

BGP

dr Pavle Vuletić

1

Internet – istorija (1)

- ARPANET (1969 – 1989) – 56kbps
 - Samo za akademske, istraživačke i vladine institucije
 - nekomercijalna
- NSFNET (1985 - 1995) – 56kbps, T1, T3
 - Troslojna arhitektura (kičma, regionalne i kampus mreže
 - Za akademske i istraživačke organizacije
 - Vlada je imala svoje nezavisne mreže
 - Početkom 90-ih se javila potreba za povezivanjem sa komercijalnim mrežama, u Evropi i Aziji su razvijene prve mreže zasnovane na istim protokolima

2

Internet – istorija (2)

- Vladine mreže su razmenjivale saobraćaj u Federal Internet eXchange – FIX (na istočnoj i zapadnoj obali SAD)
- Komercijalne mreže su razmenjivale saobraćaj u Commercial Internet eXchange – CIX
- Pojavili su se prvi ISP
- SPRINT je dobio od NSFNET zadatak da poveže NSFNET sa evropskim i azijskim mrežama

3

Današnja organizacija Interneta

- Nema više jedinstvene kičmene mreže
- Skup mreža različitih provajdera i korisnika Interneta – distribuirana arhitektura
- Provajderi razmenjuju saobraćaj u “tačkama za razmenu saobraćaja” - Exchange Point-ima
- Veliki provajderi razmenjuju saobraćaj na više mesta i imaju veliki broj “tačaka za pristup” – Point of Presence – PoP koji služe za povezivanje sa korisnicima

4

Razmena saobraćaja

- NSFNET projekat – Network Access Point - NAP
- U tačkama za razmenu saobraćaja
- Direktnim povezivanjem ISP
- Tačke za razmenu saobraćaja su obično organizovane kao neprofitne organizacije
- Gotovo sve zemlje na svetu imaju svoje tačke za razmenu saobraćaja
- Srbija od 2010 ima tačku za razmenu saobraćaja, ali privatnu, profitnu. Povezano nekoliko ISP. TS, SBB nisu.

5

Routing Arbiter

- RA – projekat finansiran od strane NSF
- U velikim tačkama za razmenu saobraćaja bi svaki provajder sa svakim morao da formira BGP konekciju
- Umesto toga svi razmenjuju rute i ostvaruju BGP konekciju sa RA

6

Inicijative posle NSFNET

- Very high speed Backbone Network Service – vBNS - 1995
 - Istraživačko naučna mreža
- Internet 2/Abilene
- TEN-34
- TEN-155
- GN/GEANT
- GN2/GEANT2
- GN3/GEANT3 – od marta 2009.

7

Organizacija Interneta po slojevima

- Neformalna podela na 3 sloja (tier):
 - Sloj 1 – Mreža (ISP) koja razmenjuje saobraćaj sa svim ostalim mrežama čime ima mogućnost da dođe do svih destinacija na Internetu.
 - Sloj 2 - Mreža (ISP) koja razmenjuje saobraćaj sa nekim mrežama, ali mora da kupi IP tranzit kako bi imala pristup do nekih destinacija na Internetu
 - Sloj 3 – Mreža koja isključivo kupuje tranzit do svih destinacija na Internetu.

8

Provajderi prvog sloja

- Mreža 1. sloja razmenjuje saobraćaj sa svim ostalim mrežama prvog sloja
- Mreže 1. sloja – ima ih 9:
 - [AOL Transit Data Network \(ATDN\)](#) (AS1668)
 - [AT&T](#) (AS7018)
 - [Global Crossing \(GX\)](#) (AS3549)
 - [Level 3](#) (AS3356)
 - [Verizon Business \(UUNET\)](#) (AS701)
 - [NTT Communications / \(Verio\)](#) (AS2914)
 - [Qwest](#) (AS209)
 - [SAVVIS](#) (AS3561)
 - [Sprint Nextel Corporation](#) (AS1239)

9

Provajderi prvog sloja

- Grupa od 9 ISP prvog sloja je vrlo zatvorena
- Postoji “peering agreement” koji postavlja izuzetno visoke kriterijume bilo kojem ISP da se uključi i da postane Tier 1 ISP
- Ukoliko neki Tier 1 ISP prodaje Internet servis po niskim cenama ili ako se ne ponaša u skladu sa dogovorima, moguće je isključivanje iz Tier 1 grupe
- Prestanak razmene saobraćaja između neka dva Tier 1 ISP deli Internet na dva dela

10

“nova” Internet topologija

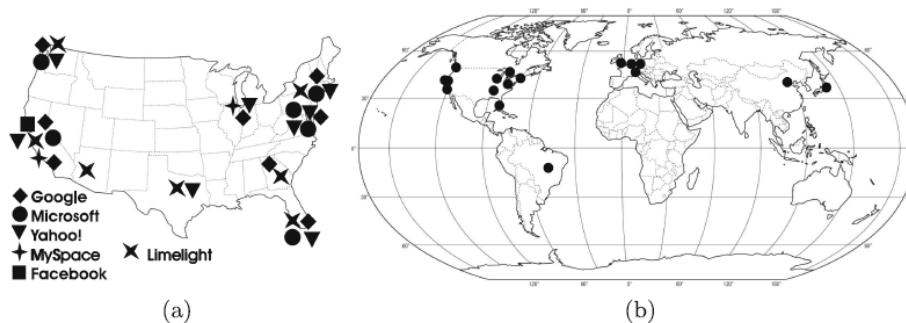


Fig. 2. (a) Location of network end-points in the United States for selected content providers. (b) Our measurement of Google's current WAN.

izvor: Gill P., Arlitt M., Li Z., Mahanti A. "The Flattening Internet Topology: Natural Evolution, Unsightly Barnacles or Contrived Collapse?", Passive and Active Measurement Conference, 2008, Cleveland OH, USA

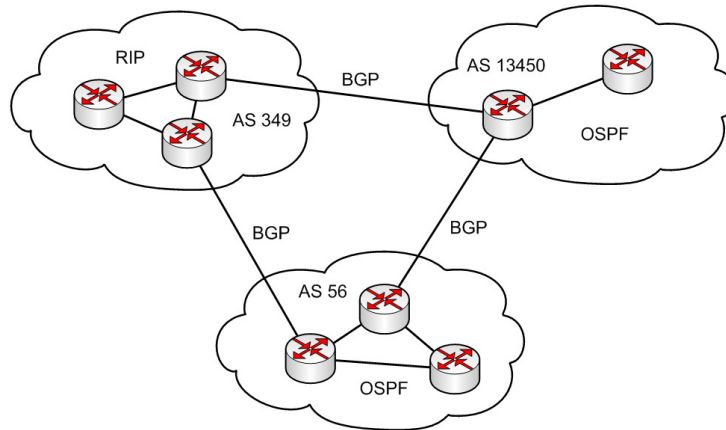
11

BGP

- Služi za razmenu ruta između autonomnih sistema (uglavnom mreže provajdera)
- U vreme NSFNET za razmenu ruta sa drugim mrežama koristio se EGP
- Sadašnja verzija BGP – BGP-4
- Interni protokoli rutiranja (unutar AS) imaju zadatak da obezbede prosleđivanje paketa tehnički najekonomičnijom putanjom
- BGP ima više netehničkih (političkih i komercijalnih) kriterijuma za izbor ruta

12

Autonomni sistemi



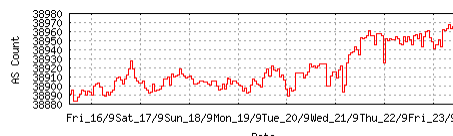
Autonomni sistemi se označavaju brojevima

AS brojevi od 64,512 do 65,535 su rezervisani za privatno korišćenje

13

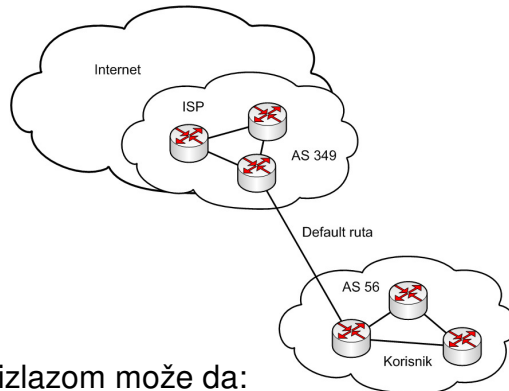
Tipovi autonomnih sistema

- AS sa jednim izlazom (stub, single-homed)
- AS sa više izlaza bez tranzita saobraćaja (multihomed nontransit)
- AS sa više izlaza i tranzitom saobraćaja (multihomed transit)
- Broj AS na Internetu (23.9.2011):



14

AS sa jednim izlazom

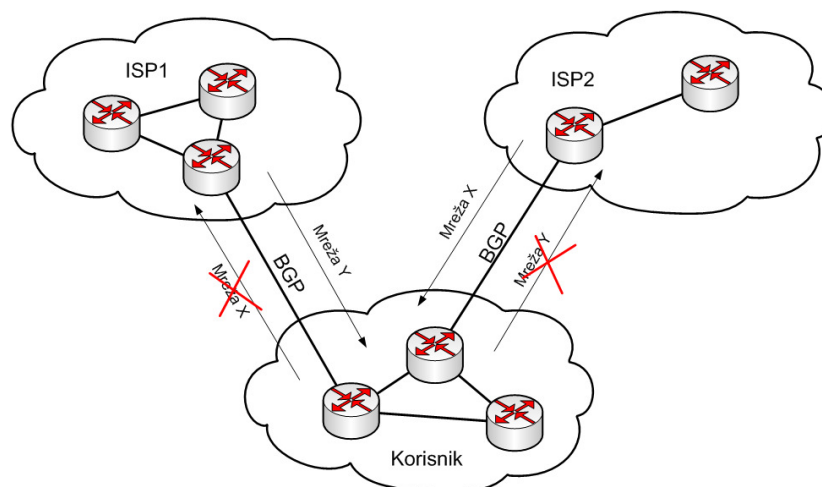


AS sa jednim izlazom može da:

- Koristi default rutu za prosleđivanje saobraćaja ka Internetu
- Bude deo IGP protokola svog provajdera
- Bude poseban privatan AS unutar provajderovog AS-a

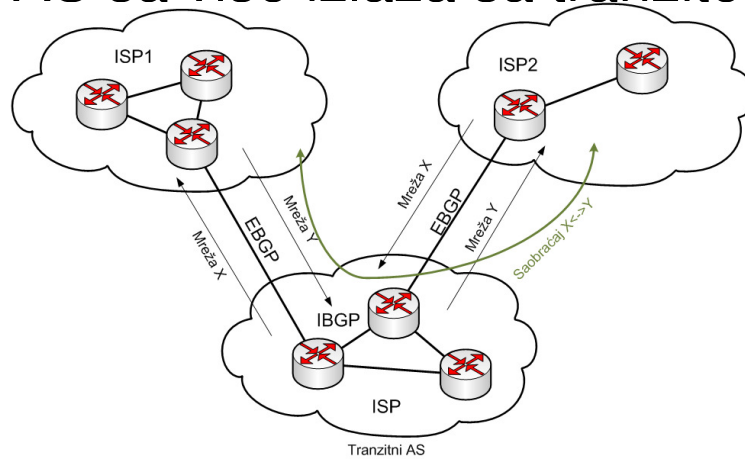
15

AS sa više izlaza bez tranzita



16

AS sa više izlaza sa tranzitom



BGP može da se koristi:

- izvan AS-a i onda je to eksterni BGP – EBGP
- unutar AS-a i onda je to interni BGP – IBGP

17

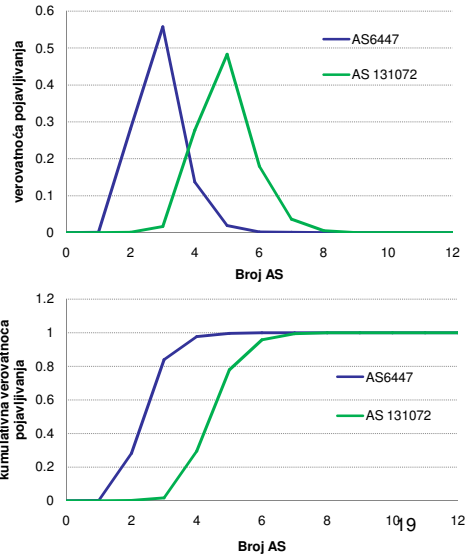
Kada se ne koristi BGP

- BGP se ne koristi u sledećim situacijama:
 - Postoji samo jedna veza sa Internetom ili ISP
 - Politika rutiranja date mreže je ista kao politika rutiranja ISP
 - Granični ruteri ne podržavaju ili nemaju dovoljno resursa za pokretanje BGP procesa
 - Mali propusni opseg između dve mreže

18

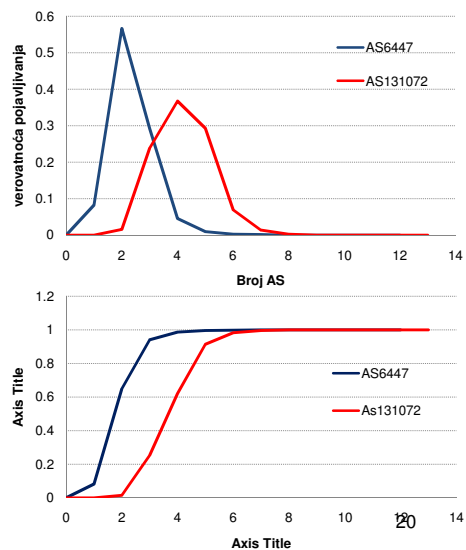
Koliko ima AS do destinacija na Internetu?

