### Internet tehnologije

# BGP - Border Gateway Protocol - drugi deo -

#### Predavač:

Prof. dr Slavko Gajin, slavko.gajin@rcub.bg.ac.rs

#### **Asistent:**

Zlatko Dejanović, zlatko.dejanovic@etfbl.net

### **BGP** atributi

- Rute (NLRI) se između BGP suseda prenose u UPDATE porukama
- Osim IP adrese i maske mreže, rute dodatno sadrže pojedine atribute
- Atributi se dele u 4 grupe:
  - Dobro poznati moraju biti podržani na uređaju
    - obavezni (Well-known Mandatory)
    - neobavezni (Well-known Discretionary)
  - Opcioni ne moraju biti podržani na uređaju
    - prenosivi (Optional Transitive)
    - neprenosivi (Optional Nontransitive)
- Atributi utiču na izbor najbolje rute (uvek samo jedna ruta)
- Konfiguracija BGP-a podrazumeva upravljanje rutiranjem, tako što se manipuliše atributima na pojedinačnim rutama:
  - Prihvatanje ruta u zavisnosti od vrednosti pojedinih atributa
  - Filtriranje ruta na osnovu pojedinih atribura
  - Izmena pojedinih atributa prilikom slanja ili primanja ruta

# BGP atributi – dobro poznati

### Dobro poznati obavezni (Well-known Mandatory - WMA)

- Moraju da postoje u svakoj BGP Update poruci
- pridružen odgovarajućoj ruti
- Sve implementacije BGP-a moraju da ga prepoznaju
- Nedostatak ovih atributa u Update poruci generiše grešku (notification)
- može da se modifikuje (npr. AS-PATH)
- uvek su prenosivi u druge AS-ove tranzitivni

### Dobro poznati neobavezni (Well-known discretionary - WDA)

- Atribut koji prepoznaju sve BGP implementacije i u skladu sa njim se ponašaju
- ne mora da bude pridružen nekom ruti
- mogu biti i prenosivi i neprenosivi (iz jednog AS-a u drugi AS)

# BGP atributi – opcioni

### Opcioni prenosivi (Optional transitive - OTA)

- Atribut koji ne moraju da prepoznaju sve BGP implementacije i da se ponašaju u skladu sa njim (opcioni)
- Ako ruter dobije ovakav atribut koji ne prepoznaje, onda treba da ga prosledi ostalim BGP susedima (prenosivost), čak i kada ga ne prepoznaje

#### Opcioni neprenosivi (Optional nontransitive - ONTA)

- Atribut koji ne moraju da prepoznaju sve BGP implementacije i da se ponašaju u skladu sa njim (opcioni)
- Ako ruter dobije ovakav atribut ne prosleđuje dalje

# Uobičajeni BGP atributi

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

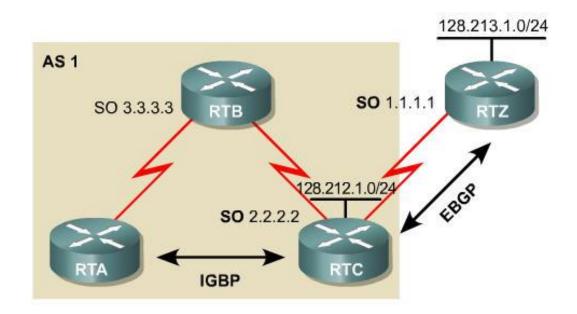
# Next hop atribut

- Dobro poznat obavezan atribut (WMA)
- Next hop IP adresa rutera koji vodi ka oglašenoj ruti (obično adresa rutera koji je oglasio rutu)
- Next hop ne mora nužno da bude susedni ruter, odnosno na direktno povezanom mrežnom segmentu
- Pravila kako BGP koristi next hop atribut:
  - U EBGP sesijama, next hop je IP adresa EBGP suseda koji je oglasio datu rutu.
  - U IBGP sesijama, ako su rute oglašene unutar samog AS, next hop je IP adresa rutera unutar AS koji je oglasio 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 (iz drugog AS-a)
- Ruter mora da ima u svojoj ruting tabeli rutu ka IP adresi označenoj u Next hop atributu, inače ruta neće biti ubačena u ruting tabelu (da bi mogao da prosledi saobraćaj)

