

Cursul #6

Border Gateway Protocol



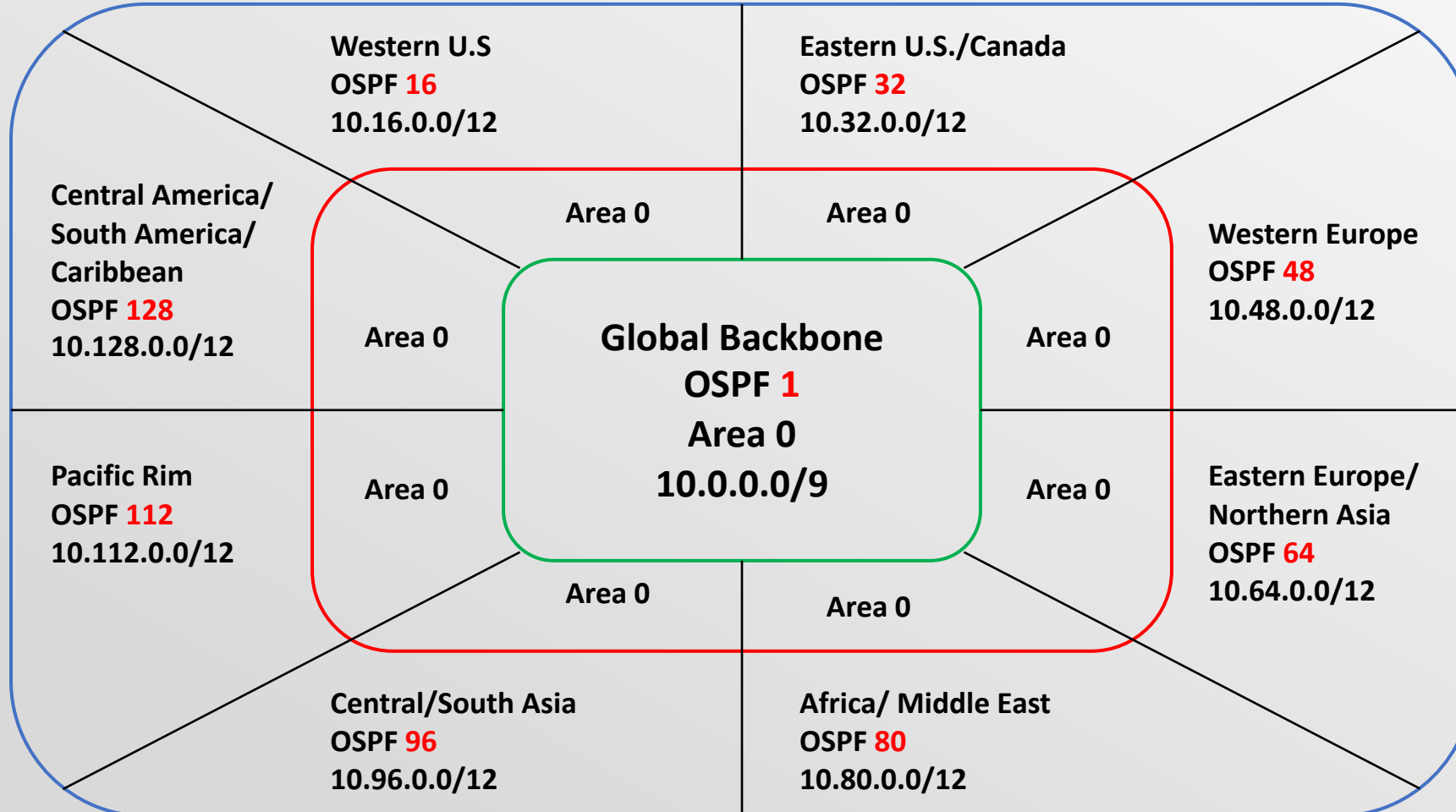
Cuprins

- Who needs BGP?
- Autonomous System
 - Alegerea unui ISP
- Concepte generale BGP
- Tabela de vecini
 - iBGP și eBGP
- Tabela BGP
 - Construirea pachetelor de actualizare
- Tabela de rutare
 - Procesul de selecție

Why BGP?



OSPF multi-area – the perfect solution



OSPF multi-area – the perfect solution

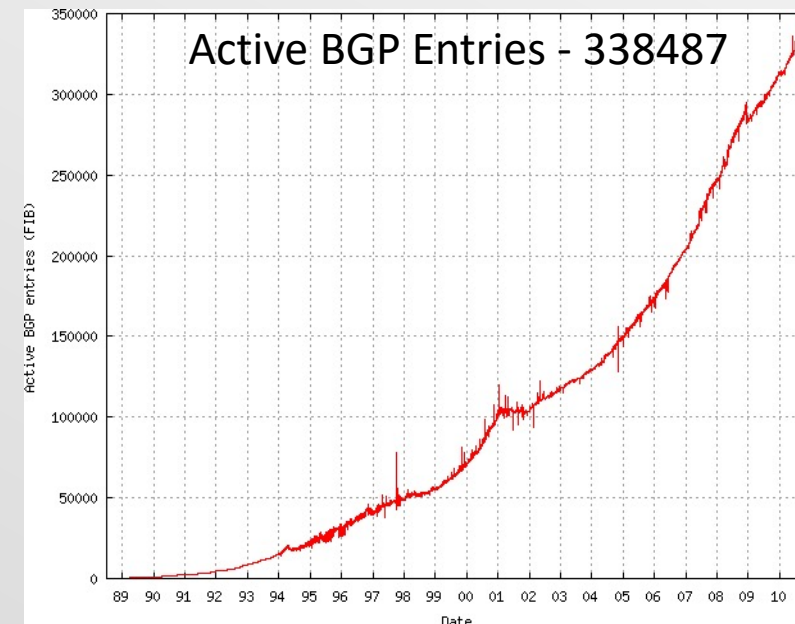
- Fiecare continent rulează un proces OSPF separat
 - O țară = o arie OSPF
 - Agregarea se face la nivel de arie 0 pentru fiecare țară
 - Fiecare continent își poate defini politicile de agregare
- Agregarea la nivel mondial se face în procese diferite
 - Fiecare continent trimite doar o rețea agregată, redistribuită în alt proces OSPF
- Necesități pentru o astfel de abordare
 - Construcție ierarhică a nivelului fizic
 - Înțelegerea perfectă între diverse țări
 - Distribuirea perfectă a spațiului de adrese

Welcome to the real world

- Fiecare țară dorește un set de politici mai complexe
 - OSPF nu poate furniza politici de rutare complexe în interiorul aceleiași proces/arie
- De foarte puține ori o rețea este construită “from the ground up”
 - De obicei, fiecare alege un set de vecini cu care să comunice
- De foarte multe ori în interiorul aceleiași zone fizice există mai mulți provideri de Internet
 - Fiecare provider trebuie identificat unic în Internet

Welcome to the real world

- Peste 300K rețele publice
 - Peste 99.000 de rețele /24.
 - IS-IS poate suporta aproximativ 30K
 - OSPF poate suporta aproximativ 7K
- www.cidr-report.org



Cum se alege un ISP?

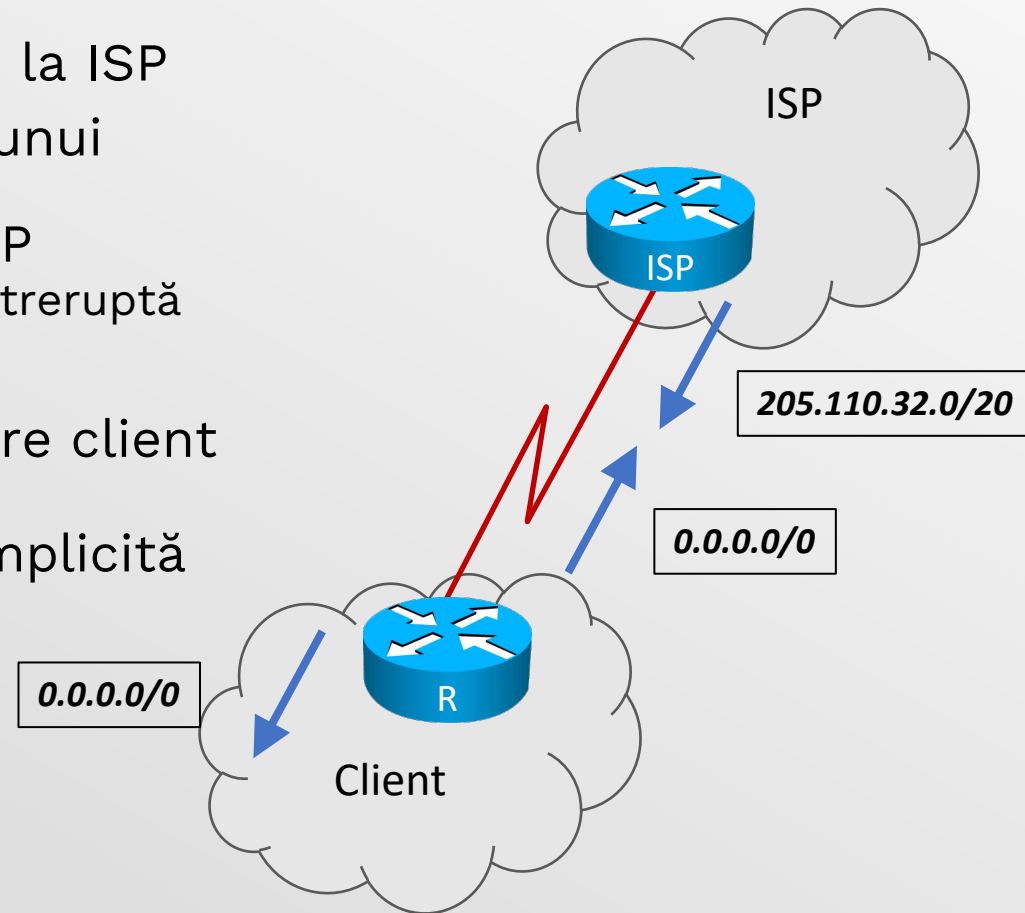


Autonomous System

- Fiecare ISP (Internet Service Provider) este unic
 - AS: număr care identifică unic în Internet un ISP
 - Spațiul de adrese folosit este închiriat de către ISP de la IANA
 - Implementează politici similare, protocoale IGP similare
- Fiecare AS reprezintă un număr între 1 și 65535
 - Intervalul 64512 – 65535 este folosit pentru AS privat
 - Numerele 0 și 65535 sunt rezervate
 - A fost realizată modificarea dimensiunii la 32 de biți (1 ian 2010)
- Informații despre AS-uri
 - www.ripe.net (pentru Europa)
 - AS2614 – RoEduNet
 - AS8708 – RDSNet
 - AS34566 – UPC Romania
 - AS9050 – Romtelecom

Single Homed

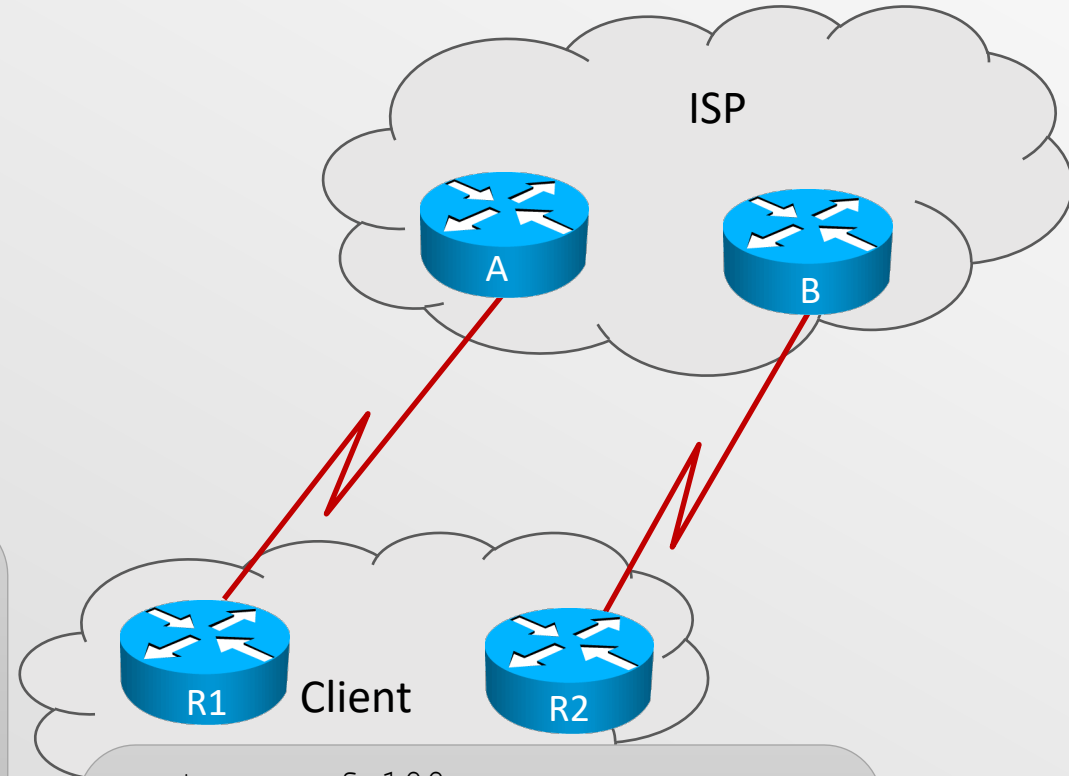
- Există o singură conexiune la ISP
- Nu este necesară rularea unui protocol de rutare între Client și ISP
 - dacă legătura fizică este întreruptă
 - ...
- ISP-ul folosește o rută către client
- Clientul folosește o rută implicită
 - Propagă o rută implicită în rețeaua internă