- Raspodela broja AS do destinacije u FIRT (dužina AS putanje bez prependinga)
- AS 6447: University of Oregon
- AS 131072: APNIC R&D



Koliko ima AS do destinacija na Internetu?

- Raspodela broja AS do destinacionih adresa u FIRT (dužina AS putanje)

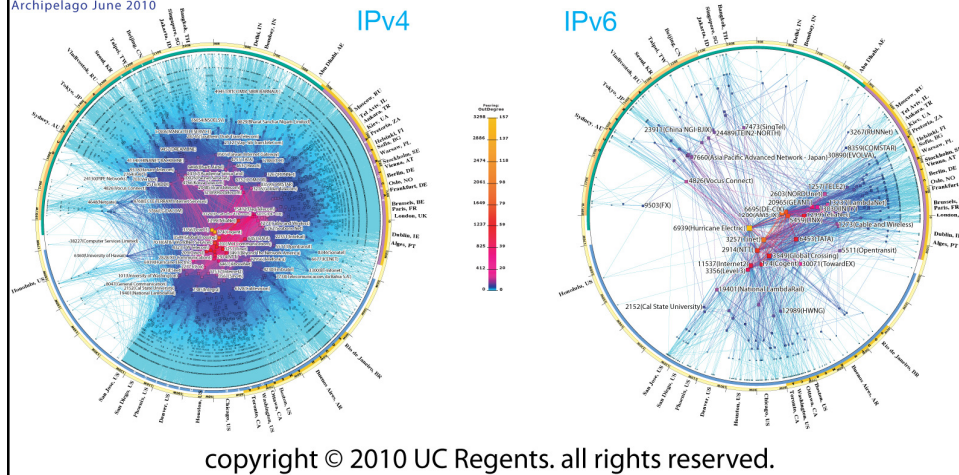


“Izgled” Interneta

- http://www.caida.org/research/topology/as_core_network/index.xml

CAIDA's IPv4 & IPv6 AS Core
AS-level INTERNET GRAPH

Archipelago June 2010



BGP – osnovne osobine - 1

- BGP je distance-vector protokol rutiranja, ali sa ugrađenim mehanizmom koji sprečava petlje u rutiranju
- BGP vrši odluke u rutiranju na osnovu pravila koja definiše administrator
- BGP-4 je trenutno aktuelna verzija i definisana je u RFC 1771 -> 4271 (+6286)
- BGP-4 je prva verzija BGP koja podržava CIDR i agregaciju ruta

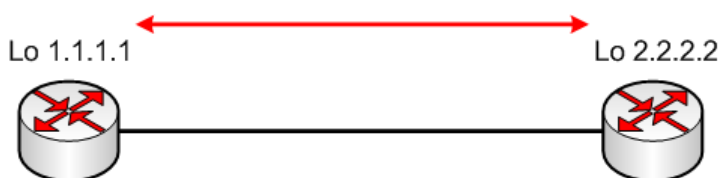
22

BGP – osnovne osobine - 2

- BGP koristi TCP po portu 179 za razmenu svojih poruka
- Pošto je BGP zasnovan na TCP protokolu, između dva rutera mora da postoji IP povezanost
- BGP proces čuva tabelu sa rutama i autonomnim sistemima iz kojih su dobijene date rute, čime obezbeđuje da ne dođe do petlji u rutiranju (kreira graf – stablo autonomnih sistema)
- Veza između dva autonomna sistema je put, a u BGP terminologiji se niz brojeva autonomnih sistema koji označavaju put do date rute zove **AS-path**
- BGP je lako proširiv i postoje brojne ekstenzije koje značajno povećavaju njegovu funkcionalnost (IPv6, VPN, Multicast) – RFC 2283 ->2858 ->4760 – Multiprotocol Extensions for BGP

23

BGP susedi



Kada dva rutera uspostave BGP konekciju, zovu se “**BGP peers**”

Svaki ruter koji ima pokrenut BGP proces i razmenjuje rute se zove “**BGP speaker**”

24

BGP - osnovni mehanizam funkcionisanja

- Ruteri razmenjuju BGP poruke kojima se uspostavlja BGP konekcija (**Open**)
- Ako postoji neslaganje u konfiguracionim parametrima (AS brojevi, IP adrese,...), BGP sesija se neće uspostaviti i šalju se odgovarajuće poruke (**Notification**)
- Kada se uspostavi BGP sesija ruteri razmenjuju sve poznate rute (**Update**)
- Nakon toga rute se razmenjuju samo kada dođe do promene BGP ruta u ruting tabelama (inkrementalno prosleđivanje poruka)
- BGP nekada razmenjuje i punu Internet ruting tabelu

25

BGP mehanizam funkcionisanja

- **Update** poruke se sastoje od: prefiksa, AS-path-a i atributa AS-path-a kojima se bliže određuje način tretiranja date rute
- BGP ruteri čuvaju broj verzije BGP ruting tabele susednih rutera.
- Broj verzije se inkrementira prilikom svake promene (ubacivanje ili izbacivanje rute)
- Ako nema nikakvih promena razmenjuju se **Keepalive** poruke.
- Keepalive poruke se šalju svakih 60 sekundi i veličine su svega 19 bajta.

26

Veličina pune Internet ruting tabele

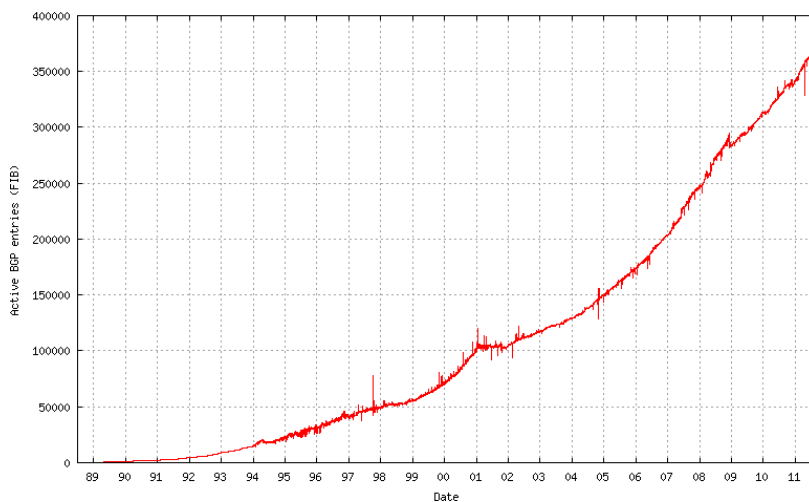
```

cisco6509#sh ip bgp summary
BGP router identifier 147.91.0.112, local AS number 13092
BGP table version is 4996862, main routing table version 4996862
174464 network entries using 17620864 bytes of memory
866812 path entries using 41608976 bytes of memory
158780 BGP path attribute entries using 8892632 bytes of memory
86314 BGP AS-PATH entries using 2795968 bytes of memory
484 BGP community entries using 24390 bytes of memory
1 BGP extended community entries using 24 bytes of memory
224552 BGP route-map cache entries using 7185664 bytes of memory
0 BGP filter-list cache entries using 0 bytes of memory
BGP using 78126518 total bytes of memory
346915 received paths for inbound soft reconfiguration
BGP activity 484815/309810 prefixes, 3868941/3001587 paths, scan interval 60 secs
    
```

Neighbor	V	AS	MsgRcvd	MsgSent	TblVer	InQ	OutQ	Up/Down	State/PfxRcd
147.91.0.3	4	13092	24441	24877	4996862	0	0	1d19h	32
147.91.8.77	4	6701	27792	24282	4996859	0	0	1w0d	27
160.99.1.12	4	13303	24253	24871	4996851	0	0	2w2d	1
193.254.1.242	4	8214	0	0	0	0	0	never	Active
194.141.252.13	4	6802	0	0	0	0	0	never	Active
194.149.130.249	4	5379	24171	24239	4996859	0	0	1d03h	5
194.177.210.40	4	5408	0	0	0	0	0	never	Active
195.111.106.254	4	1955	24264	24276	4996859	0	0	2w2d	15
195.178.34.57	4	8400	1110287	24292	4996862	0	0	2d14h	172842
195.178.35.17	4	8400	1079451	24286	4996862	0	0	2w2d	172851
195.251.4.44	4	34771	653186	48558	4996859	0	0	1w1d	174121

27

Veličina Internet ruting tabele



28

BGP vrste poruka

- Sve poruke počinju BGP zaglavljem.
- Zaglavlje ima samo tri polja: 16-bajtni **Marker**, a 2-bajtno polje **Length**, i 1-bajtno polje **Type**
- Marker služi za autentikaciju BGP speaker-a ili za detekciju gubitka sinhronizacije
 - U Open poruci Marker = sve jedinice
 - U ostalim porukama može da bude MD5 hash, ako se taj mehanizam koristi za autentikaciju BGP speaker-a
- Polje **Type** može da ima 4 vrednosti koje označavaju tipove poruka:
 - Open Poruka
 - Keepalive Poruka (samo BGP zaglavlje)
 - Notification Poruka
 - Update Poruka

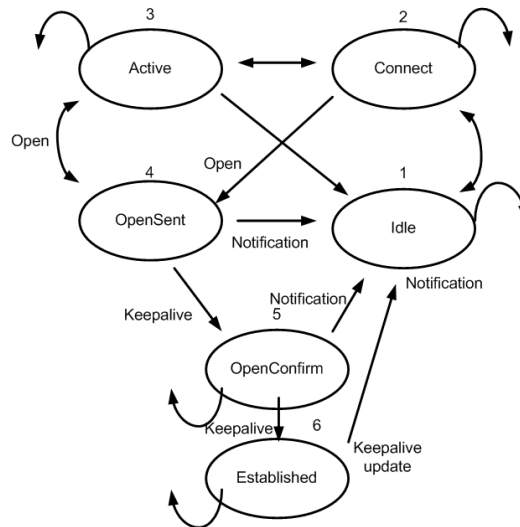
29

Poruke - detalji

- **Open** poruka – Služi za uspostavljanje BGP sesije i sadrži sledeće podatke: **BGP version number (4)**, **AS broj**, **hold time**, i **Router ID**.
 - If hold time=0 => ne šalju se keepalive paketi
 - Keepalive interval = hold time/3
- **Notification** poruke – služe za obaveštavanje suseda o eventualnim greškama. Greške se označavaju standardom definisanim kodovima.
- **Update** poruke – Sadrže informacije o novim ili o izbačenim rutama. Tri glavne komponente Update poruke su: **network-layer reachability information (NLRI)**, **path atributi**, i **povučene rute**
- **NLRI** = (network_address, prefix_length)
 - Primer: (147.91.0.0,16) = 147.91.0.0/16
- Povučene rute su u NLRI formatu

30

Mašina stanja uspostavljanja BGP sesije



31

Idle stanje

- **Idle** je prvo stanje BGP konekcije.
- BGP čeka za događaj koji će da započne uspostavljanje susedskih odnosa. To je obično nešto što uradi administrator.
- Nakon toga BGP prelazi u stanje **Connect** i resetuje **ConnectRetry** brojač i počinje uspostavljanje TCP sesije.
- Osluškuje eventualno iniciranje TCP sesije od strane BGP suseda

32

Connect stanje

- **Connect** – U ovom stanju BGP susedi čekaju na uspostavljanje TCP sesije.
- Ako se TCP sesija uspešno uspostavi, šalje se Open poruka i prelazi u **OpenSent**.
- Ako ne uspe da uspostavi TCP sesiju prelazi u **Active** stanje, i ponovo pokušava da uspostavi TCP sesiju.
- Ako istekne ConnectRetry timer, BGP ostaje u **Connect** stanju.
- Pod uticajem administrativnih aktivnosti može da se vrati u **Idle**.

33

Active stanje

- **Active** – BGP pokušava da uspostavi TCP sesiju.
- Ako se TCP sesija uspešno uspostavi, šalje se Open poruka i prelazi u **OpenSent**.
- Ako istekne ConnectRetry tajmer, vraća se u **Connect** stanje.
- **Ukoliko stanje osciluje između Connect i Active, to je indikacija da postoji problem u ostvarivanju TCP sesije**

34

OpenSent stanje

- **OpenSent** – Ruter je poslao **Open** poruku i čeka na uspostavljanje BGP sesije.
- **Open** poruka se proverava. Ako ima grešaka, šalje se Notification poruka i ruter se vraća u **Idle**.
- Ako nema grešaka BGP počinje da šalje **Keepalive** poruke i resetuje **Keepalive** tajmer.
- Kada se primi **Open** poruka od suseda, prepoznaje se da li će sa njim da se uspostavi IBGP ili EBGP sesija.
- Ako se prekine TCP sesija, BGP se vraća u **Active** stanje.

35

OpenConfirm stanje

- **OpenConfirm** – U ovom stanju ruter čeka poruke od suseda
 - Ako dobije **Keepalive** ili **Update** poruke prelazi u **Established** stanje
 - Ako dobije **Notification** poruku, vraća se u **Idle**
- Ako dođe do bilo kakve druge greške ili do prekida TCP sesije, vraća se u **Idle**.

36

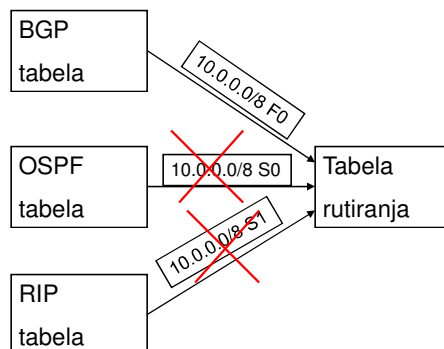
Established state

- **Established** – u ovom stanju počinje razmena Update poruka.
- U slučaju regularnog rada, BGP ostaje u ovom stanju.
- U slučaju bilo kakve greške, BGP se vraća u **Idle** stanje.

