



Diplomski studij

Informacijska i komunikacijska  
tehnologija:

Telekomunikacije i informatika

Obradba informacija

## Komunikacijski protokoli

8.

Protokol BGP (*Border Gateway Protocol*)

Ak.g. 2012./2013.

13.12.12.

## Creative Commons



### ■ slobodno smijete:



- **dijeliti** — umnožavati, distribuirati i javnosti priopćavati djelo



- **remiksirati** — prerađivati djelo

### ■ pod sljedećim uvjetima:



- **imenovanje.** Morate priznati i označiti autorstvo djela na način kako je specificirao autor ili davatelj licence (ali ne način koji bi sugerirao da Vi ili Vaše korištenje njegova djela imate njegovu izravnu podršku).



- **nekomercijalno.** Ovo djelo ne smijete koristiti u komercijalne svrhe.
- **dijeli pod istim uvjetima.** Ako ovo djelo izmijenite, preoblikujete ili stvarate koristeći ga, prerađu možete distribuirati samo pod licencom koja je ista ili slična ovoj.



U slučaju daljnog korištenja ili distribuiranja morate drugima jasno dati do znanja licencne uvjete ovog djela. Najbolji način da to učinite je linkom na ovu internetsku stranicu.

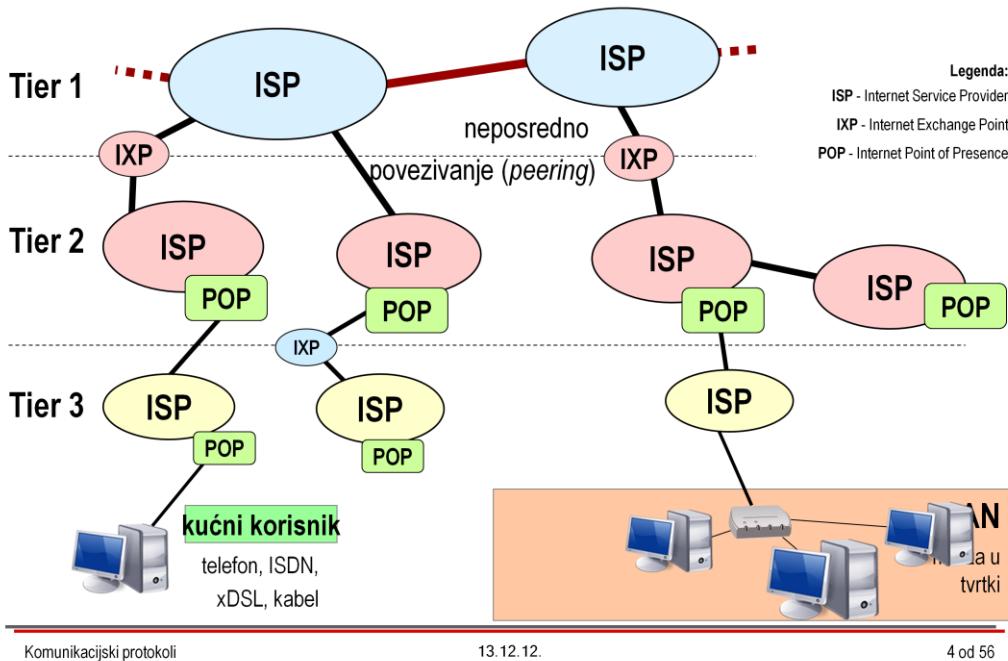
Od svakog od gornjih uvjeta moguće je odstupiti, ako dobijete dopuštenje nositelja autorskog prava.

Ništa u ovoj licenci ne narušava ili ograničava autorova moralna prava.

Tekst licencije preuzet je s <http://creativecommons.org/>.

- ◆ Protokoli usmjeravanja između autonomnih sustava
- ◆ Pojam autonomnog sustava
- ◆ Osnove protokola BGP
  - BGP poruke
  - BGP atributi
  - Komunikacija protokolom BGP
- ◆ Primjer BGP mreže
- ◆ Hrvatska akademska mreža CARNet
  - Croatian Internet eXchange (CIX)

# Hijerarhija Interneta



Komunikacijski protokoli

13.12.12.