## Next Hop atribut

#### Rurter RTA - naučio je rute preko IBGP-a:

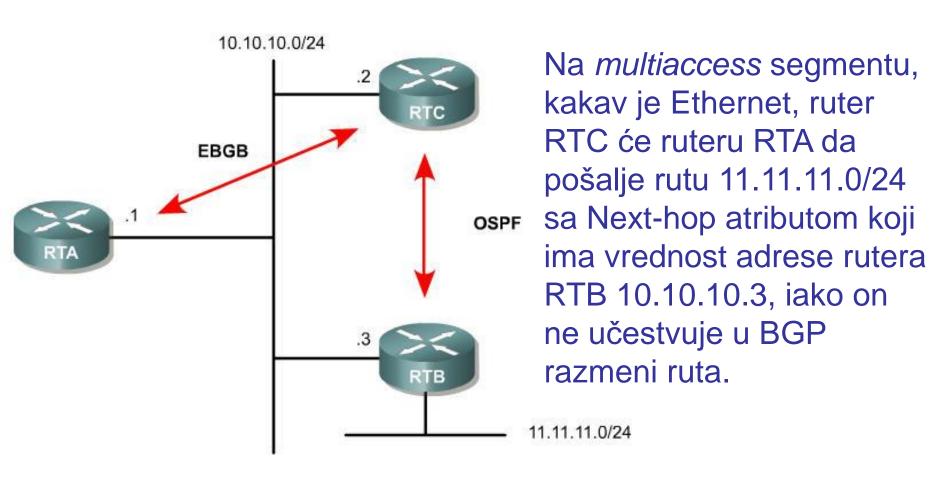
- next-hop za 128.213.1.0/24 je 1.1.1.1 (ruta je ušla AS preko EBGP)
- next-hop za 128.212.1.0/24 je 2.2.2.2 (ruta pripada AS-u)
- u ruting tabeli mora da ima rute za adrese 1.1.1.1 i 2.2.2.2 da pi prosledio pakete na mreže 128.213.1.0/24 i 128.212.1.0/24



Can reach 128.213.1.0 by way of 1.1.1.1. and can reach 128.212.1.0 by way of 2.2.2.2

Can reach 128.213.1.0 by way of 1.1.1.1

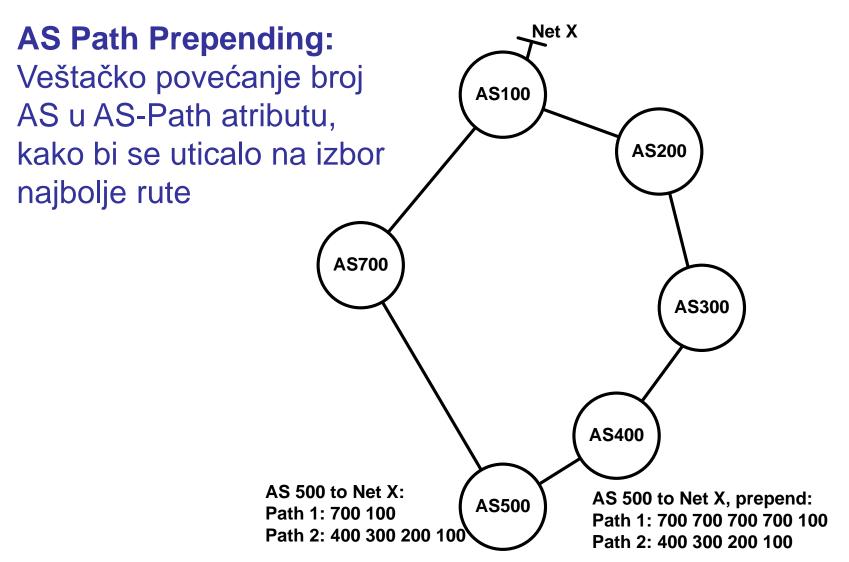
# Next Hop na Multiaccess segmentima



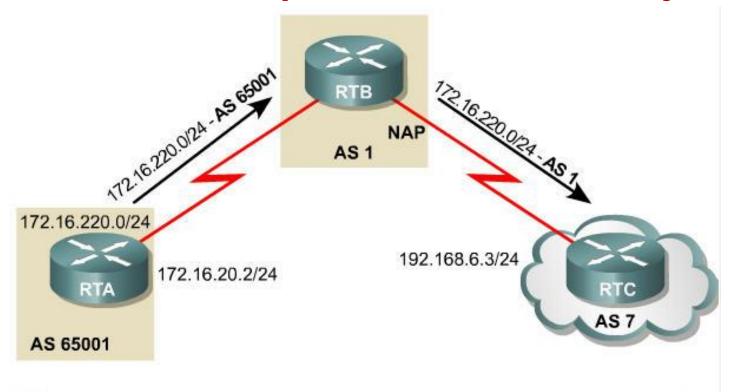
## AS\_Path Atribut

- Dobro poznat obavezan atribut (WMA)
- •AS koji oglašava neku rutu dodaje broj svog AS u ASpath atribut pridružen datoj ruti
- •Svaki sledeći AS dodaje (**prepend**) broj svog AS datoj ruti prilikom prosleđivanja narednom AS
- Detektuje i sprečava eventualne petlje u oglašavanju i rutiranju
  - Ukoliko ruter prepozna broj svog AS u AS Path atributu, ruta se odbacuje
- Dužina AS\_path osnovni kriterijum izbora najbolje putanje
  - Kraći AS Path označava rutu sa boljom metrikom koja će biti ubačena u ruting tabelu.

# AS Path prepending



# 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.

# Origin Atribut

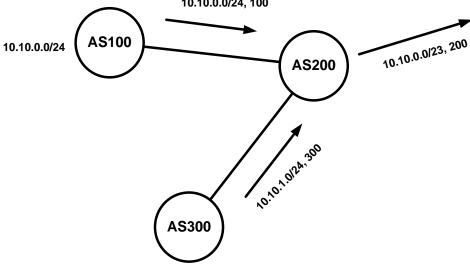
- Dobro poznat obavezan atribut (WMA)
- Origin atribut govori o poreklu rute/prefiksa
- Koristi se u izboru najbolje rute
- Vrste origin atributa:
  - IGP (0) Prefiks je dobijen direktnim oglašavanjem u konfiguraciji BGP protokola (network komanda)
  - EGP (1) Prefiks je dobijen iz EGP (prethodna verzija eksternog protokola), danas se ne koristi
  - Incomplete (2) Prefiks je dobijen redistribucijom u BPG
- Veći prioritet imaju rute sa manjom vrednosti Origin atributa

# Atomic Aggregate atribut

- Dobro poznat neobavezan 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 (ruta) agregiran od više različitih prefiksa.