37

Više protokola rutiranja na istom ruteru

U svakom ruteru postoji samo jedna tabela rutiranja!
Različiti protokoli rutiranja imaju različite, neuporedive metrike



Route Source	Default Distance
Connected interface	0
Static route	1
External BGP	20
OSPF	110
IS-IS	115
RIP	120
Internal BGP	200
Unknown	255

Ruta sa manjom distancom ulazi u tabelu rutiranja

38

Koja će od ovih ruta ući u tabelu rutiranja?

BGP	192.168.32.0/26
BGP	192.168.32.0/24
RIP	192.168.32.0/24
RIP	192.168.32.0/19
RIP	192.168.32.0/28
OSPF	192.168.32.0/19



RP	Prefix
B	192.168.32.0/26
B	192.168.32.0/24
R	192.168.32.0/28
O	192.168.32.0/19

39

Izbor ruta

Rute se biraju prema najdužem prefiksu (prema najdužoj maski)

RP	Prefix	Dis	Met	Outgoing int
R	9.0.0.0/8	120	5	S0/0
O	10.0.0.0/8	110	100	S0/1
B	10.10.0.0/16	20	0	F0/0
R	10.10.10.0/24	120	7	F0/0
O	10.10.12.0/24	110	200	S0/0

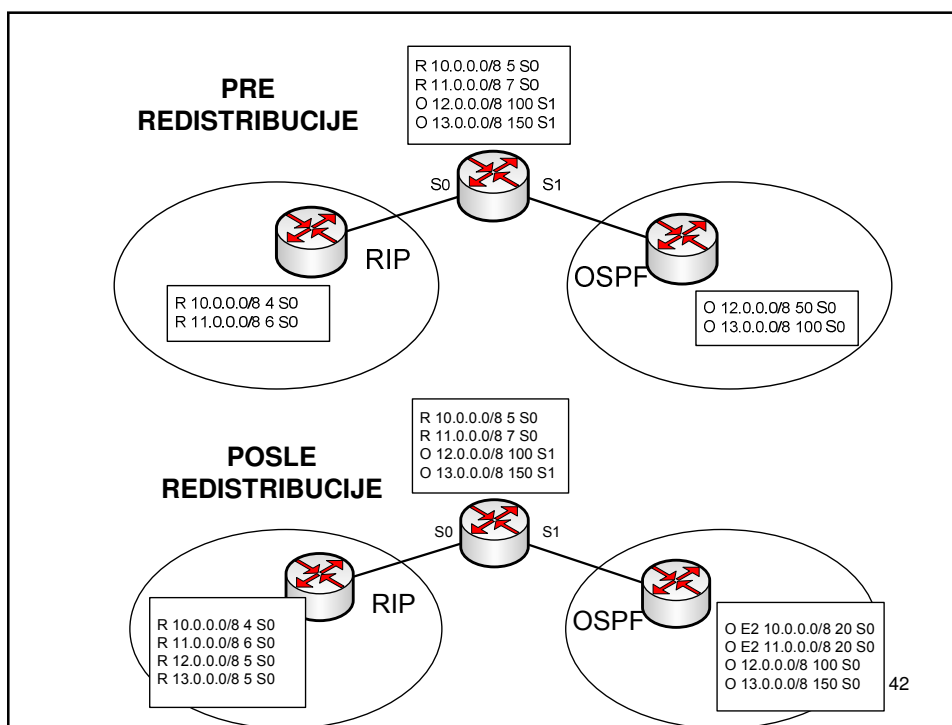
- Na koji interfejs će biti upućeni paketi za:
 - 10.10.10.1
 - 10.10.13.1
 - 10.12.12.1

40

Redistribucija ruta

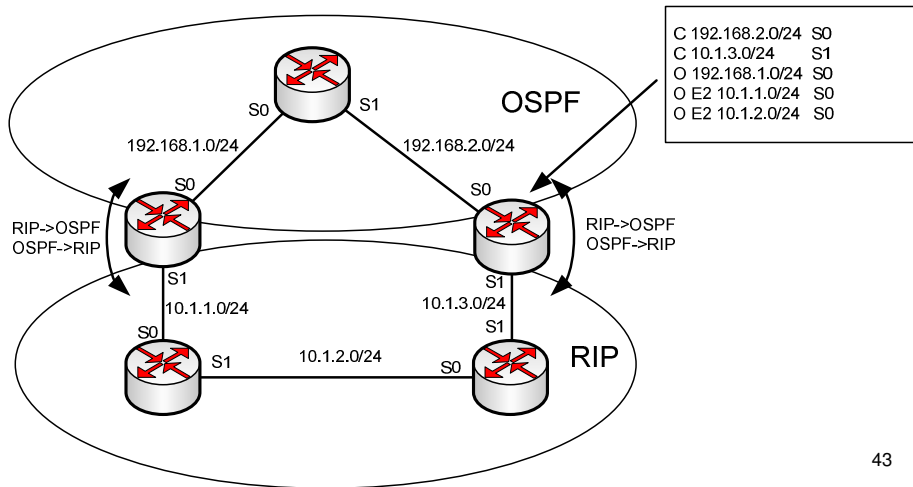
- Ruter redistribuira samo one rute koje su u tabeli rutiranja
- Redistribucija se uvek vrši u izlaznom smeru tj. ruter koji vrši redistribuciju ne menja svoju tabelu rutiranja
- Prilikom redistribucije mora da se podesi metrika prema protokolu u koji se vrši redistribucija
- Može da se vrši i redistribucije default rute

41



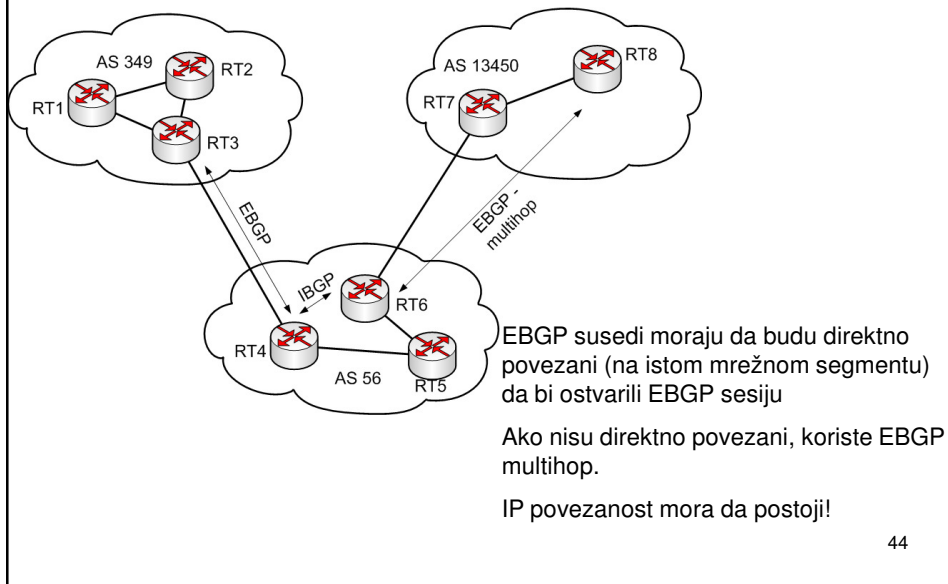
Problem kada postoji više veza

Može da dođe do neoptimalnog rutiranja ili petlji u rutiranju



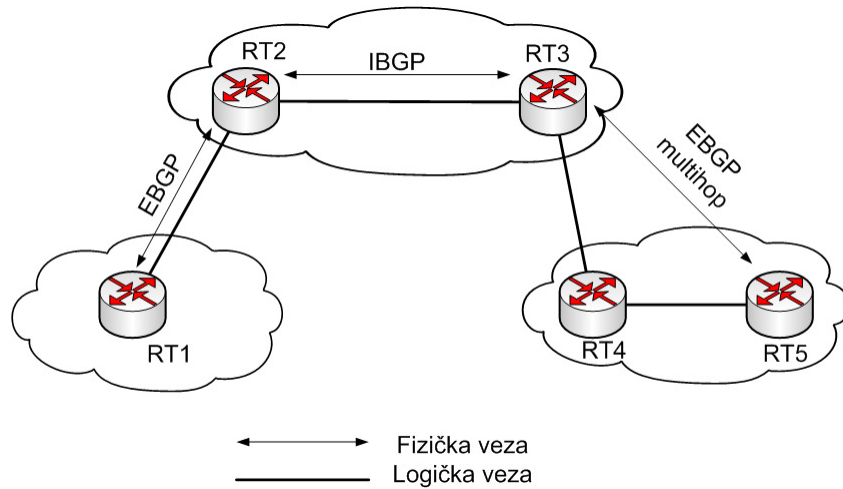
43

EBGP Multihop



44

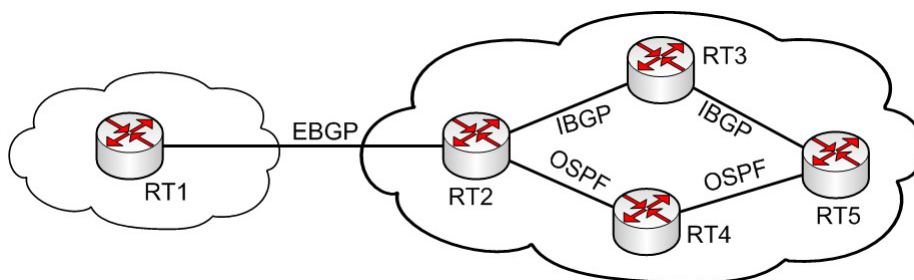
Različite vrste BGP sesija



45

EBGP i IBGP

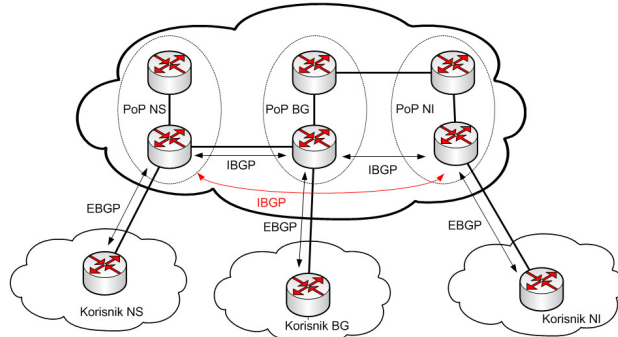
IBGP ruteri NE oglašavaju rute dobijene od rutera putem IBGP drugim IBGP susedima



IBGP se koristi za sinhronizaciju i koordinaciju rutiranja unutar AS

46

Kontinuitet BGP unutar AS



BGP ne oglašava rute dobijene putem IBGP drugim IBGP susedima

Ako bi BGP to radio, to bi stvorilo opasnost od stvaranja petlji u rutiranju (šta je u BGP protokolu zaštita od petlji?)

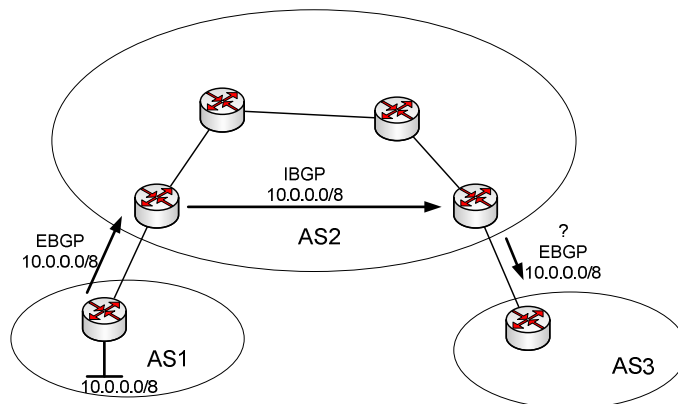
Da bi IBGP ruteri naučili sve rute unutar AS, moraju da se povežu sa svim IBGP ruterima unutar AS u potpun graf (full IBGP mesh).

Ovaj potpuni IBGP graf ne mora da bude fizički nego može da bude logički.