Multihoming to a single AS

- Două conexiuni fizice la același ISP
- Linie principală / backup
 - Linie secundară cu viteză redusă
- Load-balancing
 - Dacă liniile au aceeași viteză

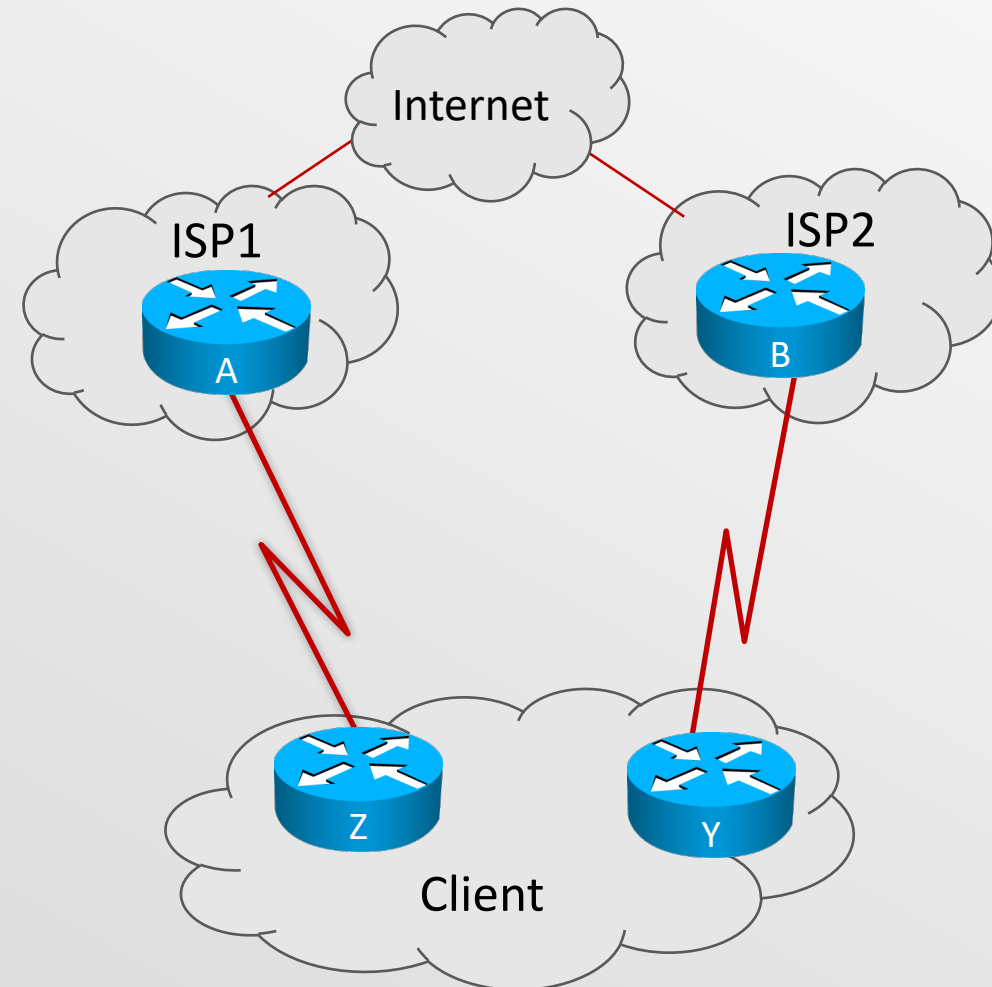
```
router ospf 100
 network 205.110.32.0 0.0.15.255
 area 0
 default-information originate
 metric 10
 ip route 0.0.0.0 0.0.0.0
 205.110.169.108
```



```
router ospf 100
 network 205.110.32.0 0.0.15.255
 area 0
 default-information originate
 metric 100
 ip route 0.0.0.0 0.0.0.0
 205.110.168.108
```

Multihoming to multiple AS

- Un ISP pierde conectivitatea la Internet
- Se pot folosi doar rute Statice din interiorul ISP-ului?
 - Care este spațiul de adrese?
 - Cine îl “anunță” în Internet?
- Pentru rețeaua clientului diferențele nu sunt majore.



Cum alegem un ISP?

- Aproximativ 56K disponibili world-wide (2010)
- Tabela de rutare BGP este publică și accesibilă
 - Google search: “bgp looking glass”
- Număr de adiacențe disponibile
- Exemplu:



Query:	Router:
<input type="radio"/> show interface <interface> <input type="radio"/> show interface status <input type="radio"/> show ip route <prefix> [netmask] <input checked="" type="radio"/> show ip bgp summary <input type="radio"/> show ip bgp neighbor <IP_addr> <input type="radio"/> show ip bgp <prefix> [netmask] <input type="radio"/> ping <IP_addr> <input type="radio"/> traceroute <IP_addr> <input type="radio"/> show environment status	<div> <div>buc-acc1</div> <div>buc-core1</div> <div>buc-peers1</div> <div>buc-rds1</div> <div>clu-acc1</div> <div>clu-core1</div> <div>cra-acc1</div> <div>cra-core1</div> <div>gal-acc1</div> <div>gal-core1</div> <div>ias-acc1</div> <div>ias-acc2</div> <div>ias-core1</div> <div>nat-br1</div> <div>nat-core2</div> <div>tgm-core1</div> <div>tim-acc1</div> <div>tim-core1</div> </div>
Argument(s): <input type="text"/>	

Back to the real world

- IGP – Interior Gateway Protocols
 - Protocoale de rutare folosite în interiorul aceluiași AS
 - Exemple: RIP, EIGRP, OSPF, IS-IS, EIGRP

- EGP
 - Protocoale de rutare folosite pentru transmiterea informațiilor de rutare între AS-uri

- Există o singură opțiune ...

Funcționare BGP



Folosire BGP

- Clientul are o singură conexiune către exterior.



- Resurse prea limitate.
- Nu există o cunoaștere bună a mecanismelor BGP.

- Conexiuni cu mai multe AS-uri.



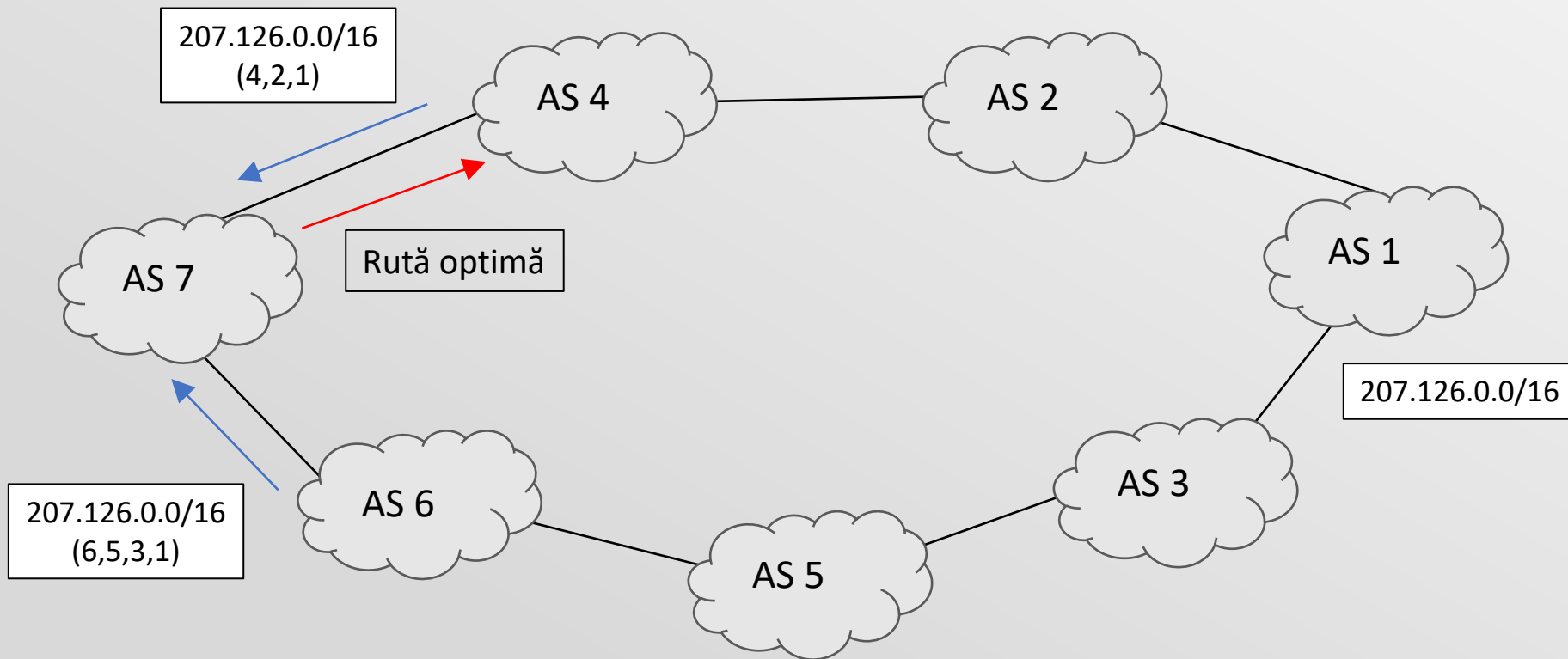
- O singură conexiune, dar politici diferite pentru diverse destinații.
- AS-ul funcționează ca un AS de tranzit.