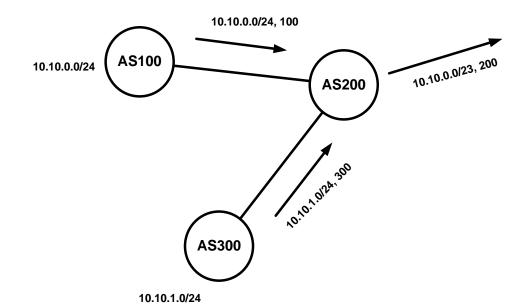
 Ruter koji je poslao prefiks sa Atomic Aggregate atributom sa vrednošću True je izvršio agregaciju ruta i ima specifičnije rute do destinacija

10.10.1.0/24



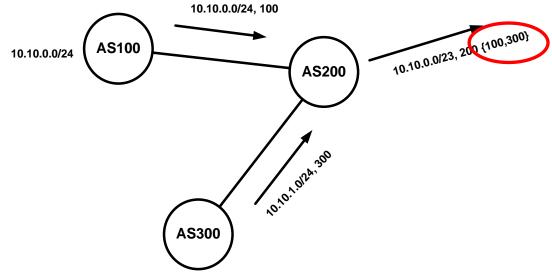
# Aggregator atribut

- Dobro poznat neobavezan atribut (WDA)
- Ovim atributom se označava onaj ruter koji izvršio agregaciju ruta.
- Kao argument ovog atributa upisuje se Router ID rutera koji je izvršio agregaciju



### **AS-Set atribut**

- Dobro poznat neobavezan atribut (WDA)
- Korišćenjem agregacije se smanjuje broj ruta u tabeli rutiranja, ali se gube neke informacije.
- AS-Set posebna vrsta AS-Path atributa, kojim se zadržavaju informacije o agregiranim rutama.
- AS Set se sastoji od agregirane rute i elemenata koji je sačinjavaju (AS brojevi pre agregacije)
- Čuva informacije o AS brojevima radi sprečavanja petlji
- Nije dobro da se koristi kada se agregira veliki broj ruta



### Local Preference Atribut

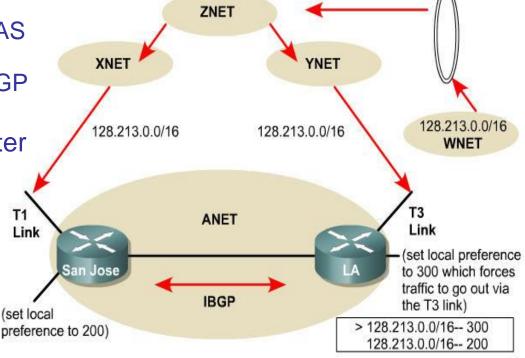
- Dobro poznat neobavezan atribut (WDA)
- Označava prioritet rute koja je ušla u AS
  - Utiče na saobraćaj koji će da izlazi iz datog AS!!!
- Veća vrednost ima prednost u ruting tabelu se ubacuje ruta koja ima viši Local Preference

Lokalan za jedan AS

 Razmenjuje unutar jednog AS putem IBGP, ali se ne prenosi putem EBGP (nije tranzitivan)

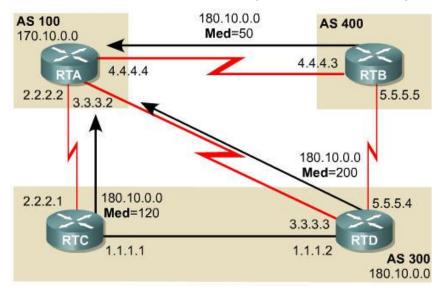
Dodeljuje ruti na ulasku u ruter

Deafult vrednost je 100



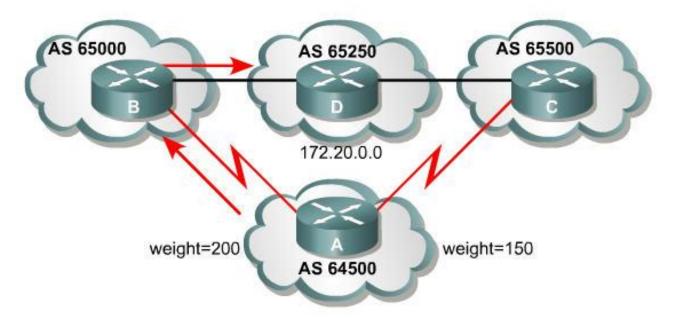
# Multiple Exit Discriminator (MED) Atribut

- Opcioni netranzitivni atribut (ONTA)
  - atribut koji uđe u AS ne napušta ga
- MED se dodeljuje rutama koje izlaze iz datog rutera
  - Kontroliše dolazni saobraćaj u AS
  - obaveštava susedne AS o željenoj putanji saobraćaja u dati AS ukoliko dati AS ima više veza sa drugim AS-om
- Namera je bila da se odslikava metrika IGP iz jednog AS i prenosi u drugi
- Niža vrednost ima prednost u odnosu na viši
- Jedan AS poredi MED vrednosti samo za prefikse dobijene iz jednog AS



