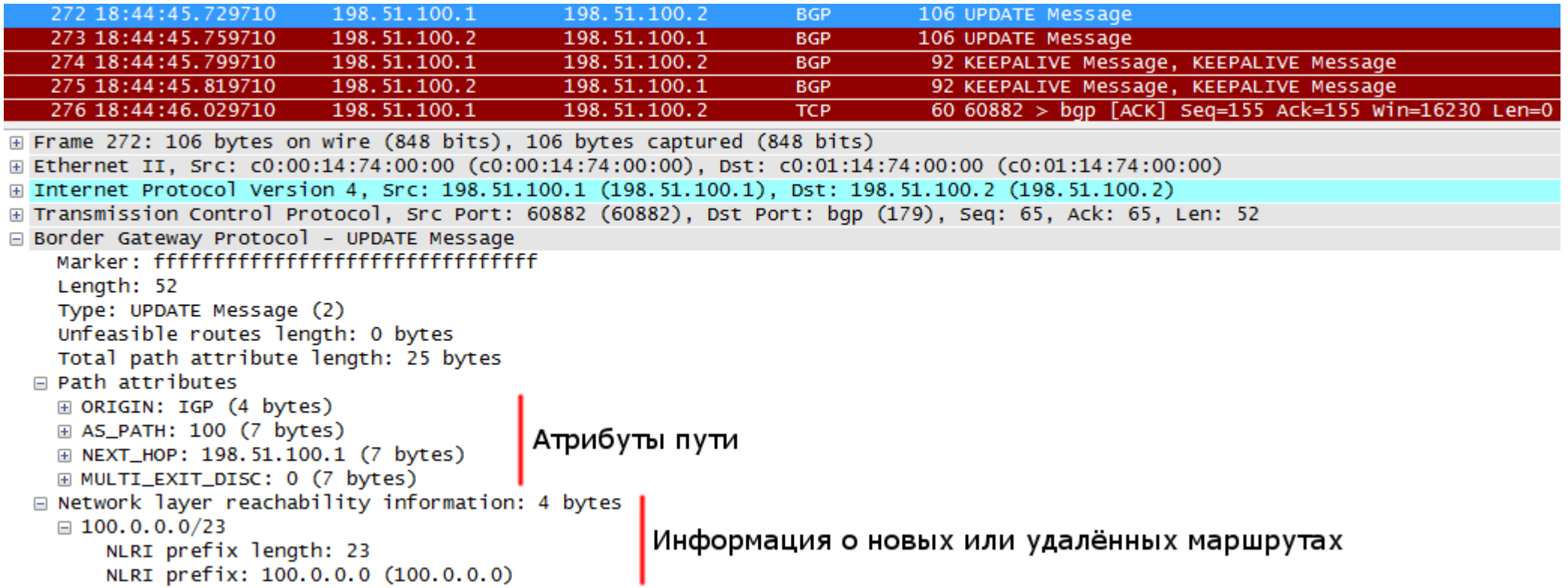
- **Path attributes**
  - Origin - who suggested this route
  - AS path - preferences, the rule for determining the output
  - next hop - IP address of BGP gateway of the next AS,
  - multi_exit_disc – how to select one of several gateways advertizing the rought
  - Local pref - prefixes of local speakers
  - Aggregate – combining several routes with a common prefix