4 od 56

Iako ne postoji "središnji autoritet" za Internet, u praksi je organizacija hijerarhijska. Razlikuju se ISP-ovi 3 razine (engl. *tier*) - Tier 1, Tier 2 i Tier 3 – s time da se svrstavanje u određenu kategoriju uglavnom temelji na obliku povezivanja s ostalima. Oblici povezivanja su *peering* (uzajamno povezivanje na istoj razini) i *transit* (prosljeđivanje prometa "prema gore" ili "prema dolje") prema drugim ISP-ovima. ISP-ovi razine 1 su najviši u hijerarhiji. Za ISP-ove razine 1 karakteristično je neposredno međusobno povezivanje svaki-sa-svakinim i uzajamna razmjena prometa 1:1 bez plaćanja. ISP-ovi razine 1 ne rade tranzit prometa trećih strana. Svi drugi ISP-ovi (razine 2 i 3) ugovaraju i plaćaju razmjenu prometa s ISP-om više razine.

Što se tiče podjele na razine, ne postoji formalna definicija ili mogućnost provjere pripadnosti konkretnog ISP-a pojedinoj razini izvan same organizacije, budući da su oblici povezivanja definirani uzajamnim ugovornim odnosima koji su poslovna tajna.

## Tier-1

ISP-ova razine 1 (Tier1) ima relativno malo, samo 10-tak (brojke se kreću od 7-14, kako prema kojem izvoru). ISP-ovi koje većina izvora navodi kao Tier 1 su: AT&T, Global Crossing, Level 3 Communications, NTT Communications, Qwest, Sprint, Verizon Business i SAVVIS. Uočimo da je većina u vlasništvu američkih tvrtki (izuzetak NTT – Japan!). S obzirom da ISP-ovi razine 1 ne rade tranzit prometa trećih strana, oni moraju biti povezani svaki-sa-svakinim. Tako mogu čuvati svoj povlašteni status – jasno je da neka nova mreža ne može postati Tier 1 bez pristanka postojećih Tier 1 ISP-ova na peering.

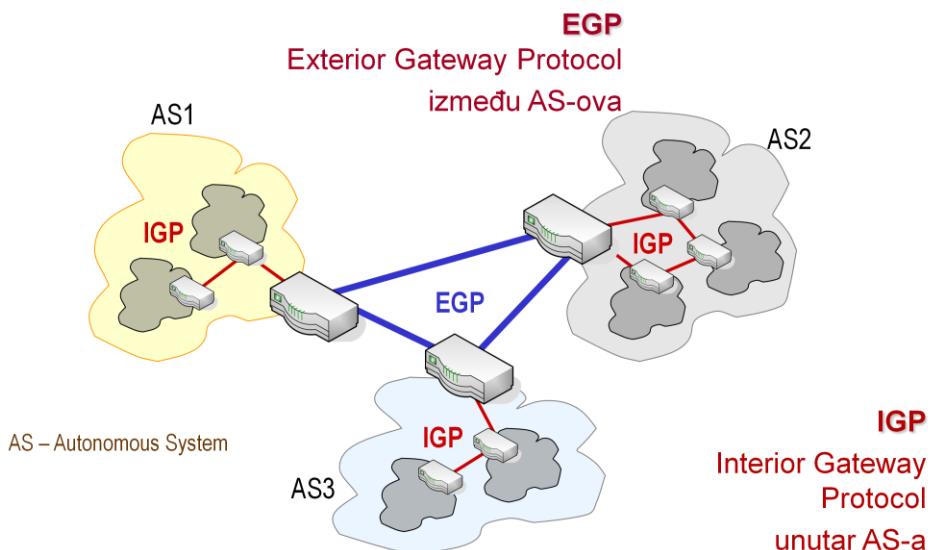
## Tier-2

ISP-ovi razine 2 imaju peering s nekim mrežama, ali i dalje moraju ugovarati tranzit s ISP-om više razine da bi mogli pristupiti dijelovima Interneta s kojima nema drugih oblika povezivanja. ISP-ovi razine 2 su, na primjer: GEANT, British Telecom, France Telecom i Deutsche Telekom.

## Tier-3

ISP razine 3 isključivo plaća tranzit drugim ISP-ovima. Primjer ISP-ova razine 3 su svi hrvatski ISP-ovi, na primjer: T-Com, VIPnet, Metronet, i CARNet.

# Klasifikacija protokola usmjeravanja



IGP protokoli: -> prošlo predavanje

- ◆ Routing Information Protocol (RIPv2)
  - ◆ temelji se na (dinamičkom) algoritmu vektora udaljenosti
- Open Shortest Path First Protocol (OSPFv2)
  - ◆ temelji se na (dinamičkom) algoritmu stanja poveznice

EGP protokol (u praksi, samo jedan!): -> danas!