## Weight Atribut

- Sličan Local Preference atributu
- Utiče na na rute koje ulaze u AS, odnosno saobraćaj koji izlazi iz AS
- Veća vrednost ima prednost nad manjom
- Lokalan je za ruter i ne razmenjuje se sa drugim ruterima
- Nije po BGP standardu, ali ga pojedini proizvođači omogućavaja na svojim ruterima (Cisco)



# Proces izbora najbolje rute u BGP protokolu rutiranja

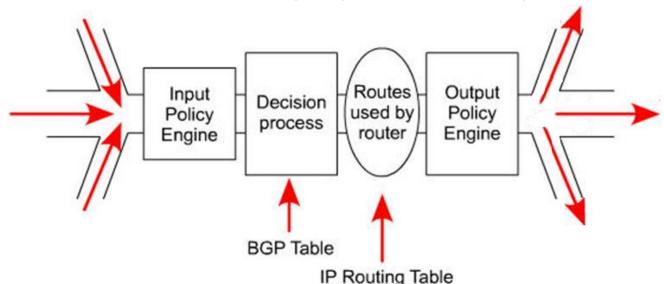
- 1. Ako Next Hop atribut za datu rutu ne postoji u ruting tabeli, ruta se ignoriše tj. ne ubacuje u ruting tabelu (sinhronizacija ruta)
- 2. Cisco: ako postoji Weight atribut, u ruting tabelu ulazi ruta sa **najvećom** Weight vrednošću
- 3. Bira se ruta sa **najvećom** vrednošću Local Preference
- 4. Bira se ruta sa **kraćim** AS-Path-om
- 5. Bira se ruta sa **nižom** vrednošću Origin atributa (**IGP < EGP < Incomplete**)
- 6. Bira se rutu sa **nižom** MED vrednošću
- 7. Prednost se daje rutama koje su dobijene preko eBGP u odnosu rute dobijene preko iBGP

## Proces izbora najbolje rute

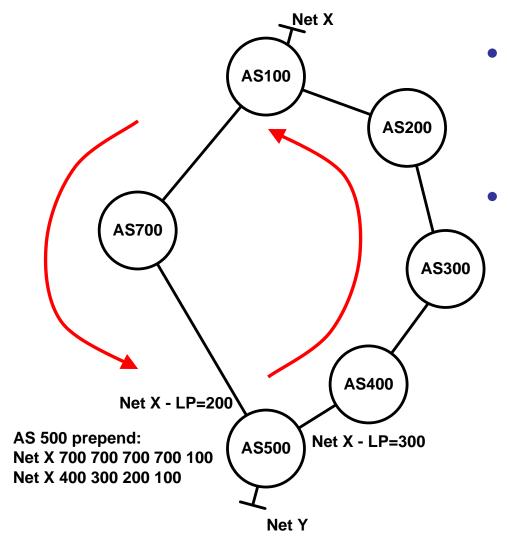
- 8. Bira se ruta čija je IGP metrika do BGP Next hopa niža
- 9. Bira se ruta koja je dobijena ranije (prva koja je stigla u ruter)
- 10. Bira se ruta dobijena od rutera sa nižim Router ID-em
- 11. Bira se putanja sa nižom vrednošću dužine klastera
- 12. Bira se ruta dobijena od suseda sa nižom adresom

#### **ZAKLJUČAK:**

Uvek će biti odabrana JEDNA najbolja ruta ka datoj mreži



### Primer



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

# Atribut Community (OTA)

- Služi za grupisanje mreža za koje se traži određeni način procesiranja od nekog nadređenog AS
- 4-bajtni parametar
- Neke predefinisane vrednosti
  - 0xFFFFFF01 No export ne oglasiti eBGP susedima
  - 0xFFFFFF02 No advertise ne oglasiti nikome
  - Uobičajen način predstavljanja: AS:COMMUNITY

# Redundansa, simetrija i balansiranje saobraćaja

#### Redundansa

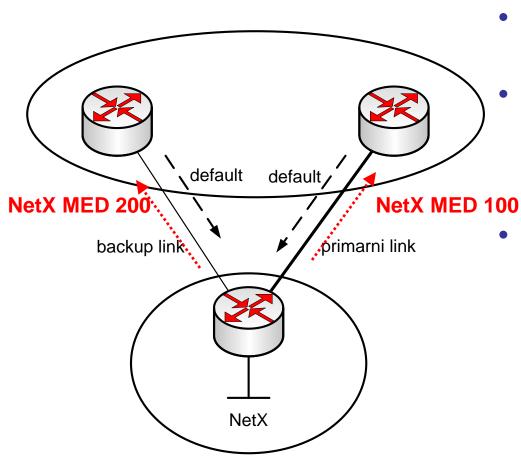
- Povećanje pouzdanosti kroz obezbeđivanje alternativnih putanja
- Pošto je uslov za dobijanje AS veza ka dva druga AS, redundansa uvek postoji
- Simetričnost saobraćaja
  - ulazni i izlazni saobraćaj između neke dve lokacije putuju istim putem
  - Jednostavnije je garantovati servis koji ISP prodaju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