47

Sinhronizacija unutar AS

- BGP mora da bude sinhronizovan sa IGP protokolom unutar AS da bi smeo da prosledi rute dobijene od IBGP eksternim susedima



48

Sinhronizacija unutar AS

- Rešenje za sinhronizaciju:
 - Redistribucija svih ruta u IGP (potencijalni problemi sa skalabilnošću i performansama IGP)
 - Interni ruteri imaju default rute ka jednom izlaznom ruteru (neoptimalno rutiranje)
 - Potpun IBGP graf unutar AS (na svim ruterima) i isključena sinhronizacija

49

sh ip bgp

```
cisco6509#sh ip bgp
BGP table version is 5011434, local router ID is 147.91.0.112
Status codes: s suppressed, d damped, h history, * valid, > best, i - internal,
               S Stale
Origin codes: i - IGP, e - EGP, ? - incomplete

   Network        Next Hop        Metric LocPrf Weight Path
* 0.0.0.0          195.178.34.57      150           0 8400 8400 i
*>                 195.178.35.17      150           0 8400 8400 i
*> 3.0.0.0          195.178.35.17      150           0 8400 8400 702 703 80 i
*                 195.178.34.57      150           0 8400 8400 702 703 80 i
*                 195.251.4.44       150           0 34771 5408 20965 3356 701 703 80 i
* 4.0.0.0          195.178.34.57      150           0 8400 8400 5400 3356 i
*                 195.178.35.17      150           0 8400 8400 5400 3356 i
*>                 195.251.4.44       150           0 34771 5408 20965 3356 i
* 4.23.84.0/22     195.178.34.57      150           0 8400 8400 5400 6461 20171 i
*>                 195.178.35.17      150           0 8400 8400 5400 6461 20171 i
*                 195.251.4.44       150           0 34771 5408 20965 1299 6461 20171 i
* 4.23.112.0/22   195.178.34.57      150           0 8400 8400 5400 174 21889 i
*>                 195.178.35.17      150           0 8400 8400 5400 174 21889 i
*                 195.251.4.44       150           0 34771 5408 20965 1299 174 21889 i
* 4.23.180.0/24   195.178.34.57      150           0 8400 8400 5400 6128 30576 i
*>                 195.178.35.17      150           0 8400 8400 5400 6128 30576 i
*                 195.251.4.44       150           0 34771 5408 20965 1299 6128 30576 i
```

50

show ip bgp summary

```
cisco6509#sh ip bgp summary
BGP router identifier 147.91.0.112, local AS number 13092
BGP table version is 5011825, main routing table version 5011825
174488 network entries using 17623288 bytes of memory
866789 path entries using 41605872 bytes of memory
158814 BGP path attribute entries using 8898456 bytes of memory
86369 BGP AS-PATH entries using 2797716 bytes of memory
484 BGP community entries using 24374 bytes of memory
1 BGP extended community entries using 24 bytes of memory
226170 BGP route-map cache entries using 7237440 bytes of memory
0 BGP filter-list cache entries using 0 bytes of memory
BGP using 78187170 total bytes of memory
346919 received paths for inbound soft reconfiguration
BGP activity 485542/310513 prefixes, 3874396/3007065 paths, scan interval 60 secs
```

Neighbor	V	AS	MsgRcvd	MsgSent	TblVer	InQ	OutQ	Up/Down	State/PfxRcd
147.91.0.3	4	13092	24571	25008	5011797	0	0	1d21h	32
147.91.8.77	4	6701	27950	24407	5011765	0	0	1w0d	27
160.99.1.12	4	13303	24378	25004	5011768	0	0	2w2d	1
193.254.1.242	4	8214	0	0	0	0	0	never	Active
194.141.252.13	4	6802	0	0	0	0	0	never	Active
194.149.130.249	4	5379	24297	24365	5011768	0	0	1d05h	5
194.177.210.40	4	5408	0	0	0	0	0	never	Active
195.111.106.254	4	1955	24389	24402	5011792	0	0	2w2d	15
195.178.34.57	4	8400	1115639	24417	5011768	0	0	2d16h	172845
195.178.35.17	4	8400	1084803	24411	5011768	0	0	2w2d	172854
195.251.4.44	4	34771	655799	48809	5011792	0	0	1w1d	174088

51

show ip bgp neighbors

```
cisco6509#sh ip bgp nei
BGP neighbor is 147.91.0.3, remote AS 13092, internal link
Description: AMREJ
BGP version 4, remote router ID 147.91.0.113
BGP state = Established, up for 1d21h
Last read 00:00:42, hold time is 180, keepalive interval is 60 seconds
Neighbor capabilities:
  Route refresh: advertised and received(new)
  Address family IPv4 Unicast: advertised and received
Message statistics:
  InQ depth is 0
  OutQ depth is 0
      Sent      Rcvd
Opens:          2         2
Notifications: 0         0
Updates:       629       194
Keepalives:   24377     24375
Route Refresh: 0         0
Total:       25008     24571
Default minimum time between advertisement runs is 5 seconds

For address family: IPv4 Unicast
BGP table version 5011868, neighbor version 5011868
```

52

show ip bgp paths

```
cisco6509#sh ip bgp paths
Address      Hash Refcount Metric Path
0x579E7DD0  0      1      0 8400 8400 1299 3343 2895 2895 2587 i
0x48A29898  0      3      0 8400 8400 1299 8928 31222 i
0x52468890  0      1     150 8400 8400 5400 209 15194 i
0x4611BA50  0      2      0 8400 702 30829 i
0x53F54BC8  0      5      0 8400 702 20485 6767 i
0x581F5450  0      3      0 8400 8400 1299 19962 30444 i
0x581EDE78  0      2      0 8400 8400 1299 3549 26315 i
0x5421C6A0  0      2     150 8400 1299 2828 5725 i
0x53CE44E8  0      2      0 8400 1299 7911 16905 1832 i
0x53CE63D8  0      4     150 8400 8400 5400 7018 16609 i
0x57BF0018  0      1     150 8400 8400 5400 7018 16609 i
0x46A26C70  0      1     150 8400 8400 5400 5511 6505 21862 i
0x568B45B0  0      1     150 8400 8400 5400 174 27429 i
0x48A2CE98  0      3      0 34771 5408 20965 1299 1239 13228 25465
```

53

Atributi putanje (Path attributes)

- Konfiguracija BGP-a podrazumeva konfigurisanje atributa ruta i putanja.
- Za svaku rutu postoje definisani atributi.
- Atributi se dele u 4 grupe:
 - Dobro poznati obavezni (Well-known mandatory) – Moraju da postoje u svakoj BGP Update poruci pridružen odgovarajućoj NLRI. Sve implementacije BGP-a moraju da koriste ove poruke. Nedostatak ovih atributa u Update poruci generiše grešku.
 - Dobro poznati neobavezni (Well-known discretionary) – Atribut koji prepoznaju sve BGP implementacije i u skladu sa njim se ponašaju, ali koji ne mora da bude pridružen nekom NLRI.
 - Opcioni prenosivi (Optional transitive) – Atribut koji ne moraju da prepoznaju sve BGP implementacije i da se ponašaju u skladu sa njim (opcionim). Ako ruter dobije ovakav atribut koji ne prepoznaje, onda treba da ga prosledi ostalim BGP susedima (prenosivost).
 - Opcioni neprenosivi (Optional nontransitive) – Atribut koji ne moraju da prepoznaju sve BGP implementacije i da se ponašaju u skladu sa njim (opcionim). Ako ruter dobije ovakav atribut koji ne prepoznaje, onda ga ne prosleđuje dalje (neprenosivost)

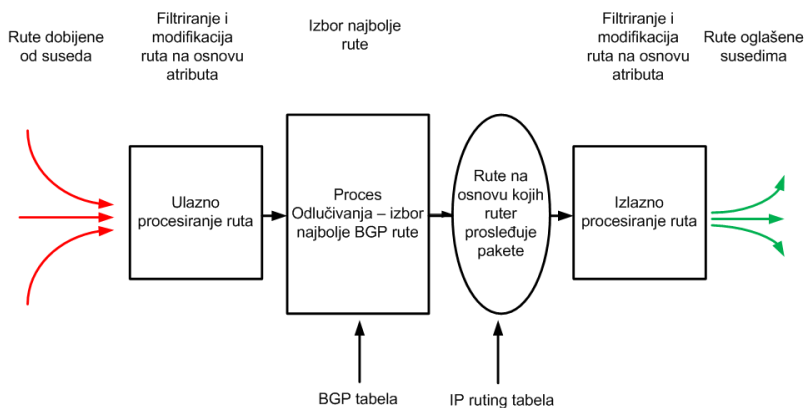
54

Atributi putanje (Path attributes)

Kôd atributa	Tip
1-ORIGIN	Well-known mandatory
2-NEXT-HOP	Well-known mandatory
3-AS_PATH	Well-known mandatory
4-MULTI_EXIT_DISC	Optional nontransitive
5-LOCAL_PREF	Well-known discretionary
6-ATTOMIC_AGGREGATE	Well-known discretionary
7-AGGREGATOR	Well-known discretionary
8-COMMUNITY	Optional transitive
9-ORIGINATOR_ID	Optional nontransitive
10-Cluster list	Optional nontransitive

55

BGP Proces rutiranja



U slučaju da postoji više BGP ruta ka nekoj destinaciji, BGP neće svojim susedima da prosledi sve te rute, već **samo najbolju**.

56

Kontrolisanje BGP rutiranja korišćenjem BGP atributa

- Uobičajeni BGP atributi
 - Next Hop
 - AS_Path
 - Atomic Aggregate
 - Aggregator
 - Local Preference
 - Weight
 - Multiple Exit Discriminator (MED)
 - Origin

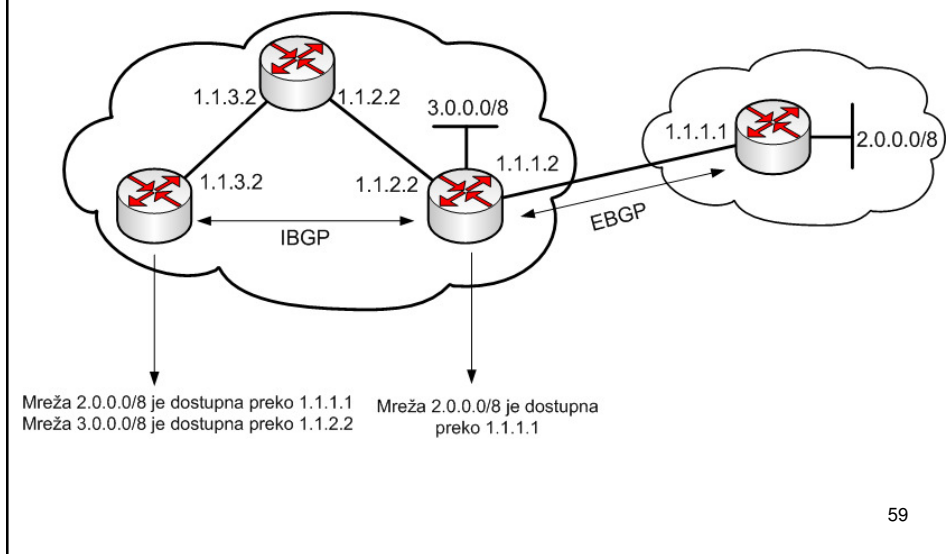
57

Next hop atribut (WMA)

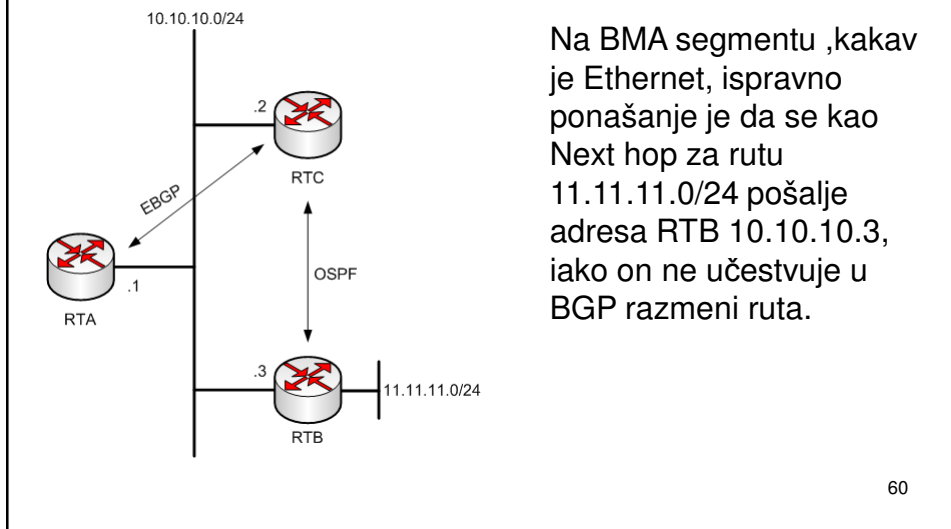
- Next hop ne mora da nužno bude na direktno povezanom mrežnom segmentu.
- Pravila kako BGP koristi next hop atribut:
 - U **EBGP** sesijama, next hop je IP adresa EBGP suseda koji je oglašio datu rutu.
 - U **IBGP** sesijama, ako su rute oglašene unutar samog AS, next hop je IP adresa rutera unutar AS koji je oglašio datu rutu.
 - U **IBGP** sesijama, ako su rute oglašene u AS iz nekog drugog AS putem EBGP, next hop koji je dobijen putem EBGP se unosi nepromenjen u IBGP
- Ukoliko ruter u svojoj ruting tabeli nema rutu ka Next Hop atributu za datu rutu, ruta neće biti ubačena u ruting tabelu.

58

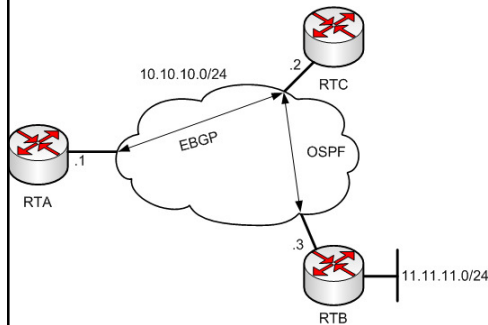
Next Hop Atribut



Next Hop na Multiaccess segmentima



Next Hop ponašanje na NBMA mrežama



Na NBMA mrežama ukoliko se ništa ne konfigurira, ruteri će se ponašati isto kao na BMA mrežama.

Međutim, to može da dovede do prekida u komunikaciji, jer možda ne postoji virtuelno kolo između datih rutera, pa treba da se konfigurira u ovom slučaju da je Next Hop za rutu 11.11.11.0 RTC.