Open message

```
   11 18:30:57.802710    198.51.100.1    198.51.100.2    TCP      60 33887 > bgp [SYN] Seq=0 Win=16384 Len=0 MSS=1460
   12 18:30:57.822710    198.51.100.2    198.51.100.1    TCP      60 bgp > 33887 [SYN, ACK] Seq=0 Ack=1 Win=16384 Len=0 MSS=1460
   13 18:30:57.832710    198.51.100.1    198.51.100.2    TCP      60 33887 > bgp [ACK] Seq=1 Ack=1 Win=16384 Len=0
   14 18:30:57.852710    198.51.100.1    198.51.100.2    BGP      99 OPEN Message
   15 18:30:57.872710    198.51.100.2    198.51.100.1    BGP     118 OPEN Message, KEEPALIVE Message
   16 18:30:57.892710    198.51.100.1    198.51.100.2    BGP      73 KEEPALIVE Message
   17 18:30:57.902710    198.51.100.1    198.51.100.2    BGP      92 KEEPALIVE Message, KEEPALIVE Message
   18 18:30:57.922710    198.51.100.2    198.51.100.1    BGP      92 KEEPALIVE Message, KEEPALIVE Message
   19 18:30:58.132710    198.51.100.1    198.51.100.2    TCP      60 33887 > bgp [ACK] Seq=103 Ack=103 Win=16282 Len=0
```

```
⊞ Frame 14: 99 bytes on wire (792 bits), 99 bytes captured (792 bits)
⊞ Ethernet II, Src: c0:00:14:74:00:00 (c0:00:14:74:00:00), Dst: c0:01:14:74:00:00 (c0:01:14:74:00:00)
⊞ Internet Protocol Version 4, Src: 198.51.100.1 (198.51.100.1), Dst: 198.51.100.2 (198.51.100.2)
⊞ Transmission Control Protocol, Src Port: 33887 (33887), Dst Port: bgp (179), Seq: 1, Ack: 1, Len: 45
⊟ Border Gateway Protocol - OPEN Message
    Marker: ffffffffffffffffffffffffffffffff
    Length: 45
    Type: OPEN Message (1)
    Version: 4
    My AS: 100
    Hold Time: 180
    BGP Identifier: 198.51.100.1 (198.51.100.1)
    Optional Parameters Length: 16
  ⊟ Optional Parameters
    ⊟ Optional Parameter: Capability
        Parameter Type: Capability (2)
        Parameter Length: 6
      ⊞ Capability: Multiprotocol extensions capability
    ⊟ Optional Parameter: Capability
        Parameter Type: Capability (2)
        Parameter Length: 2
      ⊞ Capability: Route refresh capability
    ⊞ Optional Parameter: Capability
```

Update message



Border Gateway Protocol (BGP-4)
Выбор пути

If path selection policies are not configured, selection occurs as follows :

1. Maximum importance weight (local to router Cisco).
2. The maximum value of local preference (for the entire AS).
3. Prefer local router route (next hop = 0.0.0.0).
4. The shortest path through Autonomous systems. (the short AS_PATH)
5. The minimum value of the origin code (IGP < EGP < incomplete).
6. Minimum MED value (spread between Autonomous system).
7. The eBGP path is better than the iBGP path.
8. Choose a path through the nearest IGP neighbor.
9. Select the oldest route for the eBGP path.
10. Choose a path through the neighbor with the smallest BGP router ID.
11. Select the path through the neighbor with the lowest IP address.

Full View и Default Route

Full View router learns all the routes of the Internet.
There are more than 600 000 such routes.
If you use 2 providers, you can get more than a million routes.

- ***Full View**. Full information about the structure of the Internet. To any address on the Internet you can view the path from the current device on the network*
- ***Load distribution**. This is achieved by setting, for example, route priorities for specific prefixes*
- *Full View is **required** if your AU is transit*

Default Route-only one default route comes from each provider

- *Saving equipment resources*
- *Easy maintenance*

BGP

- *All users to interact on the Internet must use BGP-4*
- *BGP-4 uses a path vector routing algorithm that easily recognizes loops*
- *BGP-4 has a complex interface that allows each AS to set its own local routing policy*
- *Each AS sets its own policies for routing, security, and local features*