# 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

## Default rute, primarni i backup link

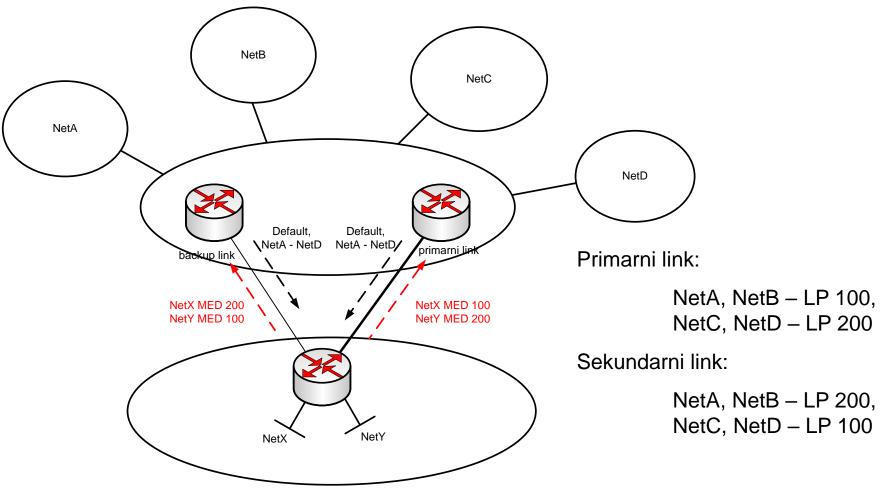


- Izlazni saobraćaj prema jednoj od default ruta
- Ulazni saobraćaj da bi se dobila simetrija
  - manji MED na primarnom linku

#### Rezultat:

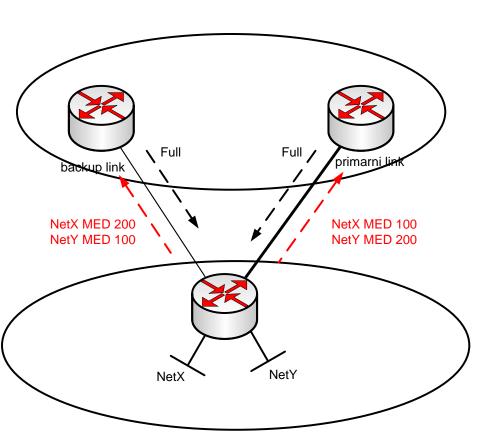
- Sav saobraćaj ide preko primarnog linka
- Sekundarni link se koristi samo u slučaju otkaza primarnog
- Neoptimalno...

# Default rute, primarni i backup link i parcijalna ruting tabela



Kako ide saobraćaj od NetX ka NetC i obrnuto? Kako ide saobraćaj od NetY ka NetC i obrnuto?

# Default rute, primarni i backup link i puna ruting tabela



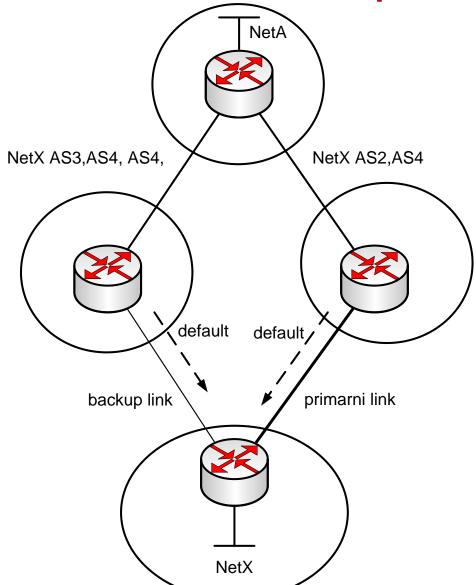
- Ako postoji primarni i backup link, sve rute sa primarnog mogu da dobiju veći LP
- Dolazni saobraćaj može da se balansira korišćenjem MED atributa
- Kako balansirati odlazni saobraćaj?

# 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

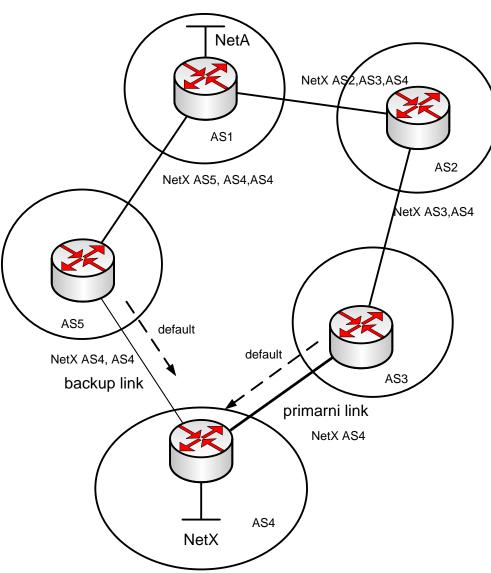
## Default rute, primarni i backup link



- Odlazni saobraćaj prema jednoj od default ruta (primarna)
- Neoptimalno rutiranje prema levom AS
- Dolazni saobraćaj ne može da se reguliše pomoću MED (jer je nije tranzitivan)
- Prepend AS na backup linku
  - postavalja se AS prepend pri oglašavanju naših mreža ruta susedima
- Ekonomski neoptimalno, jer se od ISP zakupljuje backup veza ka Internetu, koja se koristi samo u slučaju kada otkaže primarni link