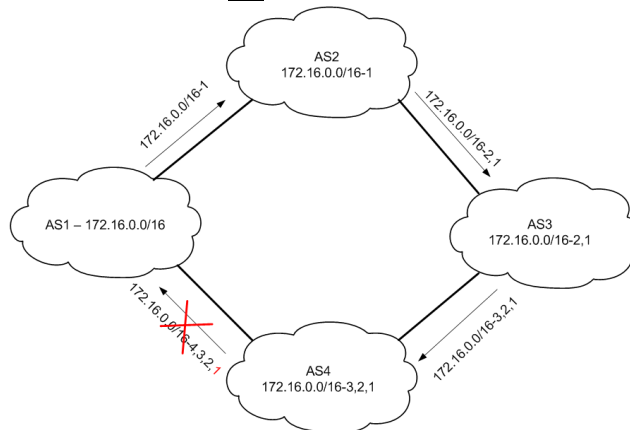
61

AS_Path Atribut (WMA)

- AS koji oglašava neku rutu dodaje broj svog AS u ASpath atribut pridružen datoj ruti.
- Svaki sledeći AS dodaje (**prepend**) svoj broj AS datoj ruti prilikom prosleđivanja narednom AS.
- Ukoliko ruter kada dobije neku rutu prepozna broj svog AS u AS Path atributu, ruta se odbacuje!!!
- BGP koristi između ostalih kriterijuma AS_Path u procesu izbora najbolje putanje.
- Kraći AS Path označava rutu sa boljom metrikom koja će biti ubačena u ruting tabelu .

62

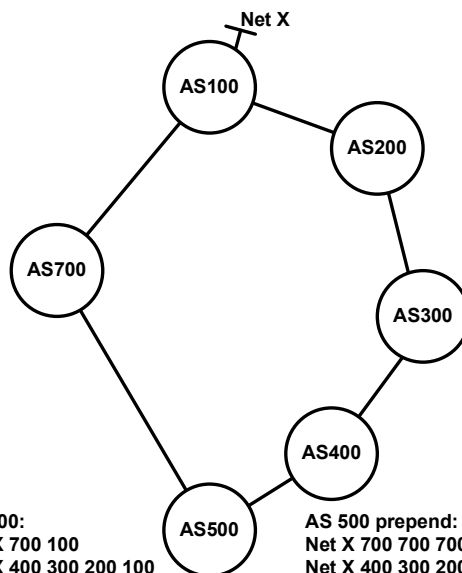
AS_Path Atribut



Često se veštački povećava broj AS u AS-Path atributu, kako bi se uticalo na izbor najbolje rute (AS Path Prepending)

63

AS Path prepending

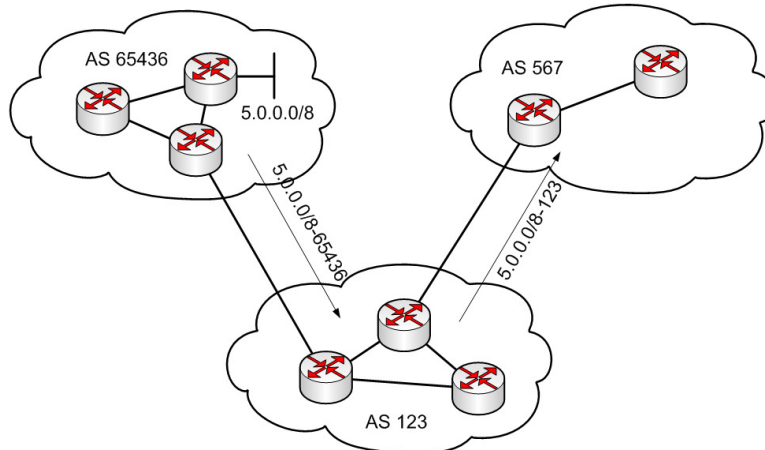


AS 500:
 Net X 700 100
 Net X 400 300 200 100

AS 500 prepend:
 Net X 700 700 700 700 100
 Net X 400 300 200 100

64

AS_Path i privatni AS brojevi



Privatni AS brojevi moraju da se skinu iz AS Path atributa pre nego što neka ruta prođe ka ostatku Interneta.

65

Origin Atribut (WMA)

- Origin atribut govori o poreklu rute/prefiksa
- Koristi se u izboru najbolje rute
- Vrste origin atributa:
 - IGP – Prefiks je dobijen iz IGP iz datog AS (eksplicitno konfigurisana ruta za oglašavanje u konfiguraciji BGP protokola)
 - EGP – Prefiks je dobijen iz EGP [RFC4271]
 - Incomplete – Prefiks je dobijen redistribucijom
- Više su cenjene rute sa manjom vrstom Origin a odnos vrsta je:
 - IGP < EGP < Incomplete

66

Origin atribut

- Origin EGP – EGP se ne koristi, pa ipak se rute sa ovim Origin-om pojavljuju u tabelama.
- RFC 4271: “ORIGIN...SHOULD NOT be changed by any other speaker”

```
amres-core-R#sh ip bgp | inc e
BGP table version is 19841950, local router ID is 147.91.0.124
Status codes: s suppressed, d damped, h history, * valid, > best, i - internal,
S Stale
Origin codes: i - IGP, e - EGP, ? - incomplete
Network Next Hop Metric LocPrf Weight Path
*>i14.96.0.0/20 147.91.0.127 100 0 1955 1299 6453 4755 55740 e
*>i14.96.16.0/20 147.91.0.127 100 0 1955 1299 6453 4755 55740 e
*>i14.96.32.0/20 147.91.0.127 100 0 1955 1299 6453 4755 55740 e
*>i14.96.48.0/20 147.91.0.127 100 0 1955 1299 6453 4755 55740 e
*>i14.96.64.0/20 147.91.0.127 100 0 1955 1299 6453 4755 55740 e
*>i14.96.80.0/20 147.91.0.127 100 0 1955 1299 6453 4755 55740 e
```

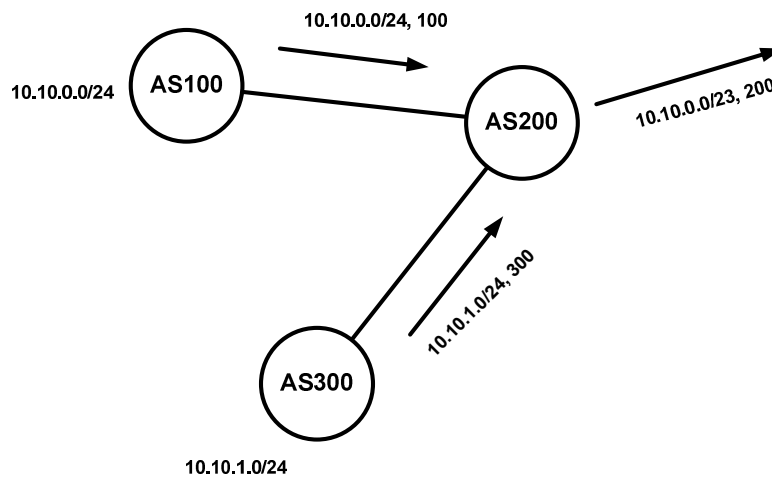
67

Atomic Aggregate Atribut (WDA)

- Koristi se kod agregacije ruta i označava gubitak informacija u AS Path atributu
- Može da ima vrednost True ili False.
- Ako je True, znači da je dati prefiks agregiran od više različitih prefiksa.
- Ruter koji je poslao prefiks sa Atomic Aggregate atributom sa vrednošću True je izvršio agregaciju ruta i ima specifičnije rute do destinacija

68

Agregacija prefiksa



69

The Aggregator Atribut (OT)

- Ovim atributom se označava onaj ruter koji izvršio agregaciju ruta.
- Kao argument ovog atributa upisuje se Router ID rutera koji je izvršio agregaciju

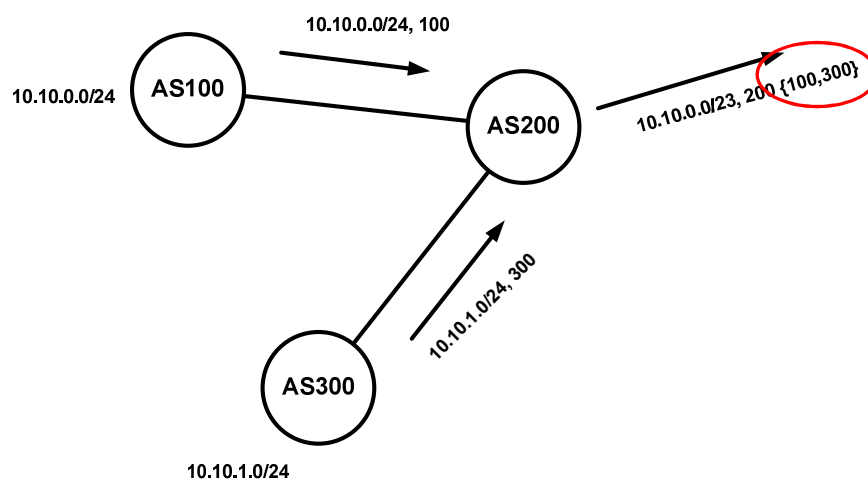
70

AS-Set

- Korišćenjem agregacije se smanjuje broj ruta u tabeli rutiranja, ali se gube neke informacije.
- Postoji mogućnost da se koristi posebna vrsta ASPath objekta koji se zove AS-Set, kojim se zadržavaju informacije o agregiranim rutama.
- AS Set se sastoji od agregirane rute i elemenata koji je sačinjavaju
- Nije dobro da se koristi kada se agregira veliki broj ruta.

71

AS-Set



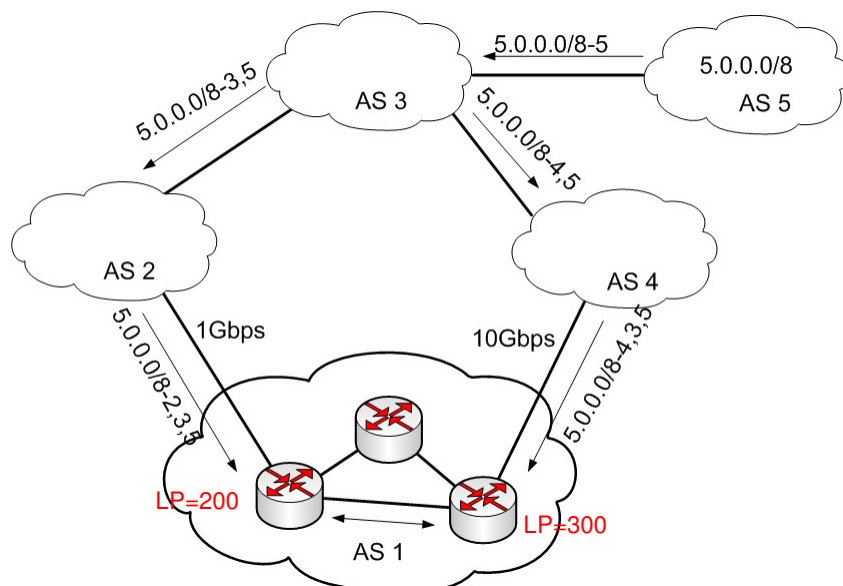
72

Local Preference Atribut (WDA)

- Local Preference atribut označava stepen prvenstva (prioritet) za datu rutu
- U ruting tabelu se ubauje ona ruta koja ima viši Local Preference.
- Local Preference, je lokalan za jedan AS.
- Local Preference se razmenjuje unutar jednog AS putem IBGP, ali se ne prenosi putem EBGP.
- Local Preference se odnosi na saobraćaj koji **izlazi** iz datog AS!!!

73

Local Preference Atribut (WDA)



4

Manipulacija Local Preference atributom

- Default vrednost za Local Preference je 100
- Local Preference može da se dodeli nekoj ruti na osnovu njene adrese, interfejsa sa kog dolazi, broja AS iz kog dolazi ili koji se nalazi u AS-Path atributu,...
- Local Preference se dodeljuje ruti nakon što ruda dođe u dati ruter

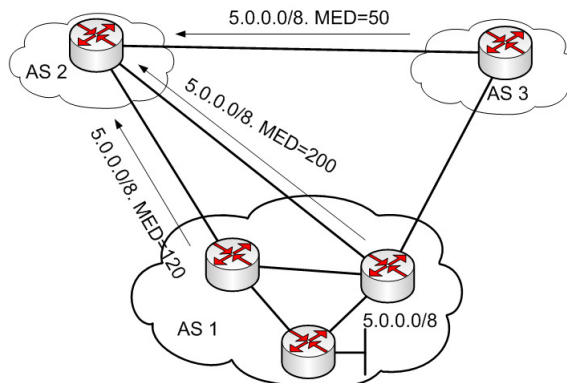
75

Multiple Exit Discriminator (MED) Atribut (ONTA)

- MED obaveštava susedne AS o željenoj putanji saobraćaja u dati AS ukoliko dati AS ima više veza sa drugim AS-om.
- **Niži MED ima prednost u odnosu na viši (zato se i zove još i metrika)**
- MED atribut koji uđe u AS ne napušta ga (netranzitivna je)
- MED se dodeljuje rutama koje izlaze iz datog rutera
- MED obično odslikava metriku IGP datog AS

76

MED Atribut

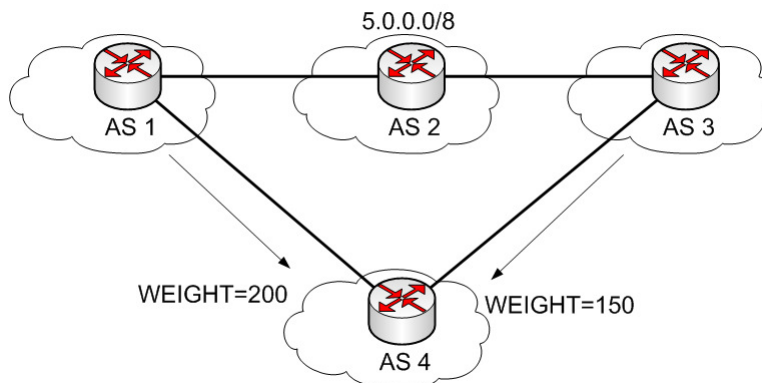


Jedan AS poredi MED vrednosti samo za prefikse dobijene iz jednog AS

MED atributi za prefiks 5.0.0.0/8 iz AS 1 i AS 3 se porede u AS 2 samo u specijalnim slučajevima