- ◆ Border Gateway Protocol (BGPv4)
  - ◆ algoritam vektora staze (engl. *vector path*)
  - ◆ sličan algoritmu vektora udaljenosti, ali uzima u obzir puteve ili "staze" kao niz AS-ova na putu do odredišta

## Border Gateway Protocol (BGP)



- ◆ Draft Standard protokol, BGP-4 (RFC 4271)
- ◆ Protokol usmjeravanja (EGP)
- ◆ Komunikacija usmjeritelja između autonomnih sustava (AS)
- ◆ Razmjena usmjerivačkih informacija između pružatelja internetskih usluga (ISP-ova) te ISP-ova i većih korisnika
- ◆ Jedini EGP u Internetu
- ◆ Koristi TCP, port 179
- ◆ Podržava sve TCP veze između svih BGP usmjeritelja između AS-ova

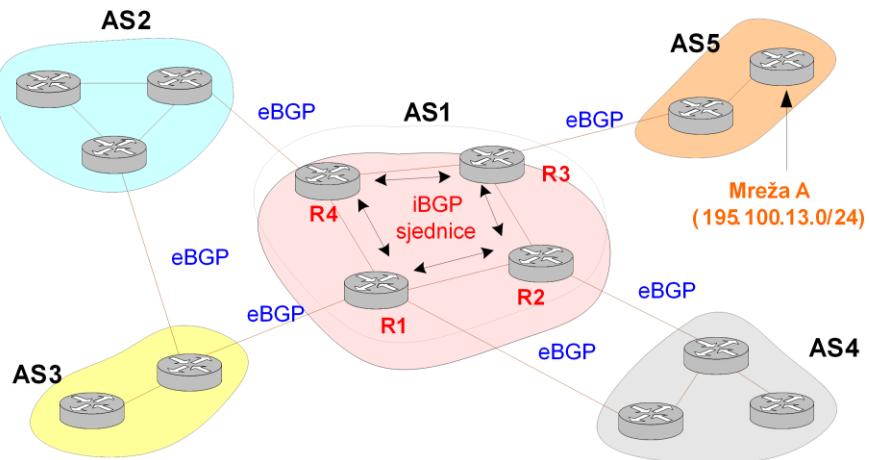
[4271](#) A Border Gateway Protocol 4 (BGP-4). Y. Rekhter, Ed., T. Li, Ed., S. Hares, Ed.. January 2006. (Format: [TXT](#)=222702 bytes) (Obsoletes [RFC1771](#)) (Status: DRAFT STANDARD)

## Border Gateway Protocol (BGP)



- ◆ Temelji se na **algoritmu vektora staza** (path-vector algorithm)
  - sličan algoritmu vektora udaljenosti, uzima u obzir stanje staze kao niz AS-ova na putu do odredišta
- ◆ Neovisan o korištenom IGP-u unutar AS-a
- ◆ Dva moda rada
  - unutarnji BGP (internal, iBGP) – između usmjeritelja unutar istog AS-a
  - vanjski BGP (external, eBGP) – između usmjeritelja različitih AS-ova

# BGP usmjeritelji



Dio R1 tablice usmjeravanja

Destination	Next hop	RP
195.100.13.0/24	R4	BGP
R3	R2	IGP

korišteni protokol  
usmjerenja

- ◆ Dvije kategorije prometa
  - lokalni – izvoriste i odredište unutar istog AS-a
  - vranxitni – izvoriste i odredište izvan istog AS-a
- ◆ Tranzitni BGP usmjeritelji
  - dopuštaju prolazak tranzitnog prometa
    - može nametnuti ograničenja
  - obično imaju **potpune tablice usmjeravanja**
    - preko 400,000 ruta (<http://bgp.potaroo.net>)
- ◆ Korištenjem CIDR i združivanja staza (*route aggregation*) smanjuju se veličine tablica usmjeravanja