Border Gateway Protocol

- Standardizat de IETF (RFC4271 – versiunea 4)
 - Protocol open-standard de tip path-vector cu numeroase implementări proprietare
 - Singurul protocol EGP implementat în Internet
- Adiacențele sunt realizate prin conexiuni TCP (port 179)
 - Un ruter este numit BGP Speaker
 - Relația de adiacență se numește peering
- Oferă suport pentru: VLSM, CIDR, agregare

Border Gateway Protocol

- BGP „descrie” calea spre o rețea ca un șir de AS-uri
 - BGP poate fi considerat un protocol de tip „path-vector”
- Un criteriu de selecție pentru ruta optimă este numărul de hop-uri
- Bucle de rutare - verifică existența AS-lui propriu în AS-PATH



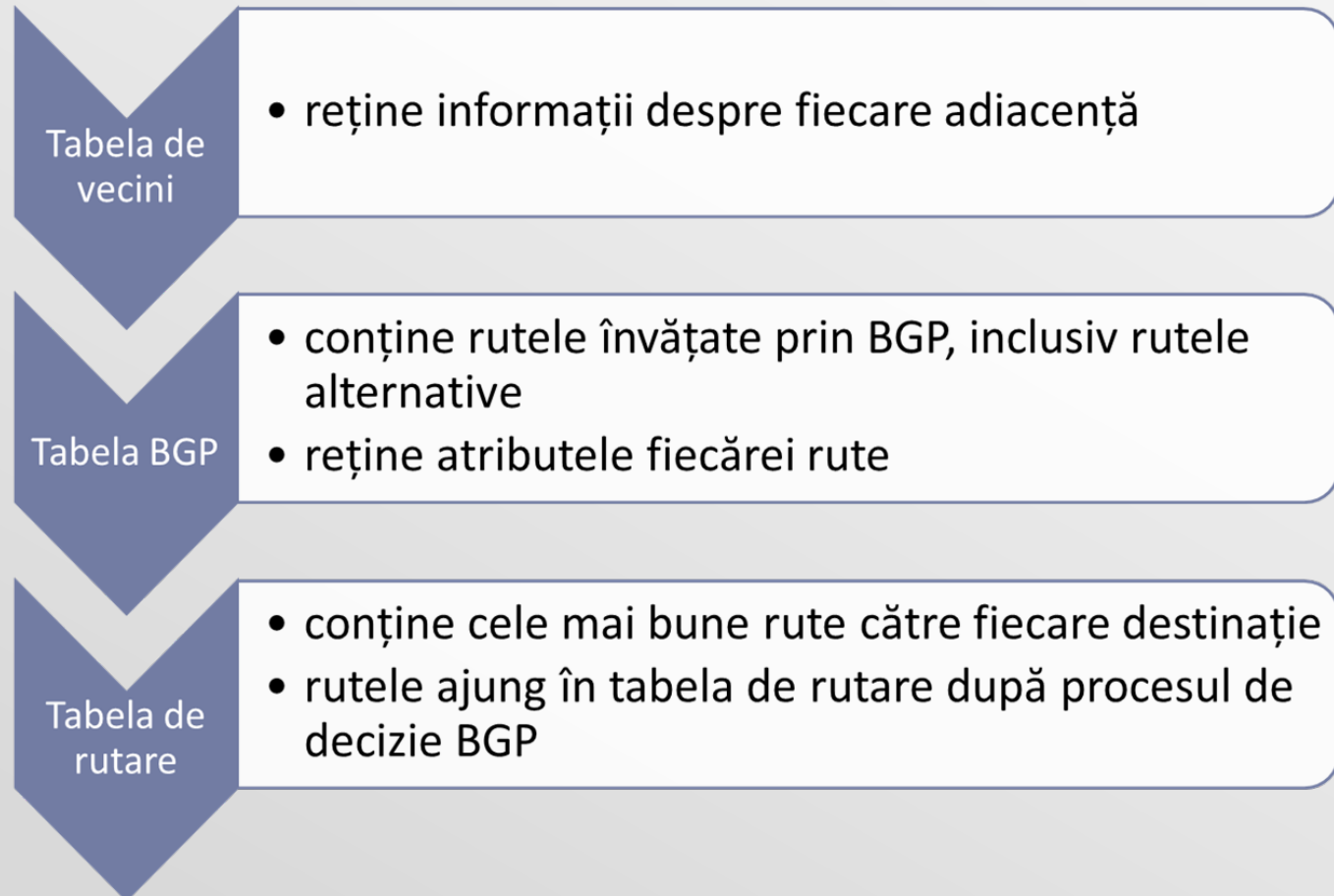
Funcționare BGP

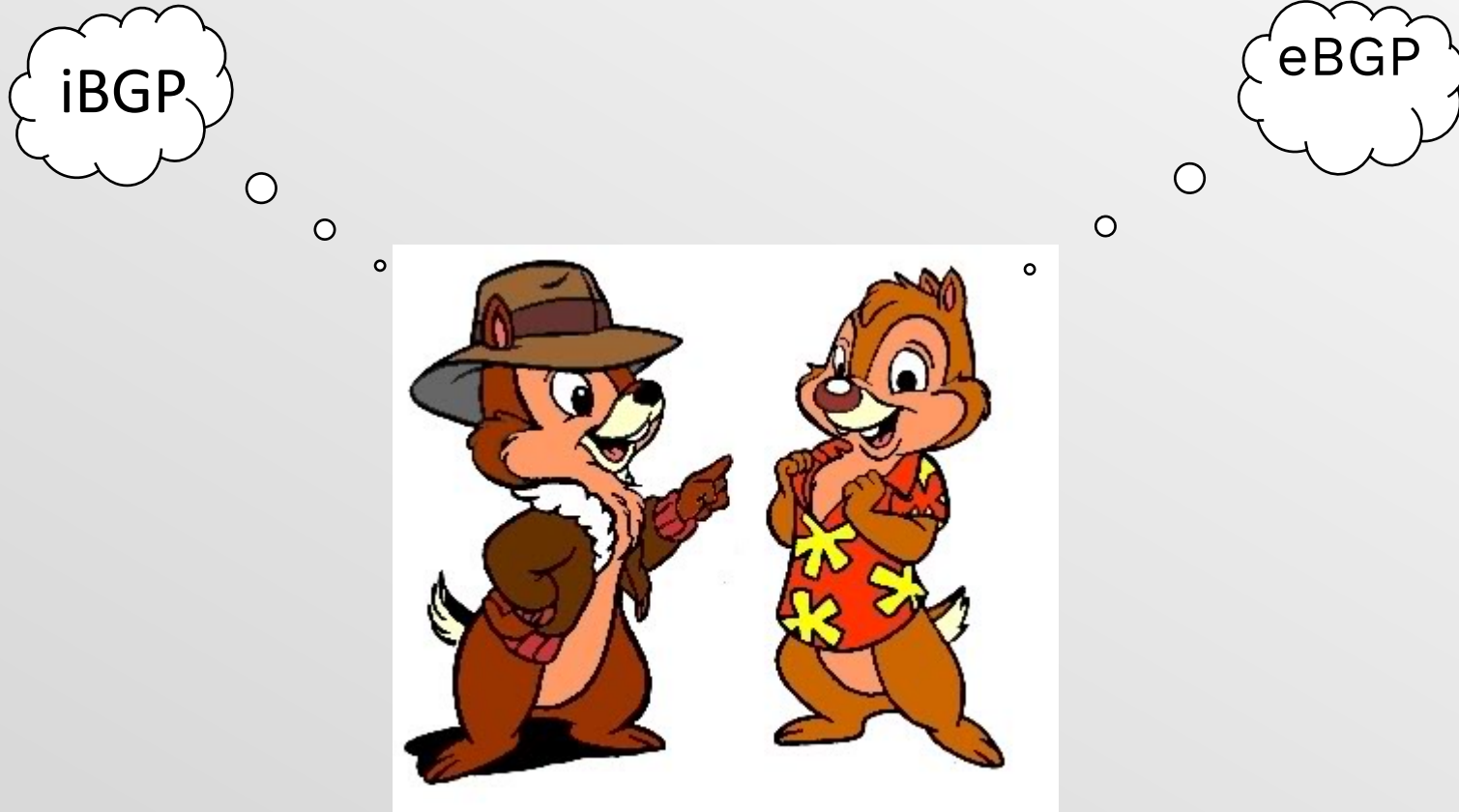
- Stabilirea adiacenței și schimbarea întregii tabele de rutare
 - Ulterior toate pachetele vor fi actualizări parțiale
- Folosirea mesajelor pentru menținerea adiacenței
 - Hello-time 60 secunde/180 secunde hold-time (Cisco)
 - Standardul BGP nu specifică o valoare implicită
- Folosirea mesajelor pentru închiderea conexiunii

Funcționare BGP

- Open
 - Folosit după stabilirea adiacenței pentru identificarea și definirea parametrilor
- Keepalive
 - Mențin sesiunile între vecini
- Update
 - Folosit pentru trimiterea/retragerea de rețele
 - Include și atributele specifice
- Notification
 - Se trimite la detectarea erorilor
 - Conexiunea BGP este imediat închisă după trimitere

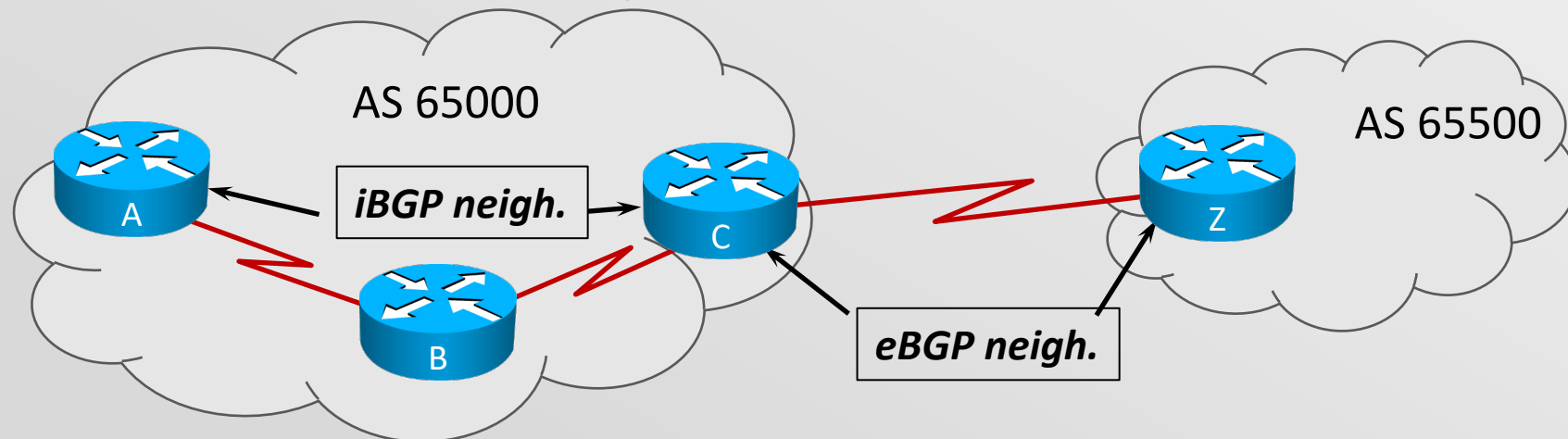
Tabele BGP





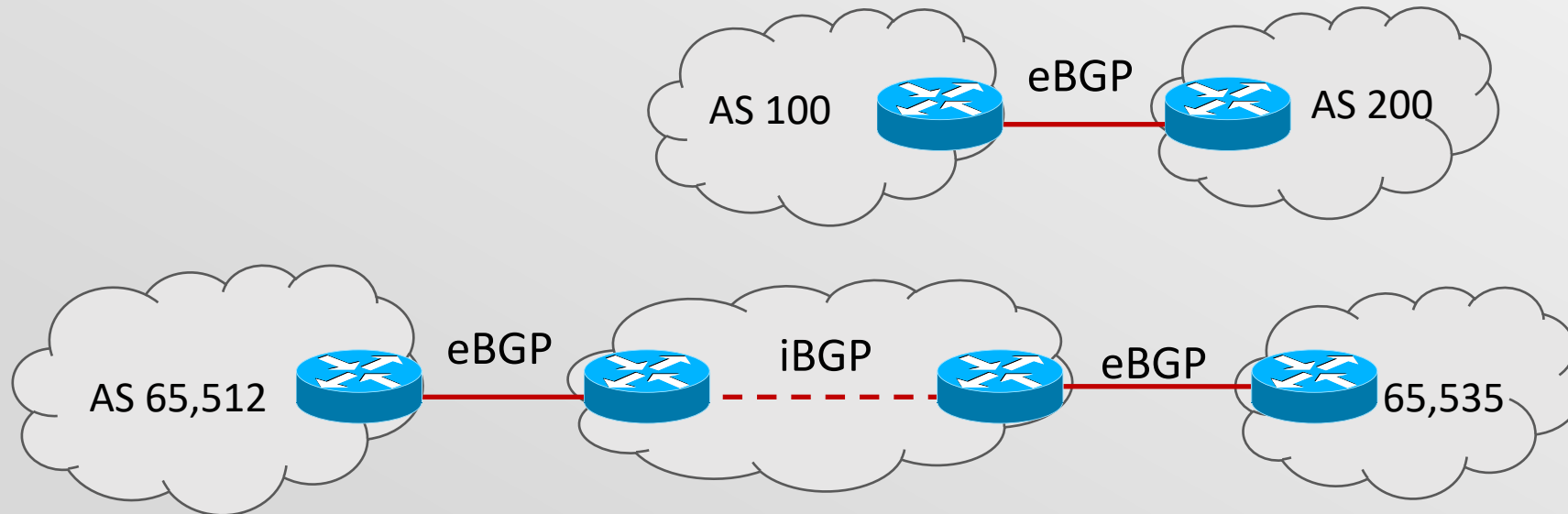
Tipuri de adiacențe

- Adiacențele BGP se formează prin configurarea explicită
- iBGP
 - Vecinii BGP se află în același AS
 - AD-ul rețelelor învățate prin iBGP este 200.
 - Nu e nevoie ca vecinii iBGP să fie direct conectați.
- eBGP
 - Vecinii BGP se află în AS-uri diferite
 - AD-ul rețelelor învățate prin eBGP este 20.
 - Vecinii eBGP sunt, de obicei, direct conectați.