77

Weight Atribut



- Ovaj atribut je Cisco specifičan atribut
- Lokalni je za ruter i ne razmenjuje se sa drugim ruterima
- Utiče na saobraćaj koji izlazi iz autonomnog sistema, a na rute koje ulaze u AS

78

Proces izbora najbolje rute u BGP protokolu rutiranja

1. Ako Next Hop atribut za datu rutu ne postoji u ruting tabeli, ruta se ignoriše tj ne ubacuje u ruting tabelu.
2. (ako postoji Weight atribut, u ruting tabelu ulazi ruta sa **najvećom** Weight vrednošću)
3. Ako su Weight vrednosti iste, u ruting tabelu ulazi ruta sa **najvećom** vrednošću Local Preference
4. Ako su Local Preference vrednosti iste, u ruting tabelu ulazi ruta koju je oglasio dati ruter

79

Proces izbora najbolje rute u BGP protokolu rutiranja

5. Ako su prethodni kriterijumi isti, u ruting tabelu ulazi ruta sa **kraćim** AS-Path-om
6. Ako su AS-Path dužine iste, ruter će odabrati rutu sa **nižom** vrednošću Origin atributa
7. Ako je i to isto, Ruter će odabrati rutu sa **nižom** MED vrednošću
8. Ako su i MED vrednosti iste, ruter će da odabere prvo rute čije putanje idu preko EBGP konekcija u odnosu na dobijene iz IBGP konekcija

80

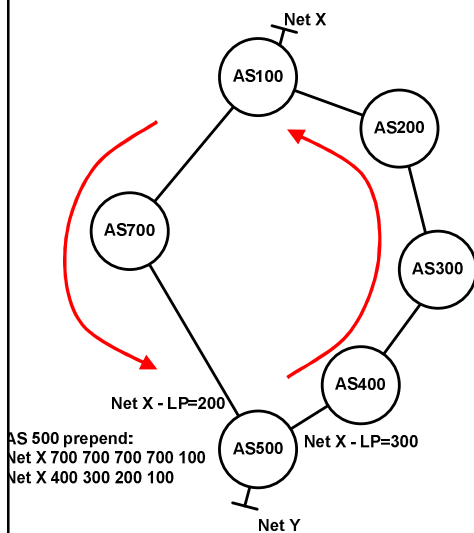
Proces izbora najbolje rute u BGP protokolu rutiranja

9. Bira se ruta čija je IGP metrika do BGP Next hopa niža
10. Ako su prethodni kriterijumi isti, bira se ruta koja je dobijena ranije (prva koja je stigla u ruter)
11. Ako su i prethodni kriterijumi isti, bira se ruta dobijena od rutera sa nižim Router ID-em.
12. Odabira se putanja sa nižom vrednošću dužine klastera.
13. Odabira se ruta dobijena od suseda sa nižom adresom.

ZAKLJUČAK: Uvek će biti odabrana JEDNA najbolja ruta ka datom prefiksu

81

Primer



- Kojom putanjom će ići saobraćaj od NetY ka NetX?
- Kojom putanjom će ići saobraćaj od NetX ka NetY?

82

Atribut Community (OT)

- Služi za grupisanje mreža za koje se traži određeni način procesiranja od nekog nadređenog AS
- 4-bajtni parametar
- Neke predefinisane vrednosti
 - 0xFFFFFFFF01 – No export – ne oglasiti eBGP susedima
 - 0xFFFFFFFF02 – No advertise – ne oglasiti nikome
 - 0x00000000 do 0x0000FFFF i 0xFFFF0000 do 0xFFFFFFFF ne mogu da se koriste slobodno
 - Uobičajen način predstavljanja: AS:COMMUNITY

83

Redundansa, simetrija i balansiranje saobraćaja

- Redundansa
 - Povećanje pouzdanosti kroz obezbeđivanje alternativnih putanja
 - Pošto je uslov za dobijanje AS veza ka dva druga AS, redundansa uvek postoji
- Simetričnost saobraćaja
 - ulazni i izlazni saobraćaj između neke dve lokacije putuju istim putem
 - ISP ne mogu da garantuju servis koji prodaju ako ne postoji simetričnost
 - Asimetričan saobraćaj – problemi sa otkrivanjem problema
- Balansiranje saobraćaja je podela saobraćaja preko više alternativnih putanja
 - Jako teško ostvarivo sa BGP protokolom u oba smera – izlaznom i ulaznom u AS

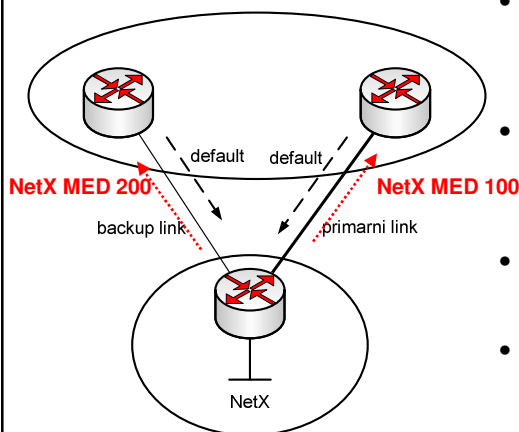
84

Više veza ka jednom provajderu

- Scenariji:
 - Default rute, primarni i backup link
 - Default rute, primarni i backup link i parcijalna ruting tabela
 - Default rute, primarni i backup link i puna ruting tabela

85

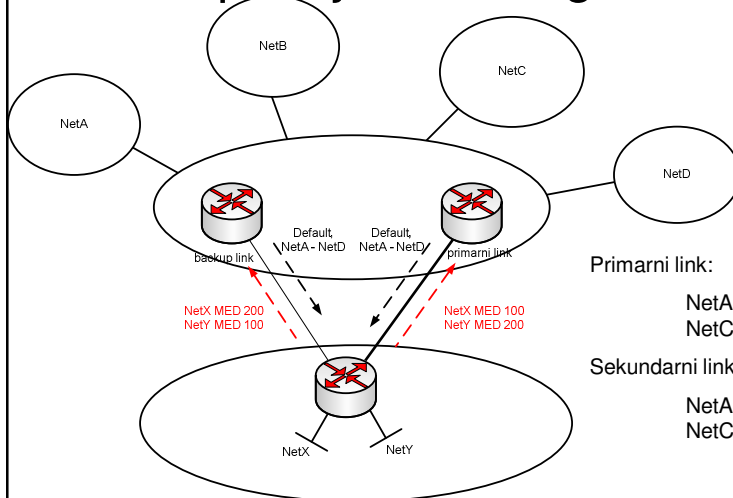
Default rute, primarni i backup link



- Izlazni saobraćaj – prema jednoj od default ruta
- Ulazni saobraćaj – da bi se dobila simetrija – MED
- Sav saobraćaj ide preko primarnog linka
- Sekundarni link se koristi samo u slučaju otkaza primarnog

86

Default rute, primarni i backup link i parcijalna ruting tabela



Primarni link:

NetA, NetB – LP 100,
NetC, NetD – LP 200

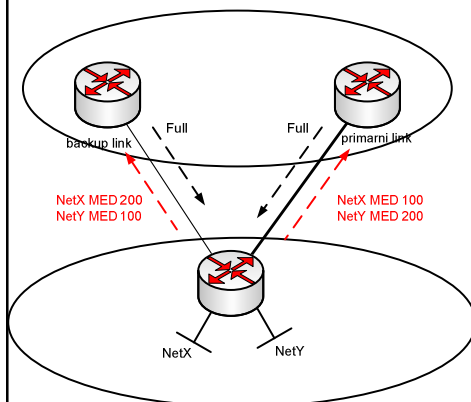
Sekundarni link:

NetA, NetB – LP 200,
NetC, NetD – LP 100

Da li mogu da se parcijalno oglašavaju NetA-NetD?
Kako ide saobraćaj od NetX ka NetC i obrnuto?
Kako ide saobraćaj od NetY ka NetC i obrnuto?

87

Default rute, primarni i backup link i puna ruting tabela



- Ako postoji primarni i backup link, sve rute sa primarnog mogu da dobiju veći LP
- Dolazni saobraćaj može da se balansira korišćenjem MED atributa
- Kako balansirati odlazni saobraćaj?

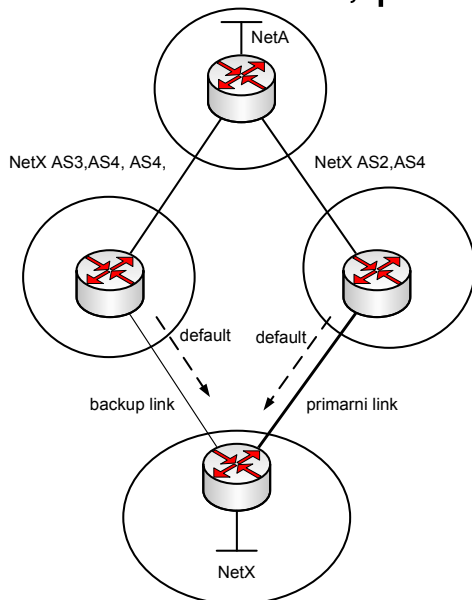
88

Više veza ka više provajdera

- Scenariji
 - Default rute, primarni i backup link
 - Default rute, primarni i backup link i parcijalna ruting tabela
 - Default rute, primarni i backup link i puna ruting tabela

89

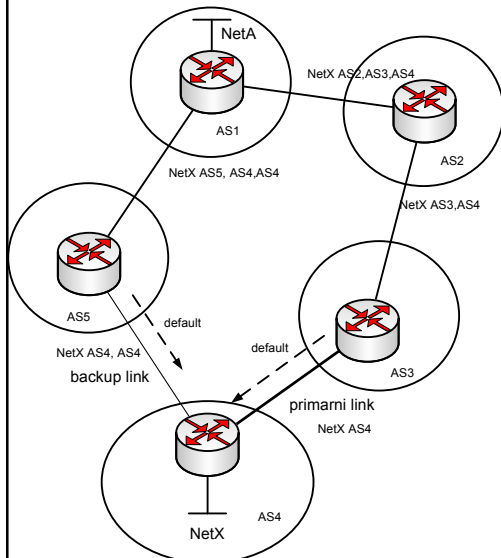
Default rute, primarni i backup link



- Odlazni saobraćaj – prema jednoj od default ruta (primarna)
- Neoptimalno rutiranje prema levom AS
- Dolazni saobraćaj ne može da se reguliše pomoću MED
- Prepend AS - backup AS
- Ekonomski neoptimalno, jer se od ISP zakupljuje veza ka Internetu koja se koristi samo u slučaju kada otkáže primarni link

90

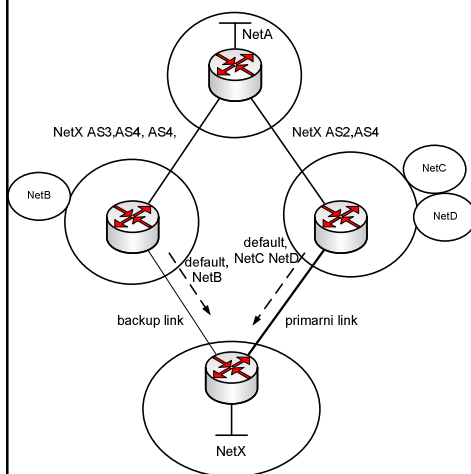
Poznavanje topologije



- Kuda će ići saobraćaj od NetA ka NetX?
- U ovakvim situacijama može da se koristi Community, ukoliko ga podržavaju udaljeni provajderi
- Teško je uticati na raspodelu dolaznog saobraćaja

91

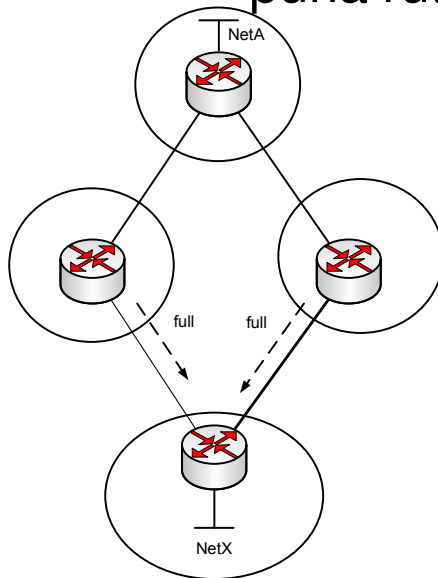
Default rute, primarni i backup link i parcijalna ruting tabela



- Saobraćaj ka mrežama NetB, C i D će biti određen kraćim AS_path
- Rešen problem neoptimalnosti rutiranja ka susednim AS
- Ostalo, kao u prethodnom primeru -većina saobraćaja preko primarnog linka
- Odnos dolaznog saobraćaja preko linkova zavisi od povezanosti ISP AS i broja prepend AS

92

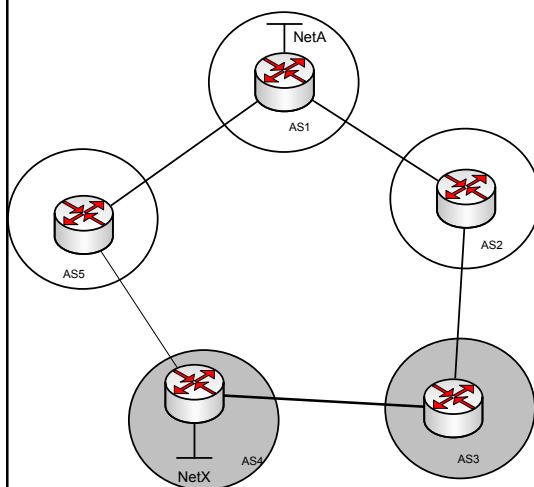
Default rute, primarni i backup link i puna ruting tabela



- Odnos odlaznih saobraćaja zavisi od povezanosti ISP AS
- Load balancing odlaznog saobraćaja moguć kada bi se delovima pune ruting tabele dodeljivali različiti LP ili AS_path
- Kako balansirati dolazni?