## Pojam autonomnog sustava AS



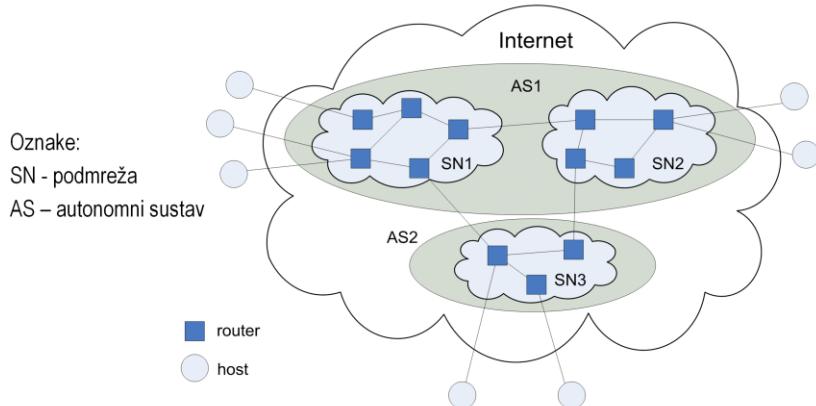
- ◆ Skup mreža i usmjeritelja temeljenih na istim načelima pod zajedničkom upravom i zajedničkom politikom usmjeravanja "prema van", odnosno prema ostalim AS-ovima
- ◆ Jedinstveni broj AS
  - Javni i privatni AS-ovi
  - Pružatelji internetskih usluga ISP-ovi (T-Com, VIPNet, Iskon, Metronet, Optima Telekom ...)
- ◆ Koriste jedinstveni IGP protokol usmjeravanja
- ◆ "Rubni" (vanjski) usmjeritelj na AS-u koji koristi eBGP naziva se "**peer**"
  - Administrativne granice, povezuje različite AS-ove
  - Izmjenjuje **informacije/poruke o putovima** s drugim AS-ovima – Network Layer Reachability Information (NLRI)

U nastavku predavanja pod pojmom **usmjeritelj** podrazumijeva se BGP usmjeritelj (rubni ili vanjski usmjeritelj) odnosno **peer**. Iznimno, ukoliko bi moglo doći do zabune, naglasiti će se radi li se o BGP usmjeritelju (*peeru*) ili običnom lokalnom usmjeritelju.

## Povezanost autonomnih sustava



- ◆ Jedinstveni broj AS-a dodijeljuje IANA
  - CARNet je AS 2108
- ◆ Pogled s motrišta usmjeravanja – skup AS-ova

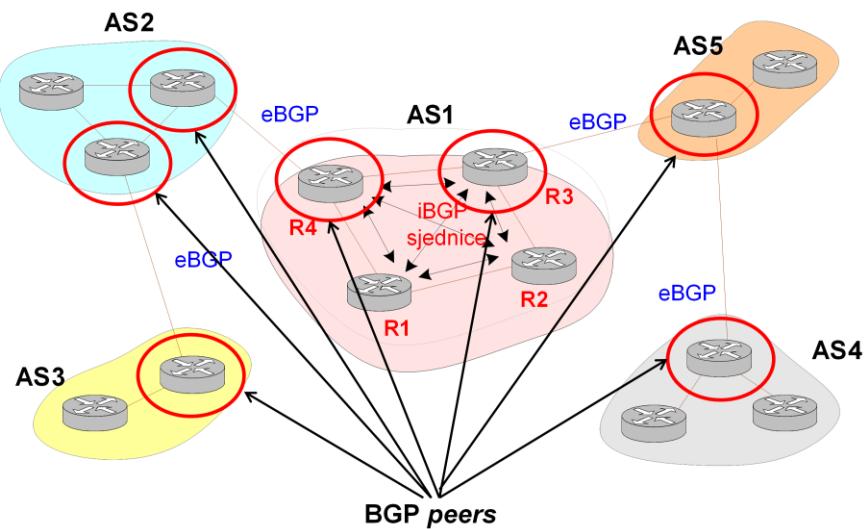


Promatrajući fizičku strukturu, Internet čine međusobno povezane *podmreže*, administrativno podijeljene u *autonomne sustave* (Autonomous System – AS).

Autonomni sustav definira se kao povezani dio mrežne topologije, tj. više podmreža s jedinstvenom i jasno definiranom politikom usmjeravanja “prema van”, odnosno prema ostalim autonomnim sustavima.

Stoga se autonomni sustav najčešće nalazi pod administracijom i u vlasništvu jednog mrežnog operatora. Primjer AS-a je Hrvatska akademska i istraživačka mreža CARNet.