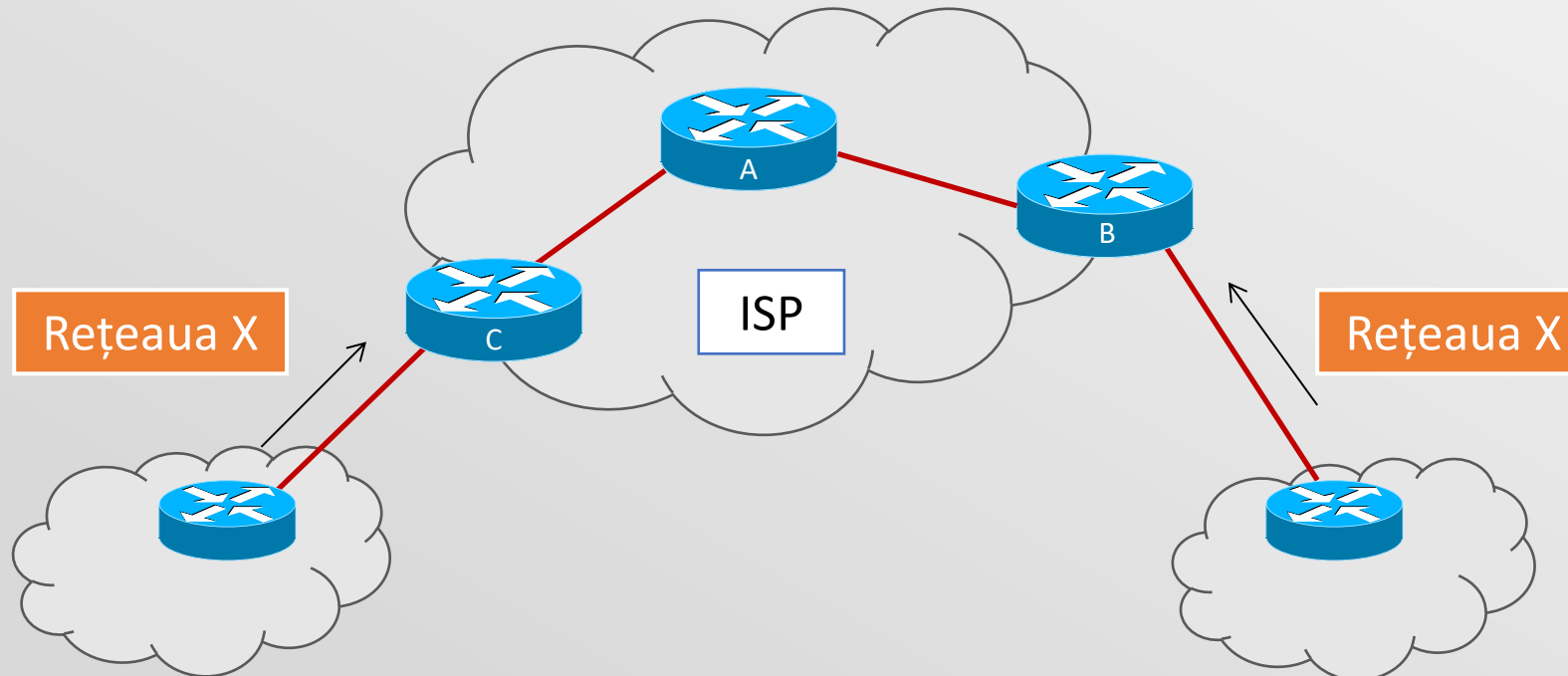
Relații de adiacențe BGP

- 2 rutere BGP (sau BGP-speakers) nu trebuie să fie direct conectate pentru a stabili adiacență
- 2 rutere BGP trebuie să aibă conectivitate de nivel 4(TCP) pentru a putea stabili adiacență
 - rute statice
 - IGP

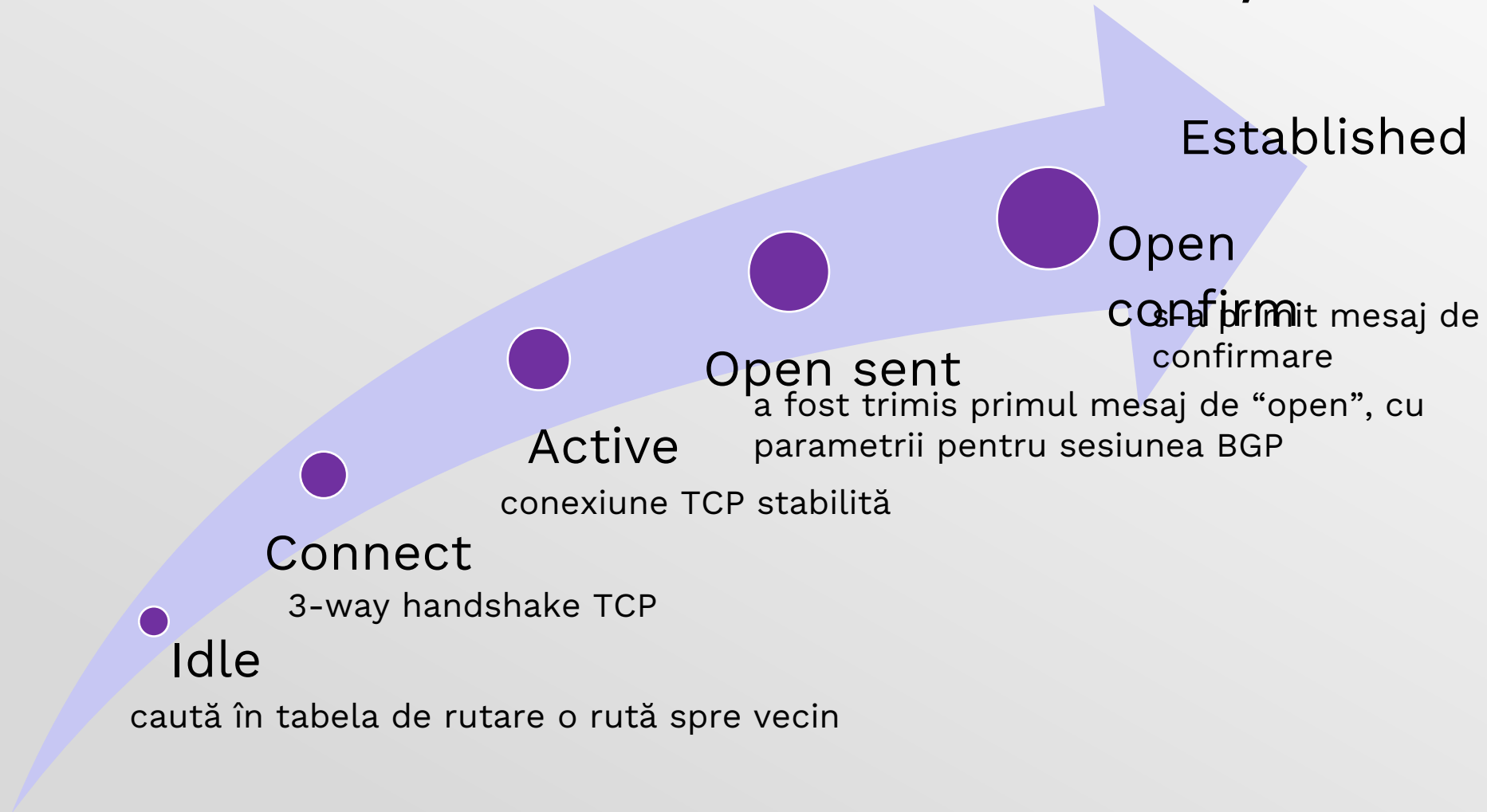


De ce eBGP? De ce iBGP?

- Principala funcționalitate eBGP:
 - Transmiterea rutelor de la un AS la altul
- Motivații pentru utilizarea iBGP:
 - Asigurarea consecvenței politicilor și rutelor BGP în cadrul unui AS
 - Necesari într-un AS de tranzit (ISP) pentru a nu crea un black-hole



Procesul de stabilire a adiacențelor

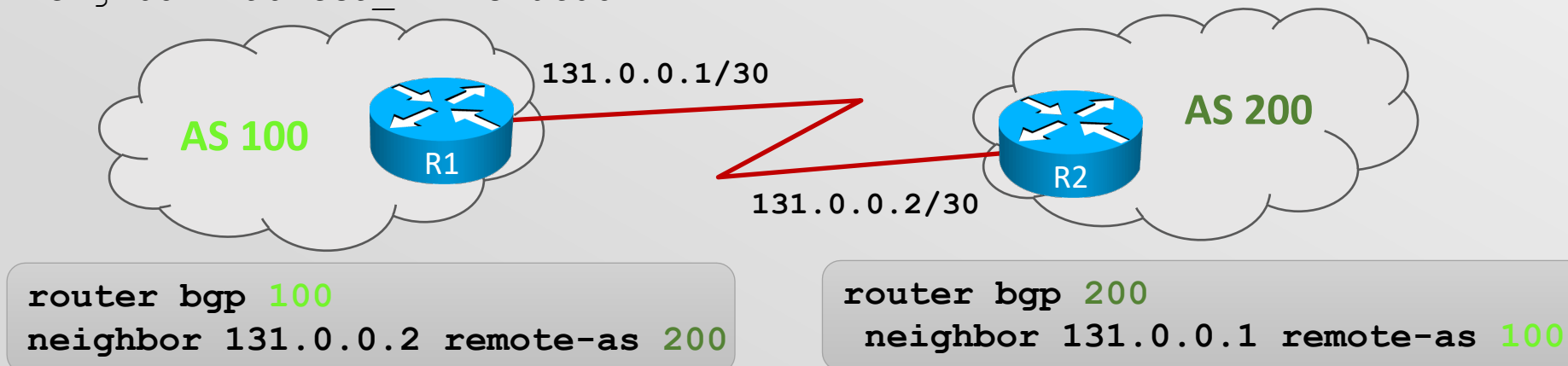


Definirea vecinilor

- Un ruter poate face parte dintr-un singur sistem autonom
 - se poate rula o singură instanță de BGP

```
neighbor <adresa_IP> remote-as <AS>
```

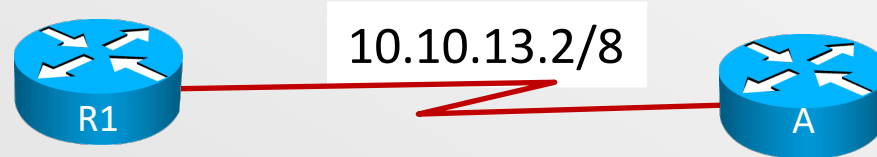
- AS al instanței de BGP de pe vecin
- Un vecin poate fi dezactivat temporar
 - neighbor <adresa_IP> shutdown
 - no neighbor <adresa_IP> shutdown



Reguli pentru stabilirea adiacențelor

- Trebuie ca un ruter să primească o cerere TCP cu adresa sursă ce există în comanda **neighbor**.
- Numărul de AS primit trebuie să corespundă cu numărul configurat cu **neighbor remote-as**.
- **RID-ul** celor două routere nu trebuie să fie egale.
 - RID = Router ID, același proces de alegere ca la OSPF
- Autentificarea trebuie configurată corespunzător.