93

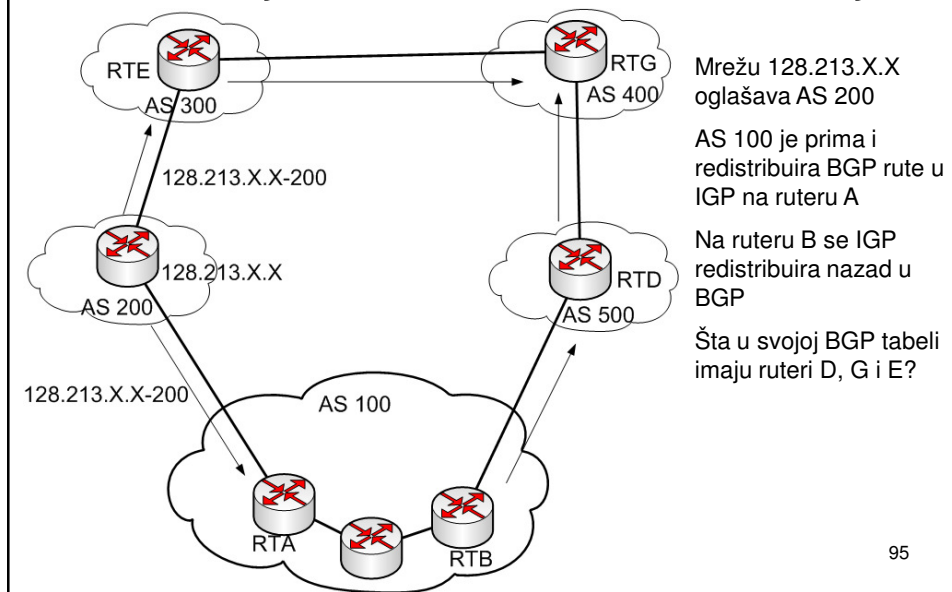
Dve mreže koje pružaju jedna drugoj backup



- Backup link između korisnika treba da prenosi saobraćaj samo u situaciji kada padne jedan od linkova ka provajderima
- LP na ruterima na osnovu AS ili community vrednosti
- AS Path manipulacije

94

Zašto nije dobro vršiti redistribuciju



iBGP – problem skalabilnosti

- iBGP ne prosleđuje drugim iBGP ruterima rute dobijene putem iBGP
- Mora da postoji potpun graf iBGP odnosa unutar AS da bi se omogućio ispravno rutiranje unutar AS
- Ukupan broj iBGP odnosa $\sim n^2$
- U velikim AS to može da predstavlja problem

Route reflector

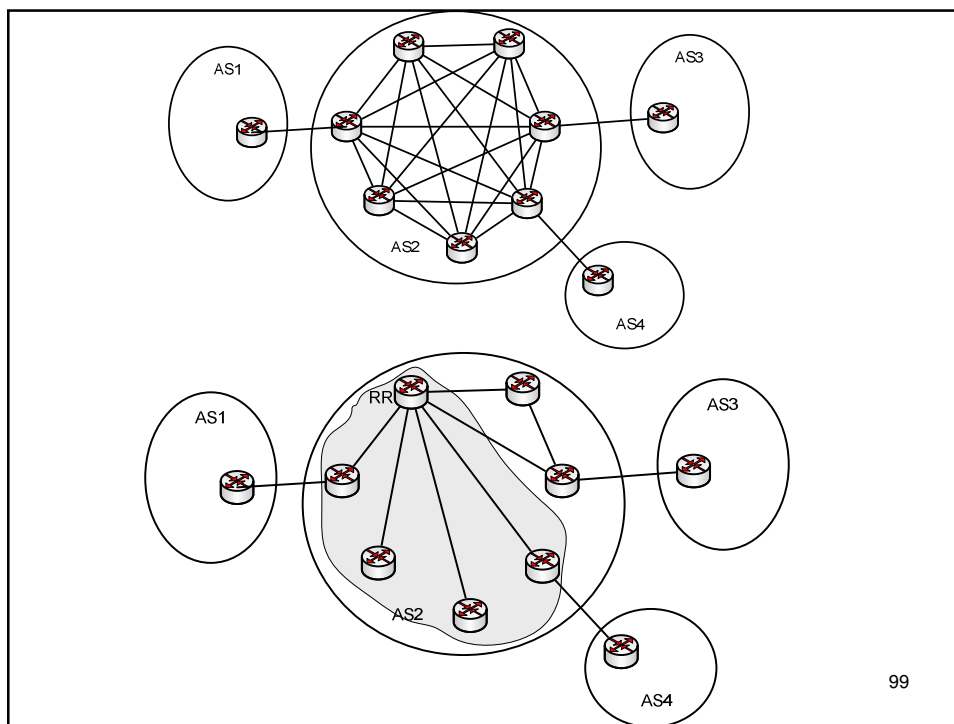
- Rešenje za veliki broj iBGP sesija je “*route reflector*”
- Route reflector je ruter koji krši pravilo ponašanja iBGP rutera – on može da prosledi iBGP rute drugim iBGP susedima
- iBGP ruteri koji koriste usluge route reflectora su klijenti.
- Kada jedan klijent pošalje UPDATE poruku nekom route reflectoru, RR prosleđuje tu poruku drugim njegovim klijentima

97

Route reflector

- Ne moraju svi ruteri u nekom AS da budu ili RR ili klijenti. Neki mogu da budu “obični” ruteri koji iBGP koriste na klasičan način
- RR prosleđuje iBGP rute samo svojim klijentima i iBGP/eBGP susedima.
- RR i njegovi klijenti čine klaster.
- Bilo koji ruter u AS može da bude RR. Izbor zavisi od administratora i performansi rutera

98



99

Pravila prosleđivanja ruta kod RR

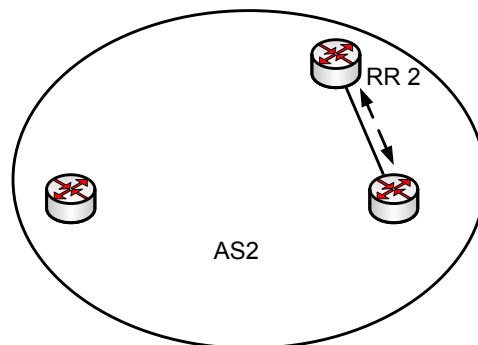
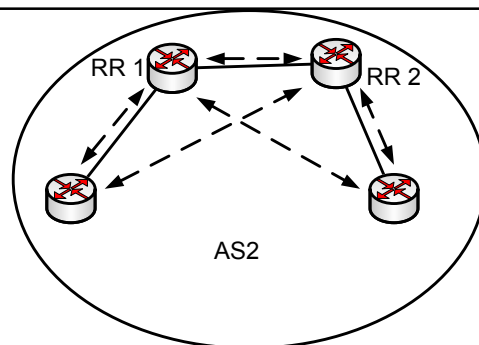
- Ako je ruta dobijena od suseda koji nije klijent datog RR, RR će reflektovati datu rutu samo klijentima
- Ako je ruta dobijena od klijenta, RR će je reflektovati svi ostalim klijentima i svim susedima koji nisu klijenti
- Ako je ruta dobijena od EBGIP suseda, reflektuje se svim klijentima i svim susedima koji nisu klijenti

100

Redundansa i RR

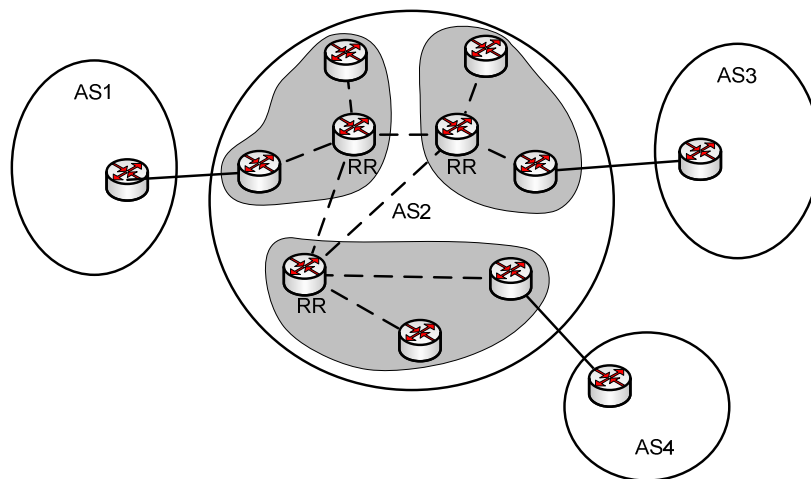
- Otkazom RR i nedostatkom potpunog grafa iBGP sesija rutiranje u AS više ne bi bilo regularno
- Moguće je da postoje dva RR u nekom klasteru čime se povećava pouzdanost mreže
- Izbor RR nije nezavisan od fizičke topologije u mreži

101



102

Hijerarhijska organizacija mreže



103

Route reflector

- RR se ponaša po svim ostalim pravilima ponašanja za iBGP rutere (ne menja Next hop,...)
- RR šalje samo najbolju rutu koju je odredio njegov BGP proces. Ovo dodatno umanjuje zauzeće memorije na ruterima klijentima u poređenju sa potpunim iBGP grafom
- Uvođenje RR otvara mogućnost stvaranje petlji u rutiranju unutar AS: postoji mogućnost da ruta koja je poslata iz nekog klastera se vrati u dati klaster (ne postoji mogućnost provere AS_PATH unutar jednog AS)
- Zbog toga su uvedeni novi atributi: ORIGINATOR_ID i CLUSTER_LIST

104

ORIGINATOR_ID, CLUSTER_LIST

- ORIGINATOR_ID (ONTA) označava Router ID onog rutera koji je poslao datu rutu.
- ORIGINATOR kao atribut dodaje RR.
- CLUSTER_LIST (ONTA) je atribut analogan AS_PATH atributu
- Svaki klaster ima svoju identifikaciju
- CLUSTER_LIST je niz identifikacija klastera unutar jednog AS kroz koje je prošla data ruta

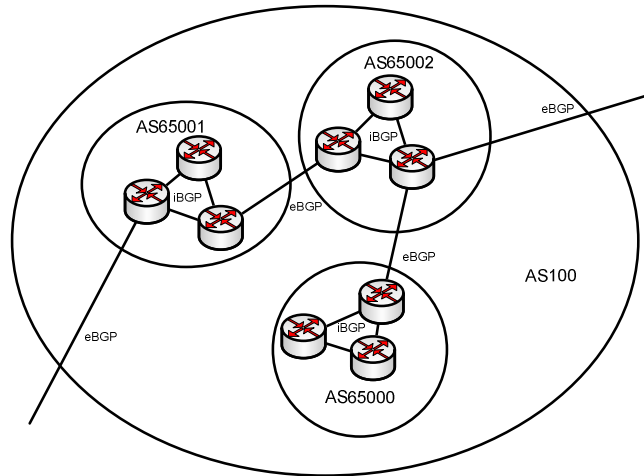
105

Konfederacije

- RFC 1965 -> 3065 -> 5065
- Drugi način za rešavanje problema velikog broja iBGP sesija
- Unutar jednog AS se formira više privatnih pod-AS koji su u konfederaciji (za spoljašnje AS se pojavljuju kao jedinstven AS)
- Pod-AS međusobno komuniciraju putem eBGP, ali je prenos atributa isti kao da je u pitanju iBGP (MED, Local preference,... se prenose između pod-AS)
- Pod-AS dobijaju brojeve za AS iz privatnog skupa AS brojeva.

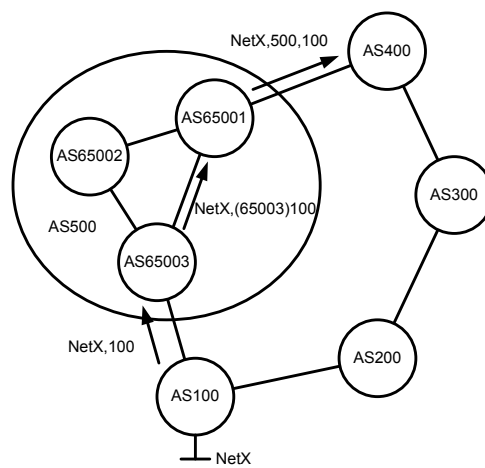
106

Konfederacije



107

Konfederacije



108

Odluke u rutiranju sa Konfederacijama (kriterijum 8.)