# Komunikacija autonomnih sustava



- ♦ Komunikacija AS-ova odvija se preko BGP usmjeritelja (peerova)

## Vrste autonomnih sustava



- ◆ *Stub AS* – ima vezu sa samo jednim AS-om te prenosi samo lokalni promet
- ◆ *Multihomed AS* – povezan je s više od jednog AS-a, ali ne prenosi tranzitni promet
- ◆ *Transit AS* – povezan je s više od jednog AS-a i u skladu sa definiranim pravilima prenosi tranzitni i lokalni promet
- ◆ Protokol BGP neovisan je o korištenom protokolu IGP

## Osnove protokola BGP



- ◆ Usmjerava internetski promet između usmjeritelja na temelju algoritma vektora staza
- ◆ Pronalazak susjednih usmjeritelja vrši se **ručno** (administrator mreže)
- ◆ Kad usmjeritelji uspostave **TCP vezu**, razmijene cijele tablice usmjeravanja
- ◆ Usmjeritelji izmjenjuju informacije o putovima (NLRI)
- ◆ Tablica usmjeravanja sadrži informacije o stazama prema AS-u u kojem se nalazi odredišna mreža

## BGP staza



- ◆ Staza se sastoji slijeda autonomnih sustava koje treba proći do odredišta
- ◆ Mogu postojati višestruke staze
- ◆ Dopušta se primjena različitih politika usmjerenja
- ◆ Svaku stazu obilježavaju **skup parametara (atributa)** koji definiraju politiku usmjerenja
- ◆ Staza kojom se usmjerava paket odabire se na temelju:
  - parametara staza
  - dostupnosti staze
  - dodatnih pravila o prihvaćanju paketa (političkih, sigurnosnih,...)
  - pravila o propuštanju paketa
  - ugovora između usmjeritelja...
  - **atributa**

## BGP RIB



- ◆ BGP staze se objavljaju pomoću poruke UPDATE
- ◆ Svaki usmjeritelj sadrži bazu staza RIB (BGP Routing Information Base)
- ◆ Baza RIB sadrži tri vrste popisa
  - Popis neobrađenih staza koji su primljeni od susjednih usmjeritelja – uzimaju se u obzir kod procesa odluke (Adj-RIBs-In)
  - Popis staza s lokalnim informacijama o usmjeravanju do kojih se dolazi primjenom vlastitih pravila usmjeravanja i provođenjem procesa odluke nad popisom neobrađenih ruta (Loc-RIB)
  - Popis staza koje se šalju susjednim usmjeriteljima slanjem *update* poruka (Adj-RIBs-Out)

## BGP poruke



- ◆ Veličina poruke je od 19 (samo zaglavlje) do 4096 okteta
- ◆ Tri polja: oznaka (*marker*), veličina (*length*) i vrsta (*type*)
- ◆ Vrste BGP poruka:
  - OPEN
    - Uspostava veze (sjednice) između susjednih usmjeritelja i izmjena početnih postavki (identificiranje međusobnih mogućnosti)
  - UPDATE
    - Razmjena informacija o stazama (objava novih i ukidanje zastarjelih)
  - KEEPALIVE
    - Održavanje sjednice između usmjeritelja, potvrda nakon poruke *open*
  - NOTIFICATION
    - Obavijesti o pogreškama i zatvaranju sjednice

Polje *oznaka poruke* (*marker*) je veličine 16 okteta i sadrži sve jedinice.

Polje *veličina poruke* (*length*) je veličine 2 okteta.

Polje *vrsta poruke* (*type*) je veličine jednog okteta.

## Poruka OPEN



- ◆ Šalje se nakon uspostave TCP sjednice između usmjeritelja
  - Identifikacija usmjeritelja i poveznice između njih
- ◆ Svaka poruka definira parametre prema mogućnostima usmjeritelja (pregovaranje o parametrima sjednice)
- ◆ Uz fiksno zaglavlje poruka *open* sadrži sljedeća polja:
  - *version* – sadrži informaciju o inačici protokola BGP (BGPv4)
  - *my autonomous system* – sadrži broj AS-a u kojem se nalazi pošiljatelj poruke
  - *hold time* – vremenska kontrola, definira vrijeme čekanja, dopušteno trajanje neaktivnosti nakon čega se prekida sjednica (brojač se ponovo pokreće kada dođe poruka *keepalive* ili *update*)
  - *BGP identifier* - identifikacija pošiljatelja (IP adresa usmjeritelja)
  - *Optional parameters length* – veličina polja *optional parameters*
  - *Optional parameters* – sadrži popis izbornih parametara

Polje *hold time* mora imati vrijednost 0 ili najmanje 3 sekunde. Predstavlja maksimalan broj sekundi koji može proći između prijama dvaju uzastopnih poruka *keepalive* i/ili *update* od pošiljatelja.