Verificarea stării de adiacență



R1#sh ip bgp summary

BGP router identifier **10.10.13.1**, local AS number **100**

BGP table version is 1, main routing table version 1

Neighbor	V	AS	MsgRcvd	MsgSent	TblVer	InQ	OutQ	Up/Down	State/PfxRcd
10.10.13.2	4	200	2	2	0	0	0	00:00:05	0

R1#sh ip bgp neighbors 10.10.13.2

BGP neighbor is **10.10.13.2**, **remote AS 200**, external link

BGP version 4, remote router ID 10.10.13.2

BGP state = **Established**, up for 00:00:11

Last read 00:00:11, last write 00:00:11, hold time is 180, keepalive interval is 60 seconds

Neighbor capabilities:

Route refresh: advertised and received(new)

New ASN Capability: advertised and received

Address family IPv4 Unicast: advertised and received

Tabela BGP



Tabela BGP

- Mai este cunoscută sub denumirile: *topology table* sau *BGP Routing Information Base (RIB)*
- Deține NLRI-urile învățate prin BGP și PA-urile asociate
 - Network Layer Reachability Information
 - IP și mască de rețea
 - Denumirile uzuale: rute BGP sau prefixe BGP
 - Path Attributes - lista de attribute
- Tabela BGP conține informații din sursele:
 - Anunțate local prin comanda `network`
 - Rețele învățate de la alți vecini BGP
 - Rețele redistribuite local prin comanda `redistribute`

Comanda `network`

- Alt comportament față de protocoalele IGP
- Specifică rețelele locale ce vor fi propagate în BGP
 - Direct conectate, Statice - definite manual
 - Învățate printr-un protocol IGP (OSPF, EIGRP, ISIS, RIP)
- NU specifică interfețele pe care se trimit pachete pentru stabilirea adiacențelor
- Dacă nu se folosește parametrul `mask`, protocolul va considera masca implicită pentru clasa rețelei
 - Rețeaua trebuie să existe în tabela de rutare (cu masca folosită în comandă)

Comanda network

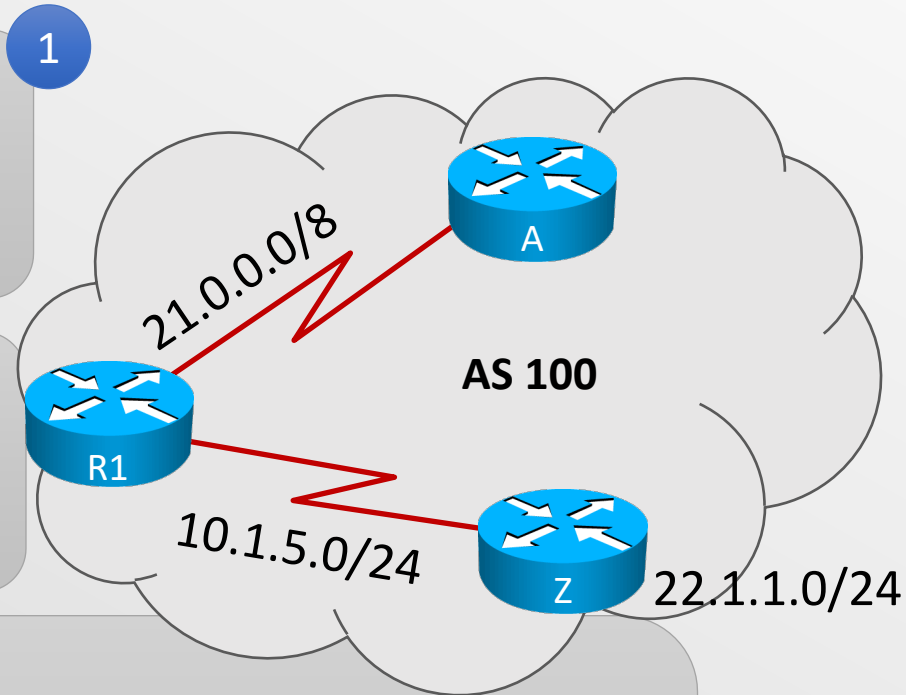
R1#show ip route | inc 21 | 22
C 21.0.0.0/8 is directly connected,
Serial 0/0/1
S 22.0.0.0/24 is subnetted, 1 subnets
S 22.1.1.0 [1/0] via 10.1.5.9

2

```
router bgp 100
network 21.0.0.0
network 22.1.1.0 mask 255.255.255.0
```

R1#show ip bgp
BGP table version is 38, local router ID is 5.5.5.5
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
*> 21.0.0.0	0.0.0.0	0		32768	i
*> 22.1.1.0/24	10.1.5.9	0		32768	i



3

Trimiterea de actualizări

- Fiecare pachet de actualizare BGP conține o listă de attribute și o listă de prefixe
 - Dacă se dorește trimiterea a două prefixe cu cel puțin un atribut diferit se vor construi două pachete de actualizare
- Pachetele de actualizare pot conține și rute ce trebuie retrase
- Se trimit doar rețelele considerate cele mai bune
 - În funcție de attributele fiecărei rețele
 - Aceste rețele vor apărea și în tabela de rutare

Clase de attribute

Well-known mandatory

Prezent în mesaje
Recunoscut de orice implementare

Origin

AS-PATH

NEXT-HOP

Well-known discretionary

Opțional în mesaje
Recunoscut de orice implementare

Local Preference

Atomic Aggregate

Optional transitive

Nu este recunoscut de orice implementare
Va fi retrimis

Community

Aggregator

Optional non-transitive

Nu este recunoscut de orice implementare
Nu va fi retrimis

Multi_Exit_Discriminator

ORIGINATOR-ID

Actualizări iBGP vs. eBGP

• iBGP

- nu sunt trimise mai departe în iBGP dacă au fost învățate prin iBGP
- nu se modifică atributul NEXT-HOP

▶ eBGP

- setează implicit valoarea câmpului TTL la 1
- modifică atributul NEXT-HOP

Impactul atributului NEXT_HOP

- Atributul NEXT_HOP este folosit pentru identificarea echipamentului către care trebuie trimis pachetul.

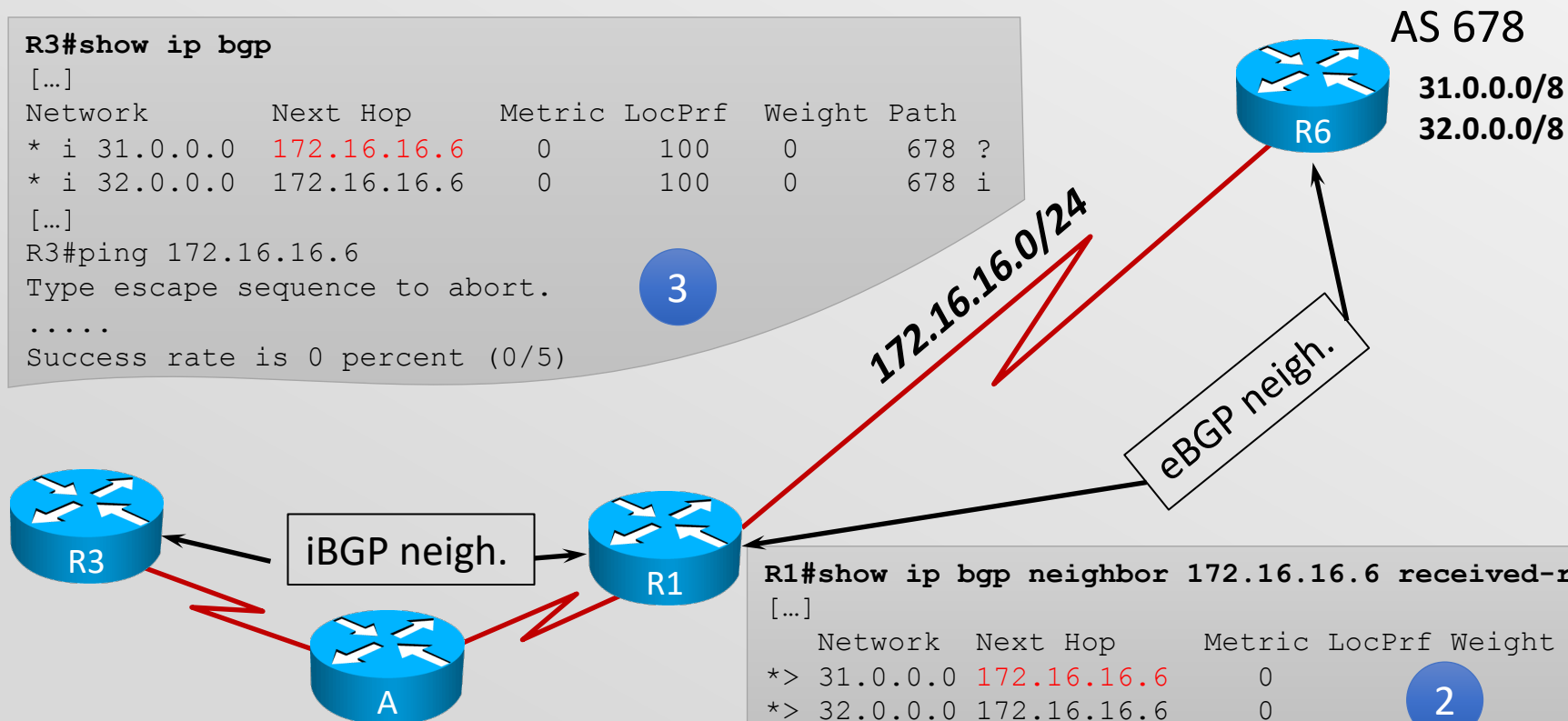
```
R6# show ip bgp neighbors 172.16.16.1 advertised-routes
[...]
```

	Network	Next Hop	Metric	LocPrf	Weight	Path
*>	31.0.0.0	0.0.0.0				?
*>	32.0.0.0	0.0.0.0				i

```
R3#show ip bgp
[...]
```

	Network	Next Hop	Metric	LocPrf	Weight	Path
* i	31.0.0.0	172.16.16.6	0	100	0	678 ?
* i	32.0.0.0	172.16.16.6	0	100	0	678 i

```
[...]
R3#ping 172.16.16.6
Type escape sequence to abort.
.....
Success rate is 0 percent (0/5)
```



```
R1#show ip bgp neighbor 172.16.16.6 received-routes
[...]
```

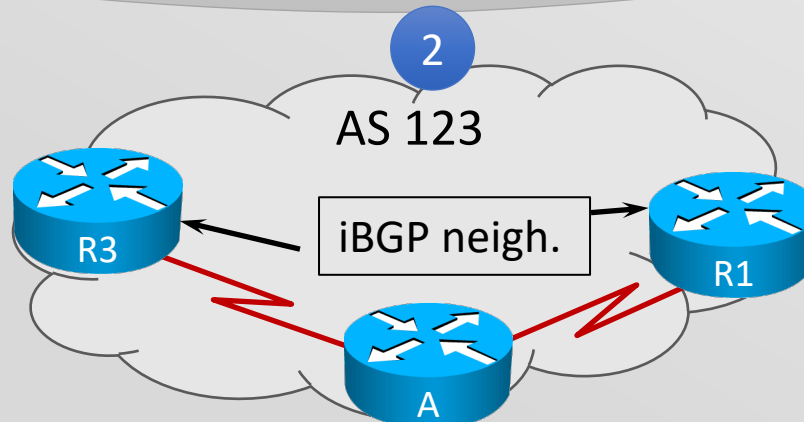
	Network	Next Hop	Metric	LocPrf	Weight	Path
*>	31.0.0.0	172.16.16.6	0			678 ?
*>	32.0.0.0	172.16.16.6	0			678 i

Impactul atributului NEXT_HOP

- Există două soluții:
 - configurarea conectivității cu adresa IP a ruterului eBGP
 - nu este recomandată anunțarea rețelei dintre ISP-uri în cadrul protocolului IGP
 - schimbarea atributului NEXT_HOP
 - folosind comanda `neighbor <adresa_IP> next-hop-self`

```
R3#show ip bgp
[...]
```