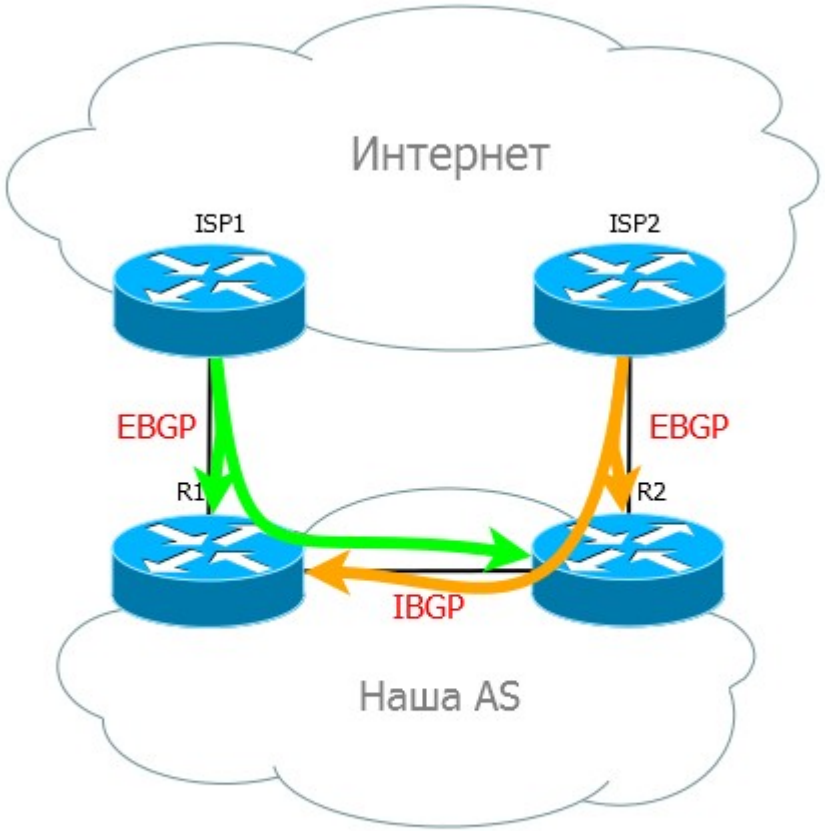
What is IBGP

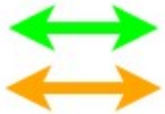
Let's start with what Internal BGP is. In fact, it is the same BGP, but inside the AS.

Reservation. *When there are several links to providers and it is undesirable to have all of them on one border router, several routers are put, and IBGP rises between them in order that on them there was always up-to-date information on all routes.*

BGP client connection. If the task is to connect the client via BGP, and you have more than one router, IBGP is used .

What is IBGP





Маршрутная информация

IBGP & EBGP differences

1) Resistance to loops generation

- *EBGP handles loops with AS-Path. If there was already an AS number of the local AS in the list, this route is discarded.*
- *When you advertize a route inside an Autonomous System, the AS-Path does not change. IBGP uses a fully connected Full Mesh.*
- *In this case, the route received from the IBGP neighbor is not announced to other IBGP neighbors.*
- *This allows all routers to have all routes and still avoid loops.*

IBGP & EBGP differences

2) Next Hop address.

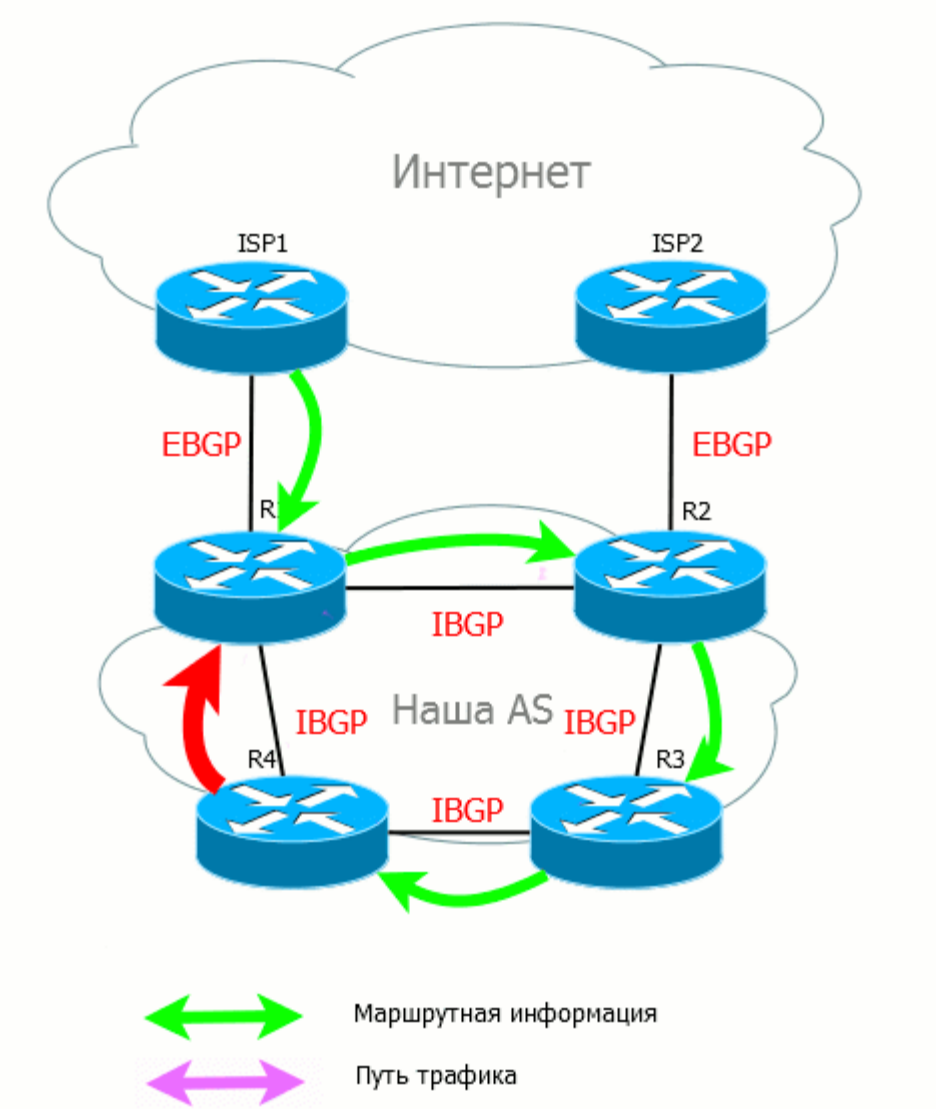
- *In External BGP, when a router sends an announcement to its EBGP neighbor, it first changes the Next-Hop address to his own one, and then sends.*
- *If the router sends an IBGP announcement to a neighbor, the next-Hop address does not change.*
- *The concept of Next-Hop is different from that used in IGP. In IBGP, it reports the exit point from the local AS.*
- *It is important that the recipient of such an announcement has a route to Next-Hop. If it is not, the route will be placed in the BGP table, but it will not be in the routing table.*