- Unutar AS se NE uzima u obzir kraći AS-Path koji se sastoji od pod-AS
- Bez konfederacija BGP u 8. kriterijumu bira eBGP ispred iBGP ruta
- Ukoliko postoji ruta ka nekoj mreži dobijena iz susednog pod-AS i ruta dobijena od eksternog AS, BGP će odabrati putanju ka eksternom AS, iako su obe eBGP
- Ukoliko postoji ruta ka nekoj mreži dobijena od iBGP (unutar pod-AS) i ruta dobijena od susednog pod-AS (eBGP unutar konfederacije), odabraće se ona ruta koja vodi van datog pod-AS (eBGP ruta van konfederacije)

109

Očuvanje stabilnosti Interneta

- Česte promene ruta koje oglašava neki BGP speaker se propagiraju po celom Internetu. (Update/Withdraw)
- To pravi nepotreban saobraćaj na mreži i opterećuje procesore rutera
- Da bi se Internet zaštitio postoji mehanizam "route flap damping"

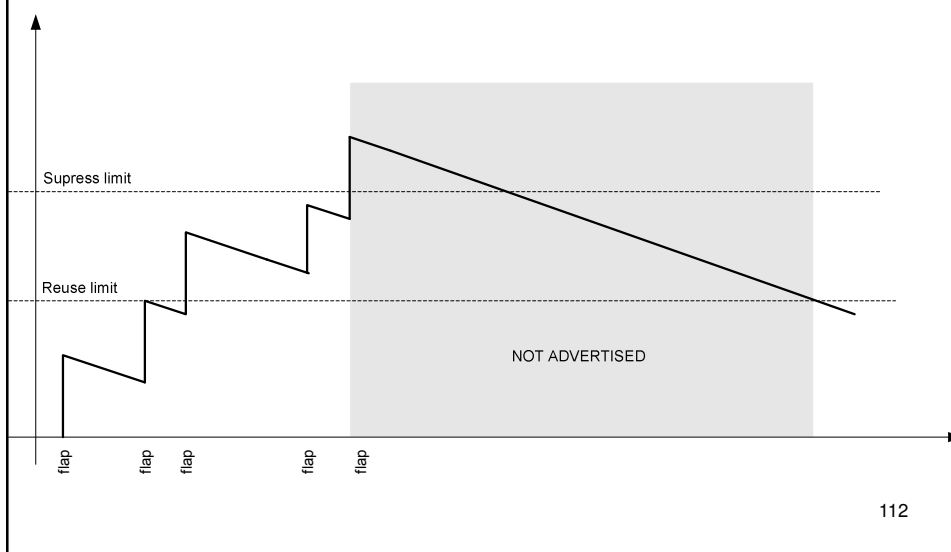
110

Route flap damping – RFC 2439

- Ruter svakoj ruti dodeljuje Penalty vrednost koja je inicijalno 0
- Kada se desi promena (route flap) date rute Penalty se povećava za određenu vrednost
- Kada nema promena, Penalty se smanjuje, tako da se za definisano vreme *half_life* se smanji na polovinu početne vrednosti
- Kada Penalty pređe *Supress-limit*, data ruta se više se ne oglašava
- Kada Penalty padne ispod *Reuse-limit*, data ruta se seponovo oglašava

111

Route flap damping



112

Kontrola ruta koje se dobijaju preko Interneta

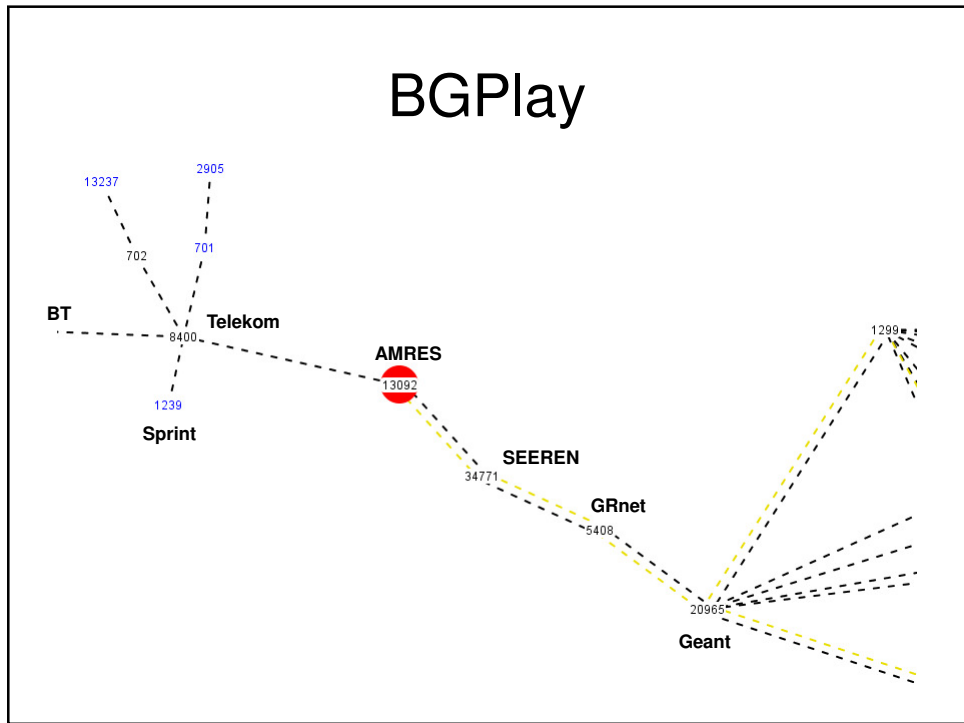
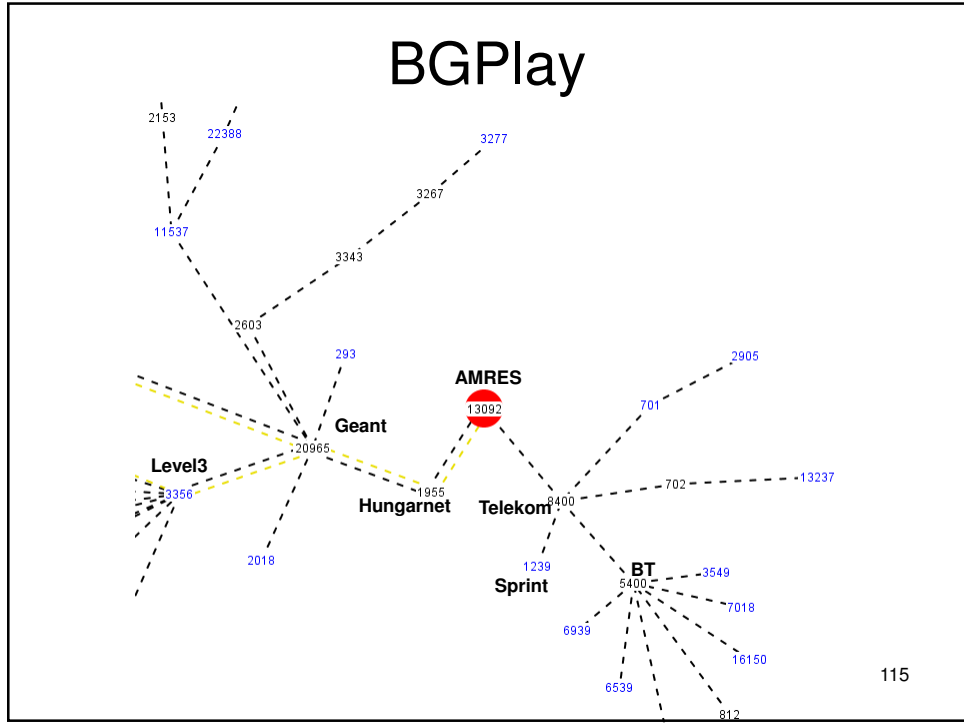
- Rute koje se oglašavaju putem BGP moraju da se pre toga unesu u bazu RIR u formi route-object-a
- ISP jednom dnevno proveravaju route-object bazu i u skladu sa njom formiraju filtre za dolazeće rute

route: 147.91.0.0/16	route6: 2001:4170::/32
descr: UNIVERSITY OF BELGRADE	descr: UNIVERSITY OF BELGRADE
origin: AS13092	origin: AS13092
mnt-by: UB-MNT	mnt-by: UB-MNT
source: RIPE # Filtered	source: RIPE # Filtered

113

Pomoćni alati za rad sa

- Looking glass
 - www.traceroute.org
- BGPlay



The screenshot shows the BT Global Services Looking Glass interface. The main content area displays the following information:

Query:	BGP Network
Router:	Austria - 12a1.at-vie
Address:	147.91.0.0

BGP routing table entry for 147.91.0.0/17, version 27825410
 Paths: (2 available, best #2)
 Multipath: eBGP
 Advertised to update-groups:
 1
 8400 8400 13092 13092 13092 13092 13092 13092 13092 13092 13092
 166.49.166.73 (metric 156) from 166.49.166.32 (166.49.166.32)
 Origin IGP, metric 0, localpref 190, valid, internal
 Community: 5400:49
 Originator: 166.49.166.73, Cluster list: 166.49.166.32, 166.49.166.64
 8400 8400 13092 13092 13092 13092 13092 13092 13092 13092 13092
 166.49.166.73 (metric 156) from 166.49.166.65 (166.49.166.65)
 Origin IGP, metric 0, localpref 190, valid, internal, best
 Community: 5400:49
 Originator: 166.49.166.73, Cluster list: 166.49.166.65

© 2006 British Telecommunications plc

117

Multiprotokolarne ekstenzije BGP protokola – RFC 2858 -> 4760

- Originalni BGPv4 je protokol je mogao da razmenjuje samo IPv4 rute
- Drugi protokoli (IPX, IPv6,...) nisu mogli da komuniciraju koristeći globalnu Internet mrežu
- RFC 2283 -> 2858 -> 4760 – ekstenzije BGP protokola koje će da omoguće prenos ruta različitih protokola mrežnog sloja

118

MBGP

- Atributi i argumenti koji su striktno vezani za IPv4:
 - Next_hop (značajan samo za nove rute, a ne i za rute koje se brišu)
 - NLRI
 - Aggregator
- Novi atributi:
 - Multiprotocol_Reachable_NLRI (MP_REACH_NLRI)
 - Multiprotocol_Unreachable_NLRI (MP_UNREACH_NLRI)

119

MP_REACH_NLRI (ONTA)

```
+-----+
| Address Family Identifier (2 octets) |
+-----+
| Subsequent Address Family Identifier (1 octet) |
+-----+
| Length of Next Hop Network Address (1 octet) |
+-----+
| Network Address of Next Hop (variable) |
+-----+
| Number of SNPAs (1 octet) |
+-----+
| Length of first SNPA(1 octet) |
+-----+
| First SNPA (variable) |
+-----+
| Length of second SNPA (1 octet) |
+-----+
| Second SNPA (variable) |
+-----+
| ... |
+-----+
| Length of Last SNPA (1 octet) |
+-----+
| Last SNPA (variable) |
+-----+
| Network Layer Reachability Information (variable) |
+-----+
```

120

MP_REACH_NLRI (ONTA)

- Address Family Identifier: Vrsta protokola mrežnog sloja za koji se šalju rute (1=IPv4, 2=IPv6)
- Subsequent Address Family Identifier: Vrsta adrese definisane pomoću NLRI (1=unicast, 2=multicast, 4=label, 127=VPN)
- Length of Next Hop Network Address: 4=32bita – IPv4, 16=128bita – IPv6 ili 32 – 2 IPv6 adrese (link-local i globalna)
- Network Address of Next Hop: Adresa sledećeg rutera ka destinaciji
- SNPA – Subnetwork Point of Attachment – Layer 2 adresa interfejsa kroz koji se dolazi do BGP suseda (da bi se izbegao ARP)
 - Number of SNPAs - Ako je upisana 0, nema ni jedno polje vezano za SNPA.
 - Length of Nth SNPA
 - SNPA of Next Hop
- Network Layer Reachability Information: NLRI u istom formatu kao za originalnu verziju BGP

121

MP_UNREACH_NLRI (ONTA)

```
+-----+
| Address Family Identifier (2 octets) |
+-----+
| Subsequent Address Family Identifier (1 octet) |
+-----+
| Withdrawn Routes (variable) |
+-----+
```

- Rute koje se više ne oglašavaju

122

MP-BGP primene

- IPv6 rutiranje (RFC 2545)
- Inter-domain Multicast
- IPv4 VPN
- IPv6 tranzicija
- MPLS distribucija labela
- S-BGP
- QoS ekstenzija za BGP

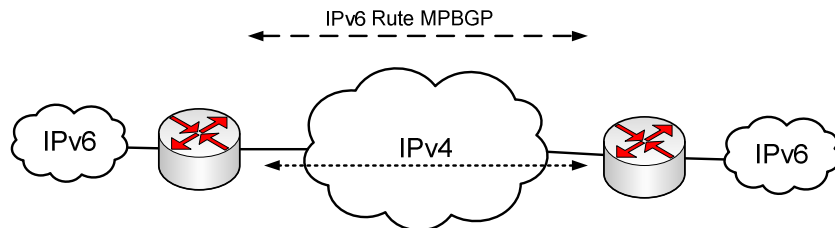
123

MP-BGP i IPv6

- TCP sesija između BGP speakera može da se ostvari bilo preko IPv4 bilo preko IPv6, bilo preko oba istovremeno (u dual stack režimu)
- Moguće je da se preko TCP BGP sesije koja je uspostavljena preko IPv4 prenose informacije o IPv6 rutama i obrnuto
- Ako se peering ostvaruje putem IPv6 na istom subnetu, onda se kao Next Hop upisuju i link-local i globalne IPv6 adrese datih interfejsa
- Šta se dešava sa Next Hop atributom ako se koriste link-local IPv6 adrese?

124

Mehanizmi povezivanja IPv6 ostrva na IPv4 mrežu



- Saobraćaj mora da se tuneluje, na primer:
 - 6to4 (**207.142.131.202** -> 2002:**CF8E:83CA**::/48)
 - ISATAP (**192.0.2.143** -> fe80::5efe:**c000:028f**)
 - MPLS (6PE)