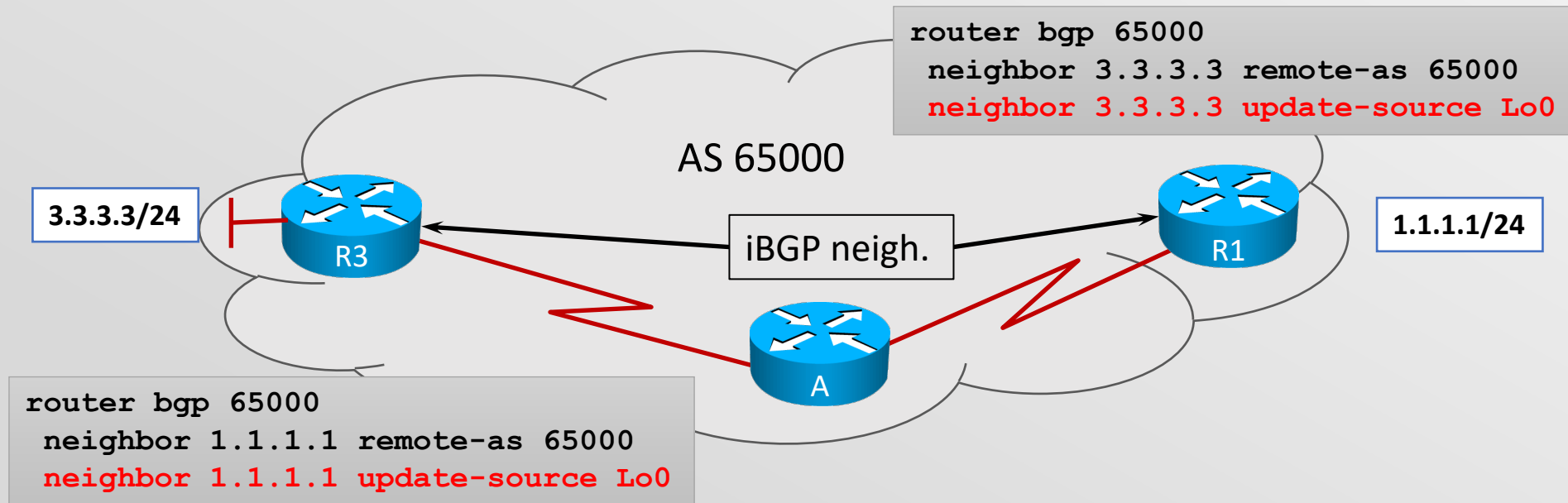
Network	Next Hop	Metric	LocPrf	Weight	Path
*>i 31.0.0.0	1.1.1.1	0	100	0	678 ?
*>i 32.0.0.0	1.1.1.1	0	100	0	678 i



```
R1#conf t
R1(config)# router bgp 123
R1(config-router)#neigh 3.3.3.3 next-hop-self
```

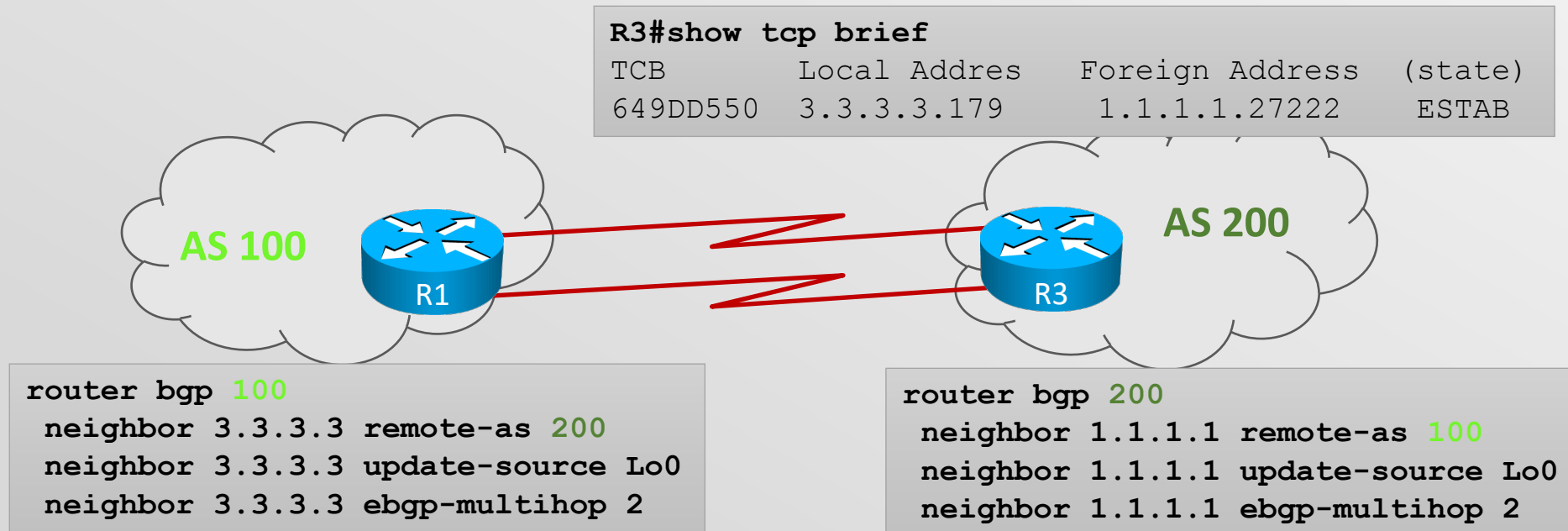
Modificarea sursei pentru actualizare

- BGP folosește implicit adresa IP a interfeței pe care se trimite actualizarea
- O interfață fizică se poate defecta
 - se recomandă folosirea unei interfețe de loopback
 - trebuie modificată și adresa vecinului (adresa destinație)



Modificarea TTL-ului

- Utilă în două cazuri
 - vecinii eBGP nu sunt direct conectați
 - vecinii eBGP folosesc alte interfețe pentru sursa pachetului
- Se poate realiza doar pentru vecini eBGP

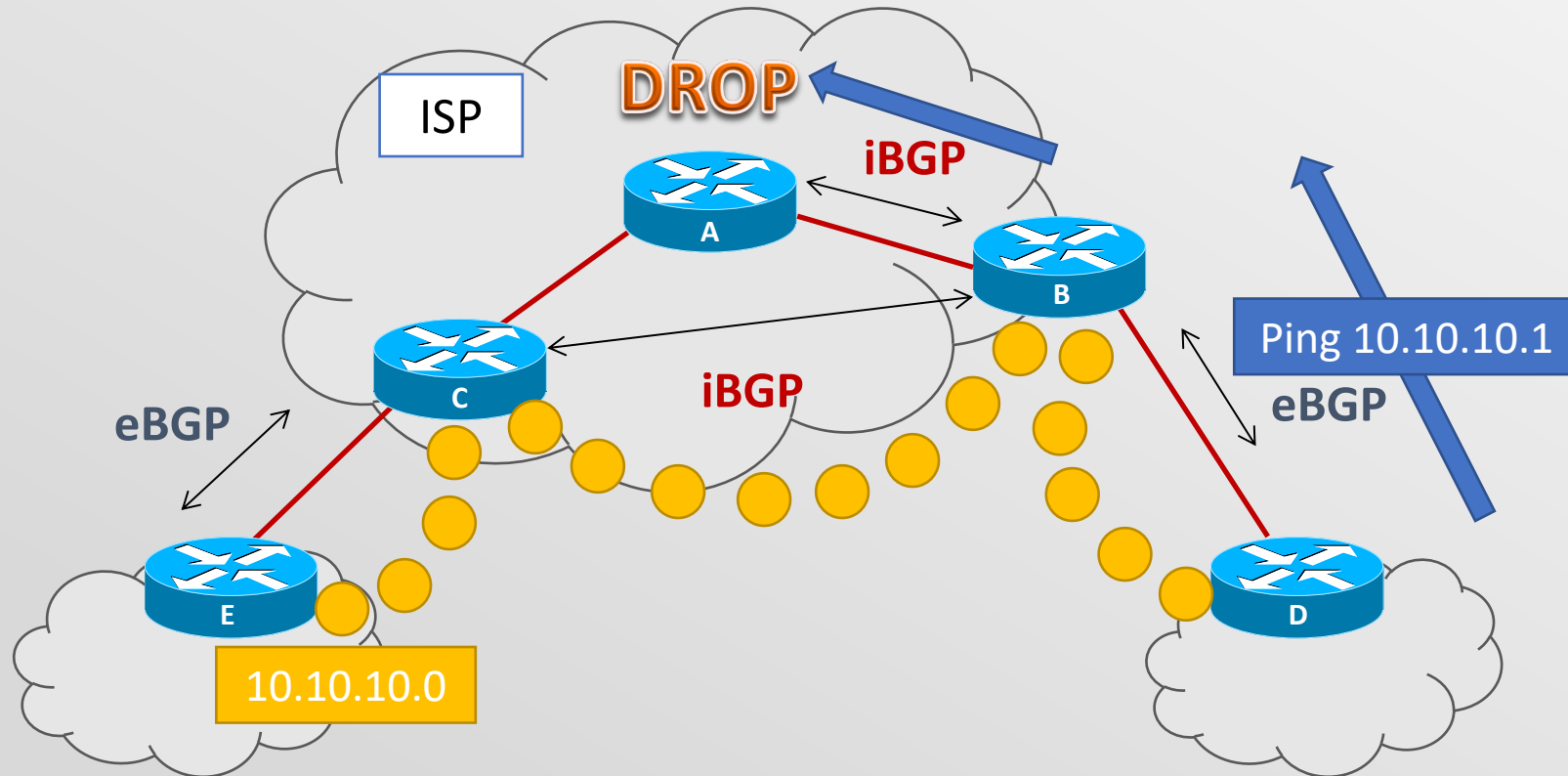


Actualizări iBGP – iBGP

- Ruterle iBGP nu transmit informațiile învățate prin iBGP către alți vecini iBGP
 - Pentru prevenirea buclelor
- De obicei numărul ruterelor ce rulează iBGP este redus
 - Se rulează iBGP doar în core
 - Se poate forma o topologie logică full-mesh
 - Adiacențe iBGP între toate ruterle
 - Tot ce se învață prin eBGP va ajunge pe toate ruterle
- Foarte greu de scalat o astfel de topologie full-mesh

Fenomenul “black-hole”

- Poate apărea în AS-urile de tranzit în momentul în care nu avem full-mesh iBGP



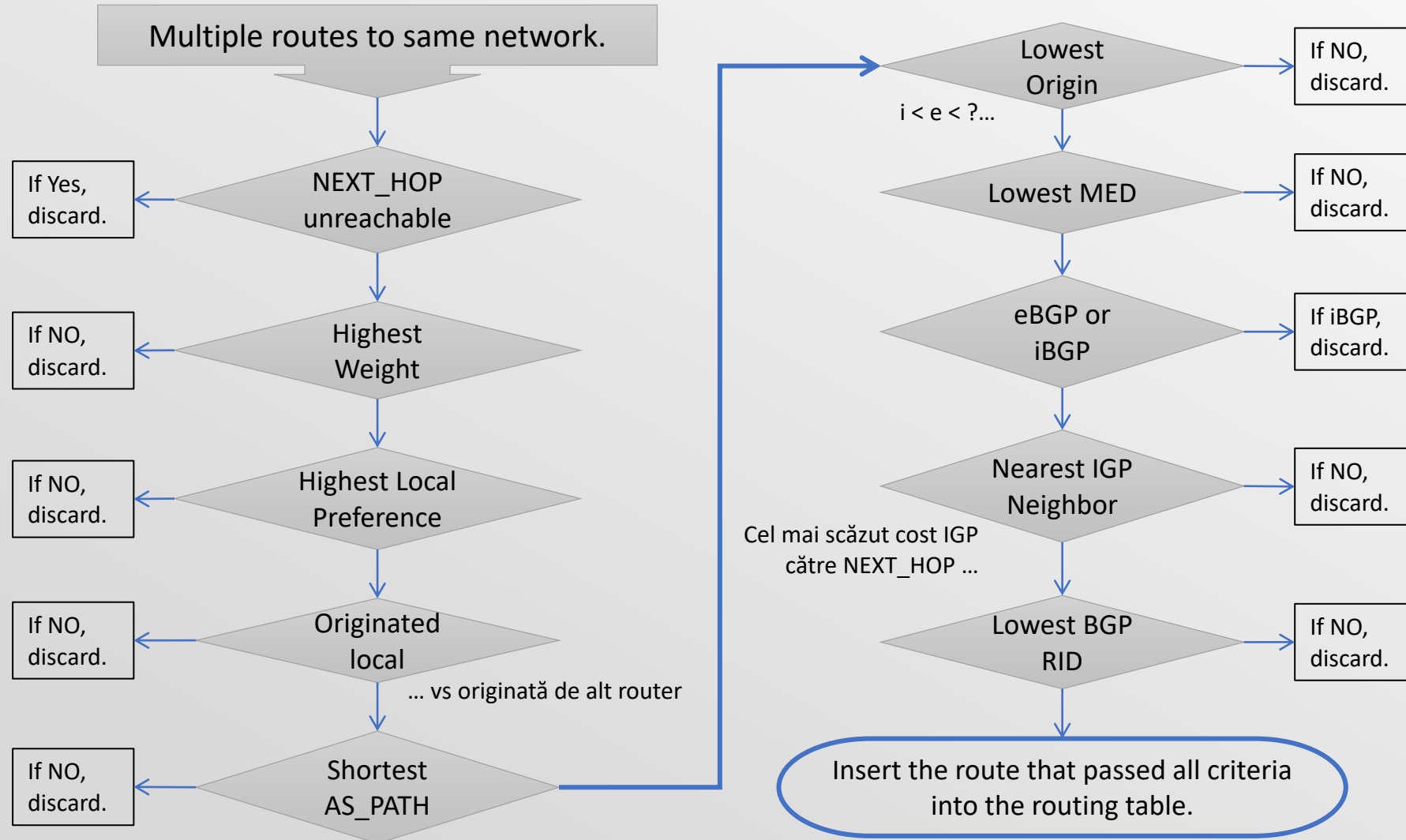
Procesul de selecție



Construirea tabelii de rutare

- BGP aplică un proces de decizie asupra tabelii BGP pentru a extrage rutele cel mai bune
 - Marcate în tabela BGP cu >
- Rutele învățate prin eBGP sunt cele mai bune
 - Nu ar trebui să existe rețelele altui ISP în interiorul rețelei
 - AD este de 20, respectiv 200 pentru iBGP
 - AD-ul se poate modifica asemănător protoalelor IGP
- În tabela de rutare adresa IP next-hop este dată de valoarea atributului NEXT_HOP
 - Se va face recursive lookup la trimiterea pachetelor