IBGP & EBGP differences

3) Connecting neighbors

- *While EBGP typically involves two neighbors connecting directly to each other, in Internal BGP, neighbors can be connected through multiple intermediate devices.*
- *in EBGP, you can also configure neighbors that are several intermediate devices away from each other, but for IBGP, this works by default.*
- *allows IBGP partnership between Loopback addresses, EBGP does not use this*

What is IBGP



- *In the case of a fully connected topology and the Split Horizon rule, a situation is excluded when even a new route from R4 can get out as a priority, this is inefficient, since the routes will be studied incorrectly, a loop may form and the traffic will not get to the destination point.*
- *R1 received the announcement from ISP1, transmits it immediately to all its neighbors: R2, R3, R4. And those, in turn, keep these announcements, but transmit only to EBGP-partners, but not IBGP, precisely because they are received from the IBGP-partner. That is, all BGP routers have up-to-date information and loops are excluded.*

Interaction with OSPF

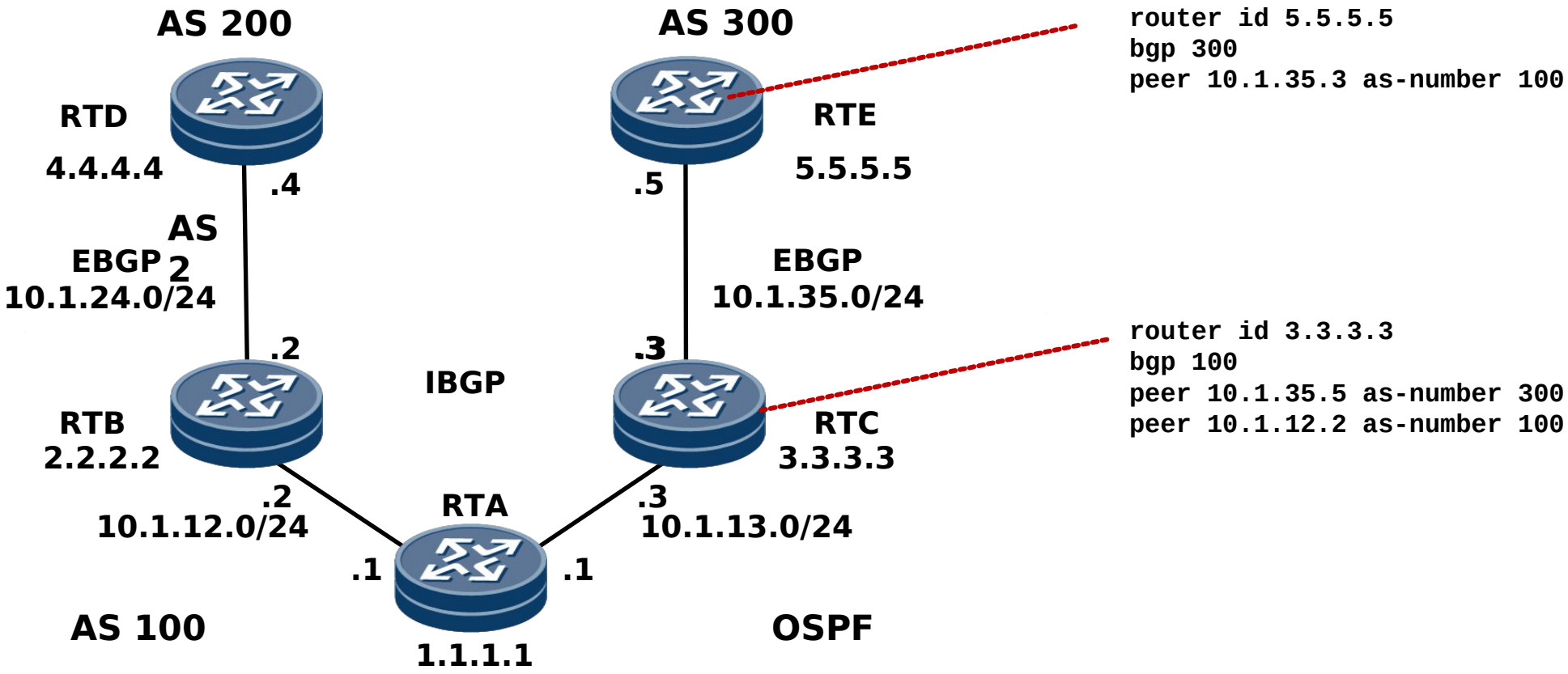
Typically, IBGP is configured using IGP existing on the network.

***IGP provides**:*

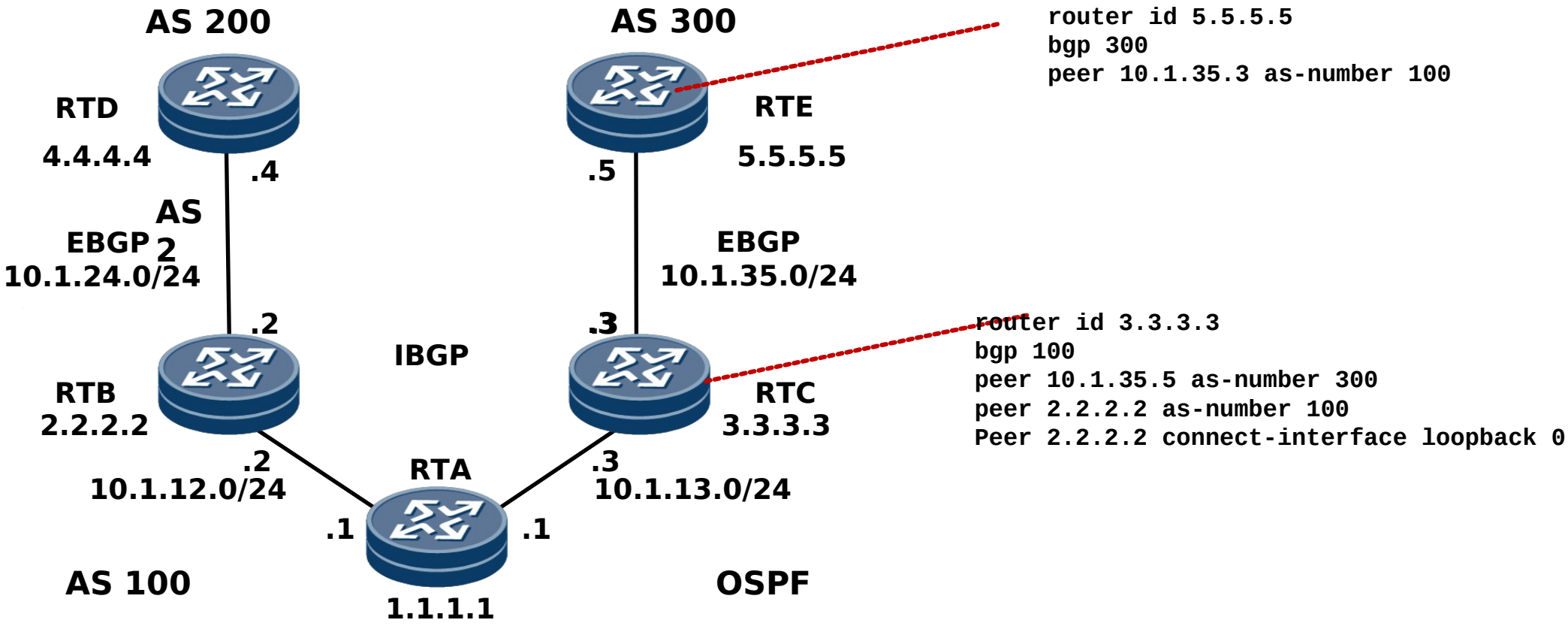
- *IP connectivity of all routers,*
- *quick response to changes in the topology*
- *transfer route information about internal networks.*