Polje *optional parameter* definiran je kao trojka *<Parameter Type, Parameter Length, Parameter Value>*.

Polje *parameter type* određuje vrstu parametra.

Polje *parameter length* određuje veličinu polja *parameter value* u oktetima.

Polje *parameter value* promjenjive je veličine i dekodira se na različite načine, ovisno o vrijednosti polja *parameter type*.

- ◆ Razmjenjivanje informacije o putovima (NLRI)
  - Ovlašavanje mogućih putova iz baze (Adj-RIBs-Out) usmjeritelja
  - Ažuriranje putova i njihovih atributa
- ◆ Parametri poruke
  - *Path attributes* (atributi puta)
    - Popis atributa koji se ažuriraju za određene putove
  - *Total path attributes length*
    - Duljina polja *path attributes*
  - *Network layer reachability information* (NLRI)
    - Popis putova (IP adresa) koji se najavljuju za ažuriranje
  - *Withdrawn routes*
    - Popis putova koji su najavljeni, ali se neće ažurirati jer više nisu valjni
  - *Total route length*
    - Duljina polja *withdrawn routes*

Svaki atribut staze zapisuje se kao trojka promjenjive veličine *<attribute type, attribute length, attribute value>*.

Polje *attribute type*, veličine 2 okteta, sastoji se od zastavica atributa (1 oktet) iza kojih slijedi kod tipa atributa (1 oktet).

Informacija o dostupnosti (*reachability information*) zapisuje se kao dvojka *<length, prefix>*, gdje *length* označava veličinu prefiksa IP adrese u bitovima, a *prefix* sadrži prefiks IP adrese iza kojeg slijedi dovoljan broj bitova potrebnih da kraj polja završi na granici okteta. Minimalna veličina poruke UPDATE je 23 okteta. Porukom UPDATE može se objaviti najviše jedan skup atributa staze za više odredišta, ako ta odredišta imaju zajedničke attribute staze. Svi atributi staze, koji se nalaze u poruci UPDATE, odnose se na sva odredišta koja su popisana u polju NLRI.

### ◆ **Keepalive**

- Služi za utvrđivanje dostupnosti usmjeritelja i održavanje sjednice
- Uobičajeno vrijeme između *keepalive* poruka je 60 sekundi
- Vremenska kontrola se ponovno pokreće nakon primitka *keepalive* ili *update* poruke

### ◆ **Notification**

- Nakon slanja poruke raskida se BGP sjednica
- Indikacija pogreške kao što su:
  - Iste vremenske kontrole (*hold timer-a*)
  - Primitak nepoznatog atributa
  - Primitak pogrešnog AS broja
  - Polje *error code* u poruci otkriva razlog pogreške

Maksimalni razmak između dvije poruke *keepalive* je  $1/3$  *hold timer-a*. Uobičajeno vrijeme vremenske kontrole je 180 sekundi. Ne smiju se slati češće od jedne poruke u sekundi. Poruka *keepalive* sastoji se samo od zaglavlja veličine 19 okteta.

Poruka *notification* minimalne je veličine 21 oktet, a uz zaglavljivo sadrži sljedeća polja (polje *error code*):

- 1 – Message Header Error,
- 2 – OPEN Message Error,
- 3 – UPDATE Message Error,
- 4 – Hold Timer Expired,
- 5 – Finite State Machine Error,
- 6 – Cease.

Polje *error subcode* sadrži detaljniju informaciju o prirodi pogreške. Ako nije definiran odgovarajući *error subcode*, tada se u to polje upisuje vrijednost 0.

Polje *data* koristi se za opisivanje uzroka slanja poruke *notification*. Promjenjive je veličine i ovisi o poljima *error code* i *error subcode*. Veličina polja *data* može se izračunati iz izraza: ukupna veličina poruke =  $21 + \text{veličina polja } data$ .