Procesul de decizie



Manipularea traficului BGP

- Procesul de decizie BGP este influențat de attributele asociate unui prefix
- Influențarea traficului BGP se poate face prin modificarea atributelor
 - În funcție de ordinea atributului în procesul de decizie
- Nu toate attributele se păstrează atunci când lista de prefixe „traversează” un anumit AS

Vizualizarea modificărilor



Vizualizarea modificărilor

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, ?

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		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	100		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	32000		0	64998 i

WEIGHT

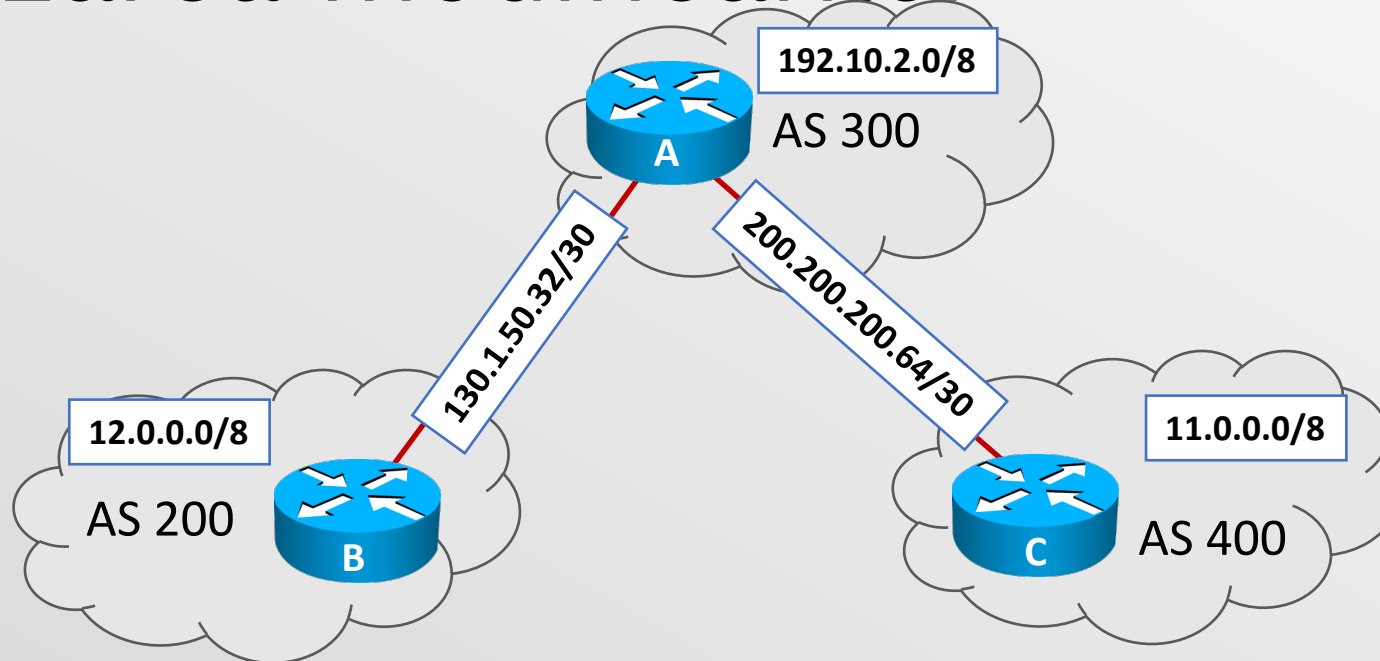
IGP origin

LOCAL_PREF

MED

AS paths

Vizualizarea modificărilor



C#show ip bgp

BGP table version is 8, local router ID is 200.200.200.66

Status codes: s suppressed, d damped, h history, * valid, > best, i - internal

Origin codes: i - IGP, e - EGP, ? - incomplete

Network	Next Hop	Metric	LocPrf	Weight	Path
*> 11.0.0.0	0.0.0.0	0		32768	i
*> 12.0.0.0	200.200.200.65			0	300 200 i
*> 193.10.2.0	200.200.200.65	0		0	300 i

Vizualizarea modificărilor

```
C#show ip bgp
BGP table version is 8, local router ID is 200.200.200.66
Status codes: s suppressed, d damped, h history, * valid, > best, i - internal
Origin codes: i - IGP, e - EGP, ? - incomplete
```

Network	Next Hop	Metric	LocPrf	Weight	Path
*> 11.0.0.0	0.0.0.0	0		32768	i
*> 12.0.0.0	200.200.200.65			0	300 200 i
*> 193.10.2.0	200.200.200.65	0		0	300 i

- **BGP table version** – incrementat de fiecare dată când tabela BGP se schimbă
- **Local router ID** – adresa RID a ruterului
- **Status codes** – Starea intrării din tabelă. Starea este afișată la începutul fiecărei linii:
 - s — intrarea este suspendată (suppressed)
 - * — intrarea este validă
 - > — intrarea este cea mai bună cale pentru rețeaua dată
 - i — intrarea a fost învățată printr-o sesiune iBGP

Vizualizarea modificărilor

```
C#show ip bgp
```

```
BGP table version is 8, local router ID is 200.200.200.66
```

```
Status codes: s suppressed, d damped, h history, * valid, > best, i - internal
```

```
Origin codes: i - IGP, e - EGP, ? - incomplete
```

Network	Next Hop	Metric	LocPrf	Weight	Path
*> 11.0.0.0	0.0.0.0	0		32768	i
*> 12.0.0.0	200.200.200.65			0	300 200 i
*> 193.10.2.0	200.200.200.65	0		0	300 i

- **Origin codes** – Originea intrării. Această informație este plasată la sfârșitul fiecărei intrări:
 - i — Intrarea a fost generată de o sesiune IGP
 - e —Intrarea a fost generată de o sesiune EGP
 - ? — Originea intrării este neclară. Această situație apare, de obicei, când rețeaua a fost redistribuită în BGP.
- **Network** – Spațiul de adrese destinație.
- **Next Hop** – adresa IP a următorului ruter. O adresă 0.0.0.0 semnifică ca ruterul are o rută non-BGP către această rețea

Vizualizarea modificărilor

C#**show ip bgp**

BGP table version is 8, local router ID is 200.200.200.66

Status codes: s suppressed, d damped, h history, * valid, > best, i - internal

Origin codes: i - IGP, e - EGP, ? - incomplete

Network	Next Hop	Metric	LocPrf	Weight	Path
*> 11.0.0.0	0.0.0.0	0		32768	i
*> 12.0.0.0	200.200.200.65			0	300 200 i
*> 193.10.2.0	200.200.200.65	0		0	300 i

- **Metric** – Dacă este precizată, valoarea sa specifică metrica interAS (MED)
- **LocPrf** – Local Preference. Valoarea implicită este 100.
- **Weight** – Importanța unei rute (Cisco proprietary)
- **Path** – Calea (AS_PATH) urmată de respectivul pachet de actualizare.

Clase de attribute

Well-known mandatory

Prezent în mesaje
Recunoscut de orice implementare

Origin

AS-PATH

NEXT-HOP

Well-known discretionary

Opțional în mesaje
Recunoscut de orice implementare

Local Preference

Atomic Aggregate

Optional transitive

Nu este recunoscut de orice implementare
Va fi retrimis către alți vecini BGP

Community

Aggregator

Optional non-transitive

Nu este recunoscut de orice implementare
Nu va fi retrimis

Multi_Exit_Discriminator

ORIGINATOR-ID

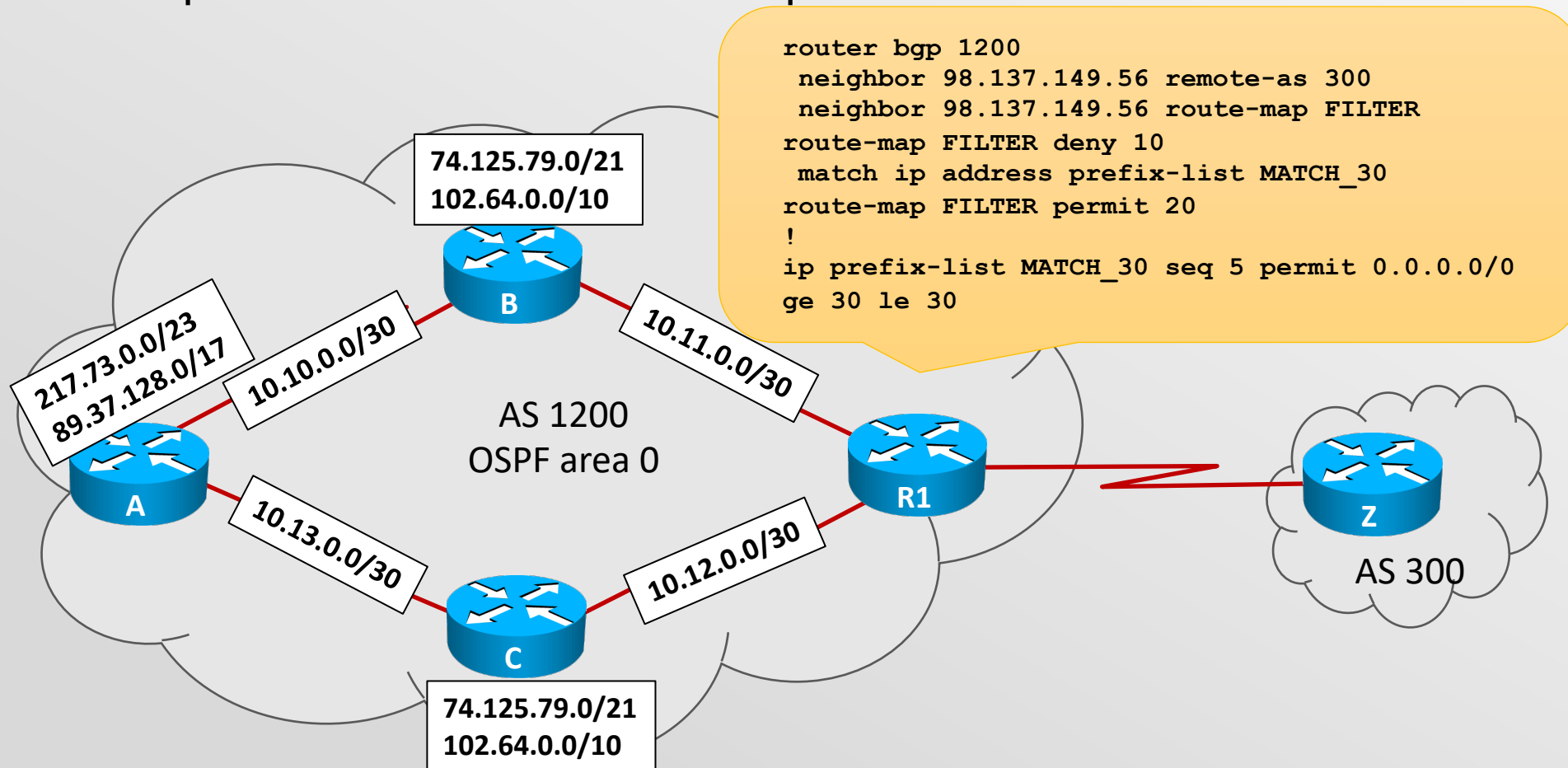
Filtrarea unor rețele

- IP Prefix Lists
 - Furnizează un mecanism de filtrare pe baza a două componente: prefixul și lungimea prefixului
 - *network/length* – toate rutele care au primii *length* biți egali cu cei definiți în *network*
 - lungimea prefixului rutelor poate fi variabilă