1. Use Loopback interfaces :
These settings will be used to determine the Router ID for both OSPF and BGP.
2. Configure internal OSPF routing

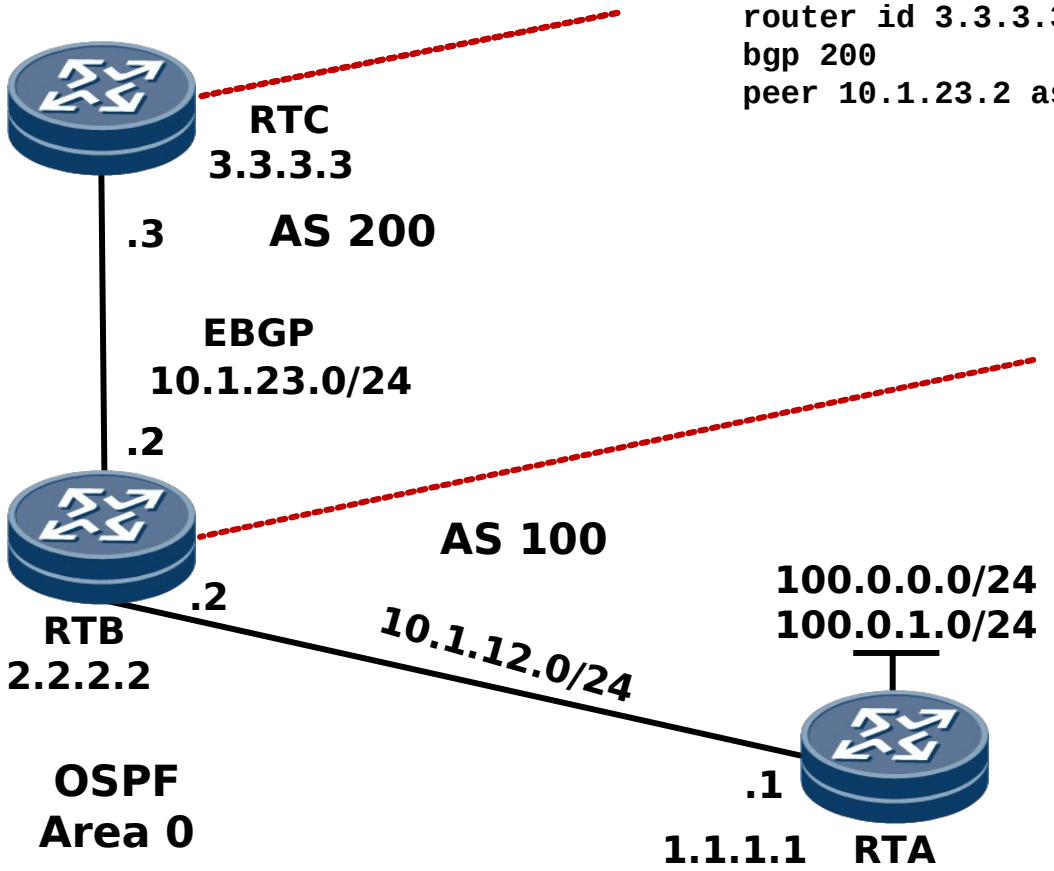
BGP - Setting up interaction with neighbors