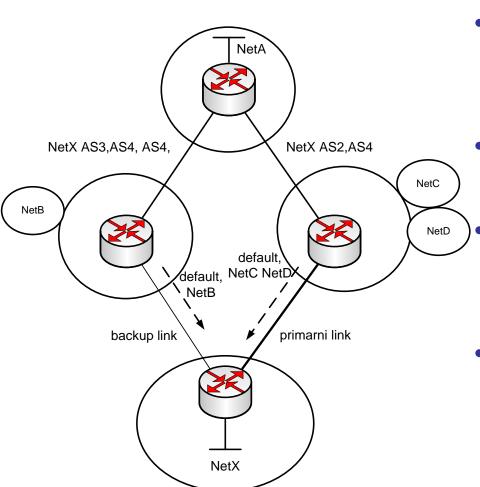
29

# Poznavanje topologije



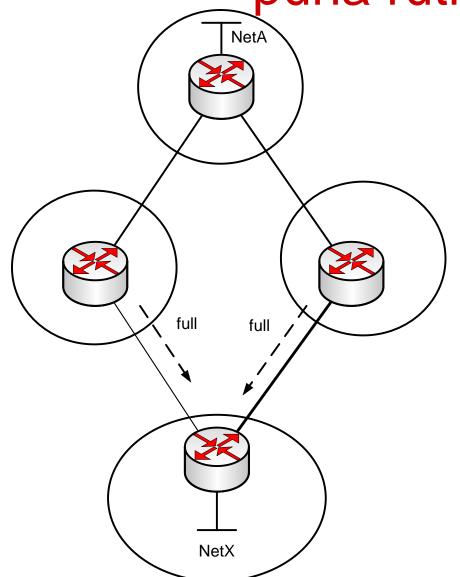
- Kuda će ići saobraćaj od NetA ka NetX?
- U ovakvim situacijama može da se koristi Community, ukoliko ga podržavaju udaljeni provajderi
- Teško je uticati na raspodelu dolaznog saobraćaja
  - postavalja se AS prepend pri oglašavanju naših mreža ruta susedima

# Default rute, primarni i backup link i parcijalna ruting tabela



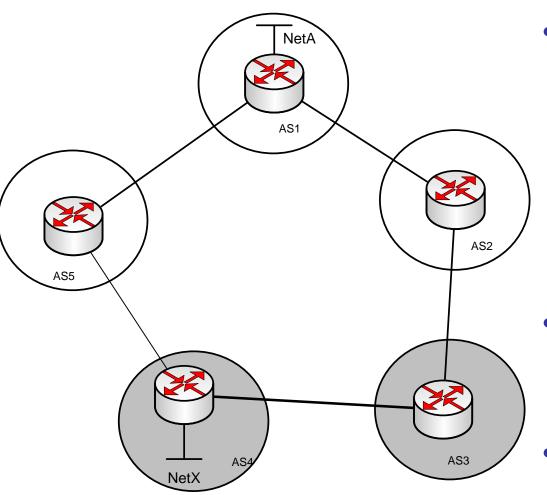
- Saobraćaj ka mrežama NetB,
   C i D će biti određen kraćim
   AS\_path
- Rešen problem neoptimalnosti rutiranja ka susednim AS
- Ostalo, kao u prethodnom primeru -većina saobraćaja preko primarnog linka
- Odnos dolaznog saobraćaja preko linkova zavisi od povezanosti ISP AS i broja prepend AS

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
- Dolazni saobraćaj se može regulisati preko prependa pri oglašavanju naših mreža
  - stepen balansiranja zavisi od topologije povezanosti drugih provajdera

# Dve mreže koje pružaju jedna drugoj backup



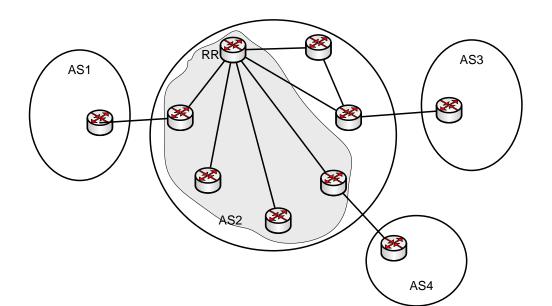
- Backup link između korisnika treba da prenosi saobraćaj samo u situaciji kada padne jedan od linkova ka provajderima
- LP na ruterima na osnovu AS ili comunity vrednosti
- AS Path manipulacije

# 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 ~ n²
- U velikim AS to može da predstavlja problem

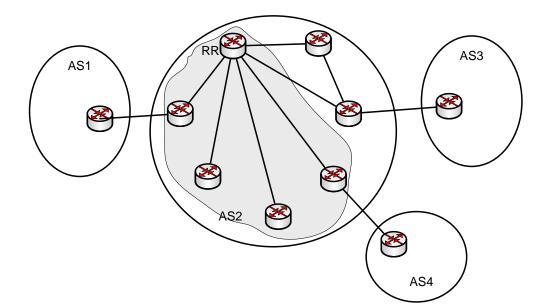
### Route reflector

- 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.
- RR i njegovi klijenti čine klaster.
- Kada jedan klijent pošalje UPDATE poruku nekom route reflectoru, RR prosleđuje tu poruku drugim njegovim klijentima