Parametru	Lungimea prefixului
Niciunul	conf-length = route-length
Doar “le”	conf-length ≤ route-length ≤ le-value
Doar “ge”	ge-value ≤ route-length ≤ 32
Ambele	ge-value ≤ route-length ≤ le-value

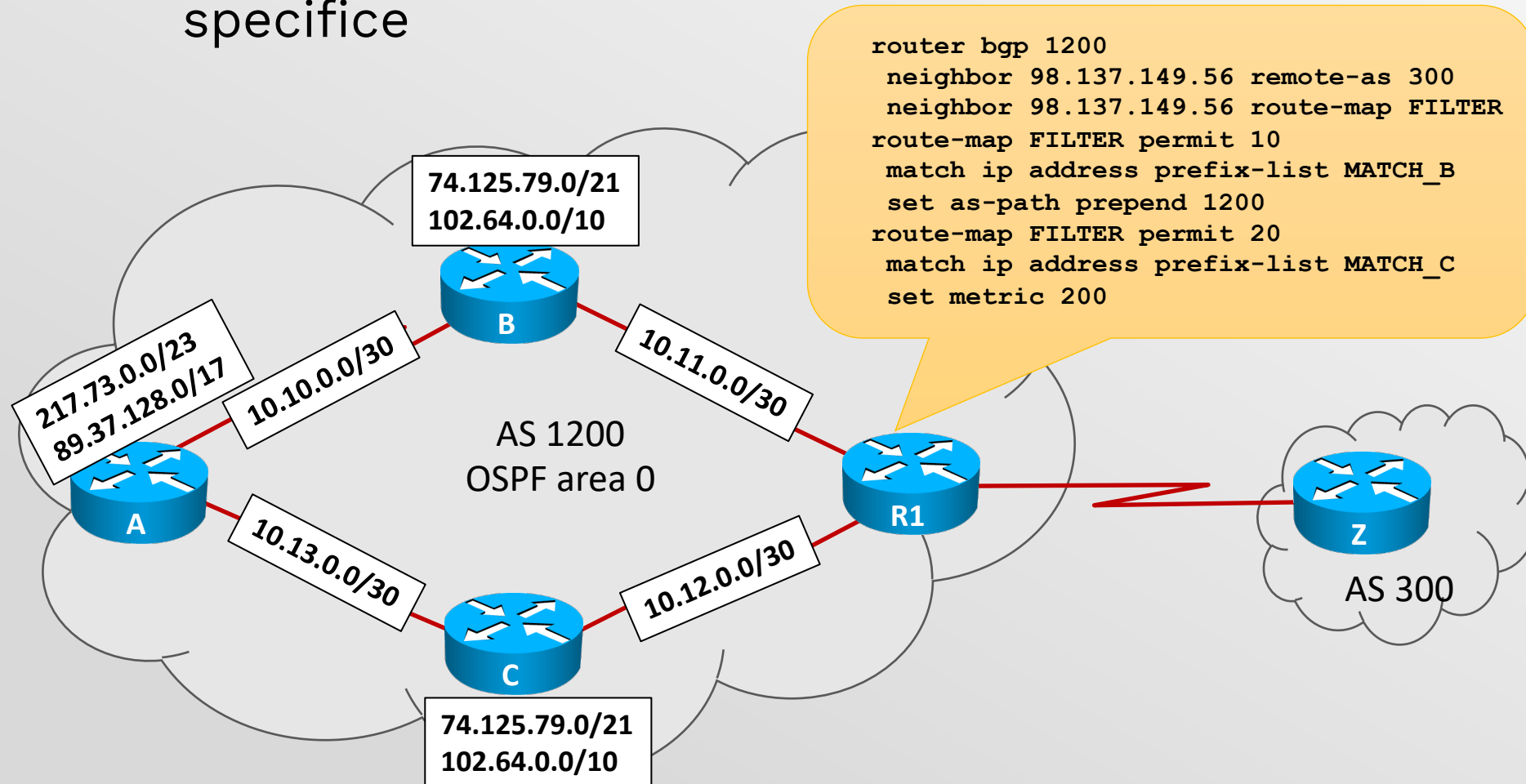
Filtrarea unor rețele

- Redistribuirea tuturor rețelelor, fără cele folosite pentru conectarea echipamentelor



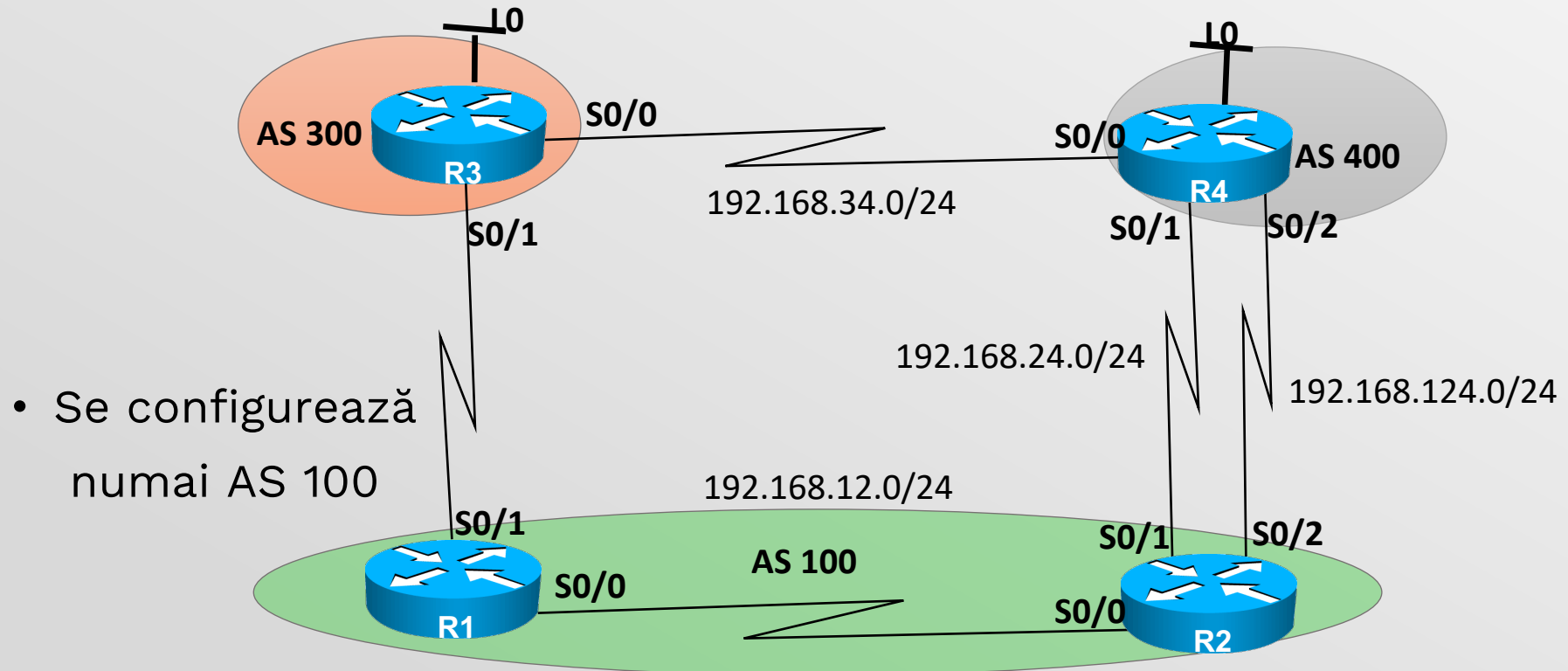
Alterarea specifică de atribute

- Se pot configura politici separate pentru prefixe specifice

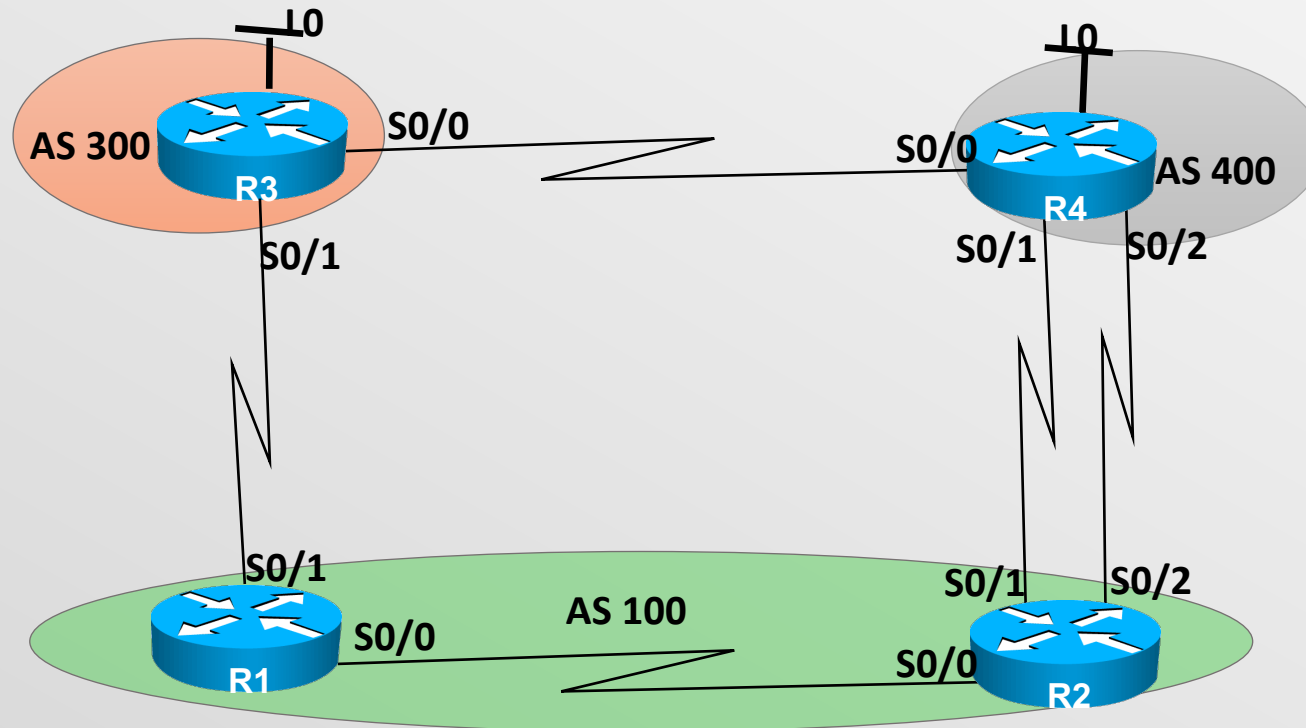


PoC – attribute BGP

- Se știe că pe toate legăturile se formează adiacențe BGP
- Se preferă ca traficul din AS 100 să iasă prin R2, S0/2 și traficul de întoarcere să fie pe aceeași cale



PoC – attribute BGP rezolvare



```

R1 (config)# route-map LOCAL_PREF permit 10
R1 (config-route-map)# set local-preference 50
R1 (config)#route-map AS_PREP permit 10
R1 (config-route-map)#set as-path prepend 100 100
R1 (config)#router bgp 100
R1 (config-router)# neighbor 192.168.13.3 route-map LOCAL_PREF in
R1 (config-router)# neighbor 192.168.13.3 route-map AS_PREP out
    
```

```

R2 (config)# route-map MED permit 10
R2 (config-route-map)# set metric 50
R2 (config)#router bgp 100
R2 (config-router)# neighbor 192.168.24.4
    route-map MED out
R2 (config-router)# neighbor 192.168.124.4
    weight 100
    
```

Sumar

BGP

Modificarea
Atributelor
BGP

Tabela
de
rutare

Tabela
BGP

Tabela
de
vecini