BGP - Setting up interaction with neighbors



BGP - Configure the advertisement of routes

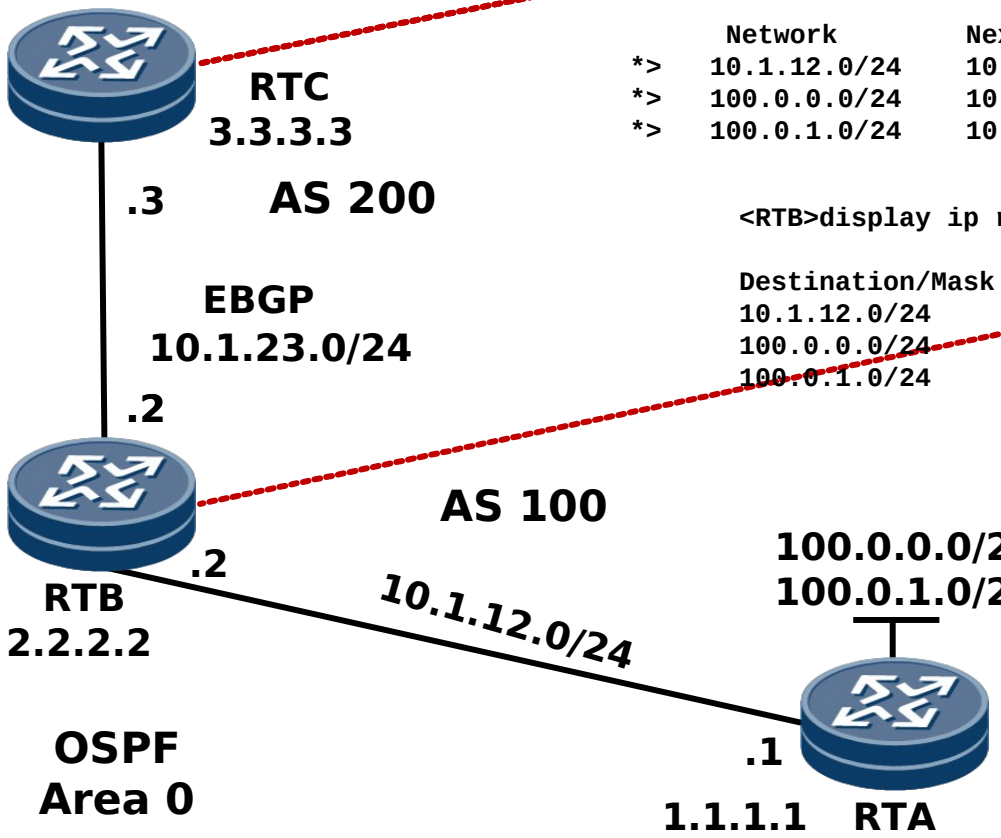



```
router id 3.3.3.3
bgp 200
peer 10.1.23.2 as-number 100
```

```
router id 2.2.2.2
bgp 100
peer 10.1.23.2 as-number 200
 ipv4-family unicast
  network 10.1.12.0 255.255.255.0
  network 100.0.0.0 255.255.255.0
  network 100.0.1.0 255.255.255.0
```

The *network* command is used to import routes that exist in the routing table into the BGP route table.

BGP — routes advertisement check




```
<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
```

Can BGP work without IGP

- *IGP and IBGP work together, each doing their job.*
- *IGP provides internal IP connectivity, response to changes in the network, notification of all nodes. He knows about the public addresses of the AS.*
- *IBGP handles external routes in the AS and their transit to and from customers. Usually he knows nothing about the structure of the internal network.*

Additional features of IBGP

Route Reflector

Route Reflector is a special IBGP router that performs the function of "reflecting " routes — he gets a route from one neighbor, and sends it to all others. That is, on IBGP routers, you need to configure a session with only one neighbor — with Route Reflector. Direct analogy - DR in OSPF.

Additional features of BGP

MP-BGP and Address Families

The Border Gateway Protocol 4 (BGP-4) transmits only IPv4 unicast routing information and cannot transmit routing information for other network layer protocols, such as IPv6 and multicast protocols.