## BGP atributi puta



- ◆ Nalaze se unutar poruke *UPDATE*
- ◆ Omogućavaju usmjeriteljima primjenu vlastite politike usmjeravanja
  - Dobro poznati obavezni (*well-known mandatory*)
  - Dobro poznati neobavezni (*well-known discretionary*)
  - Izborni tranzitni (*optional transitive*)
    - odnose se na sve AS-ove (globalni)
  - Izborni lokalni (*optional non-transitive*)
    - odnose se na AS koji ih prima
- ◆ Svaka staza može imati jedan ili više izbornih atributa kao dodatak dobro poznatim atributima

## Atributi staze

---



- ◆ 1 - ORIGIN
- ◆ 2 - AS path
- ◆ 3 - Next hop
- ◆ 4 - Multi-Exit Discriminator (MED)
- ◆ 5 – Local preference
- ◆ 6 – Atomic aggregate
- ◆ 7 - Aggregator

## Atribut **ORIGIN**

---



- ◆ Definira porijeklo staze (*origin*)
- ◆ Dobro poznati obavezni atribut staze
- ◆ Generira ga usmjeritelj od kojeg staza potječe
- ◆ Polje atributa je veličine jednog okteta
  
- ◆ **IGP-0:** put koji se prenosi unutar polja NLRI potječe iz istog AS-a u kojem se nalazi usmjeritelj
- ◆ **EGP-1:** put koji se prenosi unutar polja NLRI potječe iz drugog AS-a
- ◆ **Incomplete-2:** put koji se prenosi unutar polja NLRI dobiven je nekim drugim načinom (nije poznat)

## Atribut AS path



- ◆ Definira stazu - listu AS-ova (segmenata puta) koje treba proći do odredišta
- ◆ Dobre poznati obavezni atributi staze
- ◆ Svaki segment puta AS-a zapisan je trojkom:
  - tip (skup AS-ova koje je poruka *update* prešla),
  - duljina (broj prijeđenih AS-ova),
  - vrijednost (brojevi AS-ova)
- ◆ Korisno kod višestrukih staza - različiti atributi za isto odredište
- ◆ Izbjegavanje petlji
- ◆ Filtriranje – zabrana usmjeravanja paketa kroz određeni AS
- ◆ Preferiranje staze

## Atribut *Next hop*

---



- ◆ Definira IP adresu usmjeritelja na koji prvo treba usmjeriti paket kako bi došao do odredišta (sljedeći skok)
- ◆ Dobro poznati obavezni atribut staze
- ◆ Kod ažuriranja puta atribut se modificira samo ako dolazi od BGP usmjeritelja (*peera*) eBGP vezom

## Atribut MED



- ◆ Višeizlazni diskriminirajući atribut MED (Multi-Exit Discriminator)
- ◆ Služi za odabir jednog od više ponuđenih staza prema istom AS-u
- ◆ Izborni lokalni atribut
- ◆ Usmjeritelji daju savjete svojim susjedima kojim putem poslati pakete prema njima (dolazni promet)
- ◆ Atribut predstavlja savjet samog AS-a kojem se šalju paketi
  - preferira se određeni put prema vlastitom AS-u
- ◆ Kada prema AS-u postoji više mogućih staza, odabire ona staza koja ima najmanju vrijednost atributa MED (uzimajući u obzir i ostale attribute odnosno kada se put ne može odabrati na temelju ostalih atributa)

## Atribut *Local preference*

---



- ◆ Određuje politiku usmjerenja odlaznog prometa
- ◆ Dobro poznati neobavezni atribut staze
- ◆ Atribut se izmjenjuje između **lokalnih usmjeritelja** istog AS-a
  - Atribut ne utječe na ostale autonomne sustave
- ◆ U slučaju kada postoji više izlaznih staza iz AS-a odabire se onaj koji ima veću vrijednost atributa što označava veći prioritet staze

## Atribut *Atomic aggregate*

---



- ◆ Združena staza do odredišta
- ◆ Dobro poznati neobavezni atribut staze
- ◆ Združivanje staze omogućava da više putova sa svojim karakteristikama može biti združeno i objavljeno kao jedan put u svrhu reduciranja broja putova
  - Smanjen broj staza koji se moraju pohranjivati i izmjenivati između usmjeritelja
- ◆ Vrijednost atributa obično odgovara polju *tip* u atributu AS *path* i sastoji se od liste AS-ova od kojih je napravljen združenja staza

## Atribut Aggregator

---



- ◆ Izborni tranzitni atribut
- ◆ Uključen u poruke koje su nastale združivanjem staza
- ◆ Združivati se mogu staze koji imaju iste atribute
- ◆ Daje do znanja da je usmjeritelj združio rutu i zapisuje svoj AS broj i IP adresu