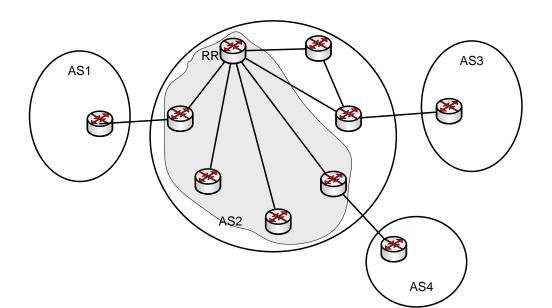
### Route reflector

- 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.
- Bilo koji ruter u AS može da bude RR. Izbor zavisi od administratora i performansi rutera



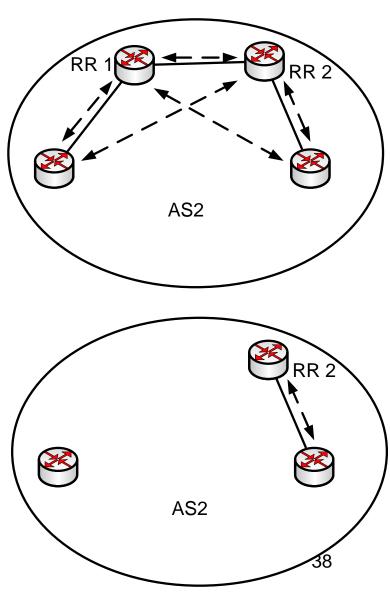
#### Pravila prosleđivanja ruta kod RR

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



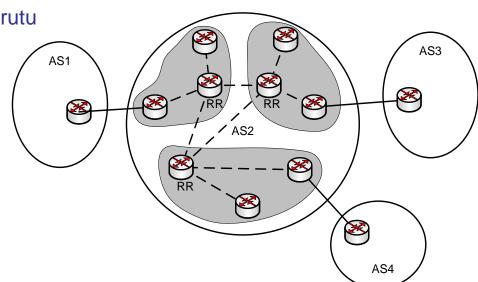
#### Redundansa i RR

- 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



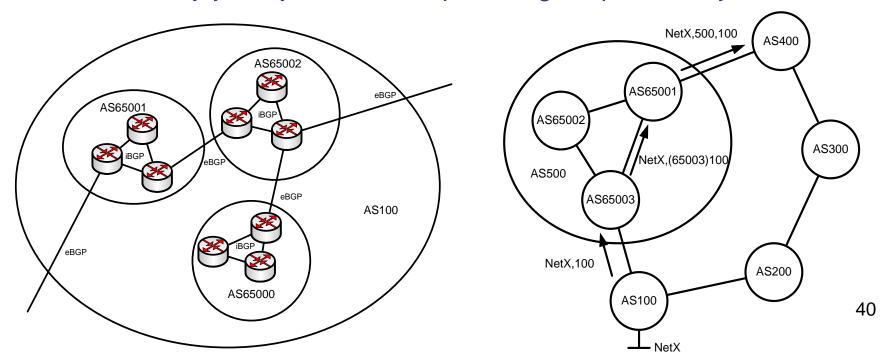
### Hijerarhijska organizacija mreže

- 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
- Problem mogućnost stvaranje petlji unutar AS:
  - postoji mogućnost da se ruta koja je poslata iz nekog klastera vrati u dati klaster (ne postoji mogućnost provere AS\_PATH unutar jednog AS)
- Zbog toga su uvedeni novi atributi: ORIGINATOR\_ID i CLUSTER\_LIST
- ORIGINATOR\_ID
  - Router ID onog rutera koji je poslao rutu
- CLUSTER\_LIST
  - niz identifikacija klastera unutar jednog AS kroz koje je prošla data ruta



#### Konfederacije

- 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 i Local preference i drugi netrazitivni atributi se prenose između pod-AS
- Pod-AS dobijaju brojeve za AS iz privatnog skupa AS brojeva.



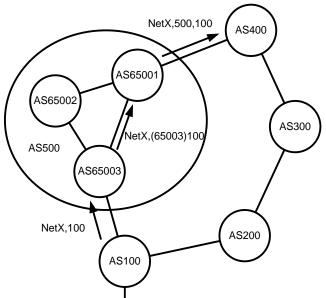
# Proces izbora najbolje rute sa konfederacijama

- Bez konfederacija BGP u 8. kriterijumu izbora ruta bira eBGP ruta ispred iBGP ruta
- Sa konfederacijama BGP odlučuje na sledeći način:

 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 unutar konfederacije ili eBGP od drugog AS a)

drugog AS-a)

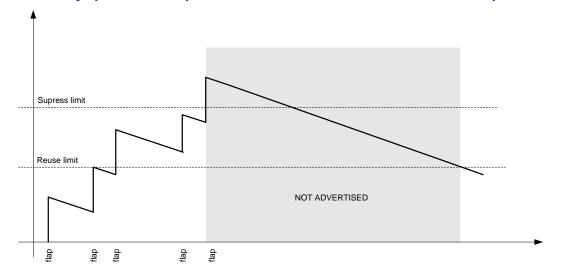


#### Očuvanje stabilnosti Interneta

- 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"