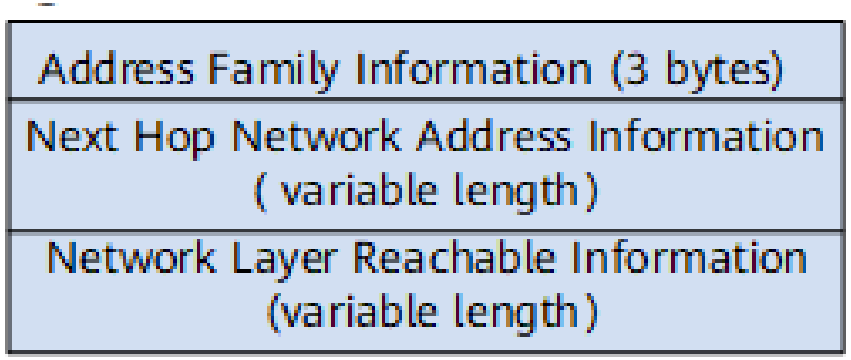
To support multiple types of network layer protocols, the IETF extended BGP-4 to Multiprotocol Extensions for BGP-4 (MP-BGP). The current MP-BGP standard is RFC 4760.

As an enhancement of BGP-4, MP-BGP provides routing information for various protocols, such as IPv6 (BGP4+) and multicast

Additional features of BGP

MP-BGP and Address Families

Multiprotocol Reachable NLRI (MP_REACH_NLRI) is used to advertise reachable routes and information about the next hop. MP_REACH_NLRI is coded as one or more 3-tuples of the form <Address Family Information, Next Hop Information, Network Layer Reachability Information>.



Address Family Information: consists of a 2-byte Address Family Identifier (AFI) and a 1-byte Subsequent Address Family Identifier (SAFI):

The AFI identifies a network layer protocol. Defined values for this field are specified in RFC 1700 (Address Family Number). For example, 1 indicates IPv4; 2 indicates IPv6.

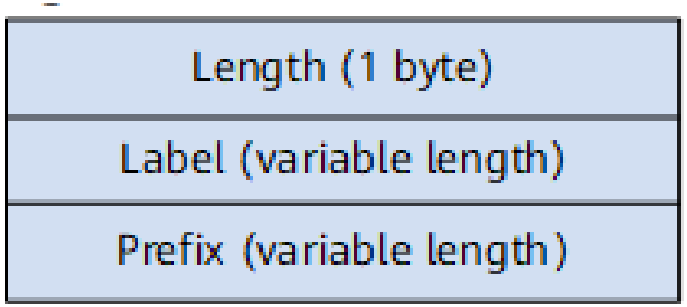
The SAFI indicates the type of the NLRI field.

If the AFI is 1 and the SAFI is 128, the address in the NLRI field is a BGP-VPNv4 address.

Additional features of BGP

MP-BGP and Address Families

Network Layer Reachable Information: is a variable-length field that lists NLRI for the routes being advertised.



Descriptions of each part of the NLRI field are as follows:

Length: indicates the total bits of the label and prefix.

Label: consists of one or more labels. The length of a label is 3 bytes.

Prefix: In a BGP/Multiprotocol Label Switching (MPLS) IP VPN, the prefix field consists of a route distinguisher (RD) and IPv4 address prefix.

VPNv4 update messages exchanged between provider edges (PEs) or autonomous system boundary routers (ASBRs) carry MP_REACH_NLRI. An Update message can carry multiple reachable routes with the same routing attributes.

Additional features of BGP

BGP-LS

BGP-link state (LS) enables BGP to report topology information collected by IGPs to the controller.

Without BGP-LS, the router uses an IGP (OSPF or IS-IS) to collect network topology. This method has the following disadvantages:

- *The controller must have high computing capabilities and support the IGP.*
- *The controller cannot gain the complete inter-domain topology information and therefore is unable to calculate optimal E2E paths.*
- *Different IGPs report topology information separately to the controller, which complicates the controller's analysis and processing.*

BGP-LS has the following advantages:

- *Reduces computing capability requirements and spares the necessity of IGPs on the controller.*
- *Facilitates route selection and calculation on the controller by using BGP to summarize process or AS topology information and report the complete information to the controller.*
- *Requires only one routing protocol (BGP) to report topology information to the controller.*