## Route flap damping – RFC 2439

- Česte promene ruta se propagiraju po celom Internetu, što pravi nepotreban saobraćaj i opterećuje procesore rutera na Internetu
- "Route Flap Damping" zaštita od česti promena
  - 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 ne oglašava
  - Kada Penalty padne ispod Reuse-limit, data ruta seponovo oglašava



# Multiprotokolarne ekstenzije BGP protokola – RFC 2858

- 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, RFC 2858 ekstenzije BGP protokola koje će da omoguće prenos ruta različitih protokola mrežnog sloja
- 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)

#### 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

descr: UNIVERSITY OF BELGRADE

origin: AS13092

mnt-by: <u>UB-MNT</u>

source: RIPE # Filtered

route6: 2001:4170::/32

descr: UNIVERSITY OF BELGRADE

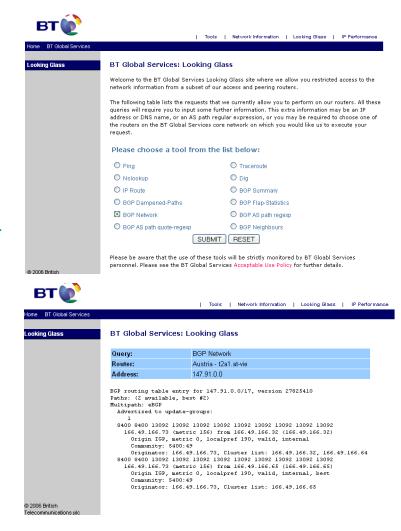
origin: AS13092

mnt-by: UB-MNT

source: RIPE # Filtered

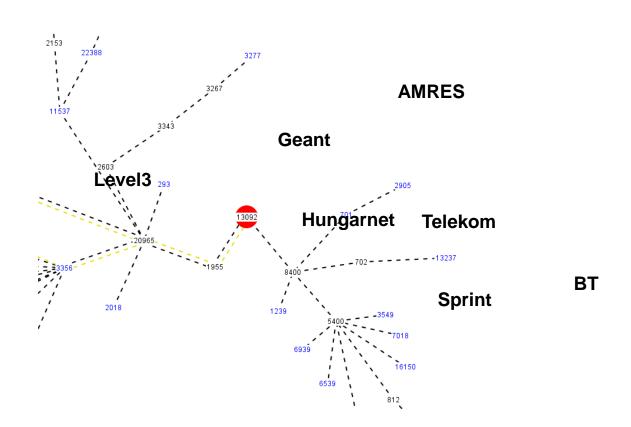
## Pomoćni alati - Looking glass

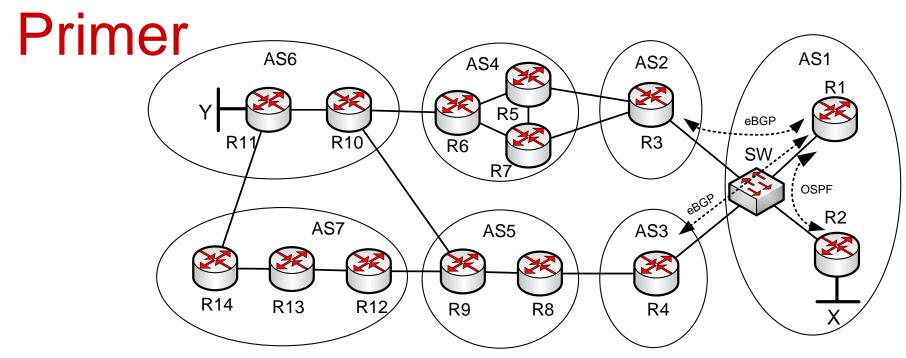
- Looking glass http://traceroute.org/#Looking%20Glass
  - <a href="http://bgp.he.net/">http://bgp.he.net/</a>
  - <a href="http://integra.net/lg/">http://integra.net/lg/</a>
  - http://lg.telekom.rs
  - http://lg.eastlink.ca/
  - http://lg.enta.net/
  - http://lg.df.ru/
  - http://merry.netsys.more.net/lg/index.cgi
  - http://lg.lan.switch.ch/lg/lg.cgi
  - <a href="http://lg.gin.ru/lg/">http://lg.gin.ru/lg/</a>
  - https://www.pch.net/tools/looking\_glass



### Pomoćni alati - BGPlay

Praćenje AS topologije tokom vremena





- Na ruteru R2 je postavljena default ruta koja pokazuje na ruter R1.
- Ruter R4 kada oglašava svoje rute ruteru R1 dodaje u AS-Path jednu dodatnu oznaku AS3 (AS prepending).
- Na ruteru R3 je pridružen MED parametar za rute koje se oglašavaju ka R5 na 100, a za rute koje se oglašavaju ka R7 na 200
- Na ruteru R3 je podešeno da sve rute koje dolaze od R5 imaju Local Preference 100, a sve rute koje dolaze od R7 imaju Local Preference 200.
- Na ruteru R10 je podešeno da sve rute koje dolaze od R6 imaju Local Preference 70, sve rute koje dolaze od rutera R9 imaju Local Preference 50,
- Na ruteru R11 je podešeno da sve rute dobijene od R14 imaju Local Preference 100.

#### Literatura

 Sam Halabi "Internet routing Architectures", Cisco press