## Primjer združivanja staza



### ◆ Pronaći združenu stazu za sljedeće putove (adrese)

- 192.168.98.0 (11000000.10101000.01100010.00000000)
- 192.168.99.0 (11000000.10101000.01100011.00000000)
- 192.168.100.0 (11000000.10101000.01100100.00000000)
- 192.168.102.0 (11000000.10101000.01100110.00000000)
- 192.168.104.0 (11000000.10101000.01101000.00000000)
- 192.168.105.0 (11000000.10101000.01101001.00000000)  
                        11000000.10101000.01100000.00000000

192 .168. .96 .0

Združena staza je 192.168.96.0/20

Maska podmreže 255.255.240.0

## Primjer združivanja staze (2)



- ◆ Oprez: rješenje sadrži i staze koji ne pripadaju združenom putu
  - 192.168.96.0, 192.168.97.0, 192.168.101.0, 192.168.103.0
- ◆ Združena staza mora isključiti prve dvije staze koje ustvari ne pripadaju združenoj stazi
  - 192.168.96.0, 192.168.97.0
- ◆ Združena staza mora sadržavati prvu adresu koja pripada združenoj stazi
- ◆ Rješenje:
  - 192.168.98.0/20

## Primjer združivanja staze (3)



- ◆ Združena staza 10.10.1.32/27
- ◆ Pripada li združenoj stazi adresa 10.10.1.44/27?
- ◆ Pripada li združenoj stazi adresa 10.10.1.90/27?

Rješenje:

10.10.1.32/27 = 00001010.00001010.0000001.00100000

10.10.1.44/27 = 00001010.00001010.0000001.00101101 (pripada!)

10.10.1.90/27 = 00001010.00001010.0000001.01011010 (**ne** pripada!)

## Algoritam usmjeravanja



- ◆ Algoritam usmjeravanja odlučuje o “najboljoj stazi” za zadano odredište na temelju **procesa odluke** (*decision process*)
- ◆ Nema definiranog pravila već se primjenjuje vlastita politika lokalnog administratora AS-a definirana u PIB (Policy Information Base)
- ◆ Promatraju se neobrađene staze objavljeni u RIB-u (Adj-RIBs-In)

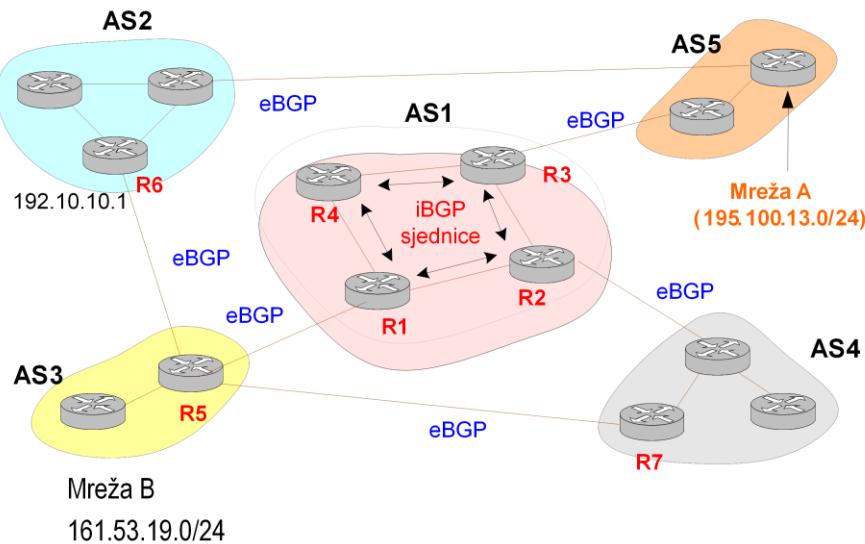
## Primjer algoritma usmjeravanja



### ◆ Proces donošenja odluke o staze:

- Odaberite stazu s najvećom vrijednosti atributa *local pref*. Ako se staza ne može odrediti na temelju ovog kriterija priđi na sljedeći korak
- Odaberite stazu kojida je domaćeg porijekla (*origin*), dobivena iz vlastitog AS-a. Ako se staza ne može odrediti na temelju ovog kriterija priđi na sljedeći korak
- Odaberite stazu s najkraćim atributom *AS path*. Ako se staza ne može odrediti na temelju ovog kriterija priđi na sljedeći korak
- Odaberite stazu s manjom vrijednosti atributa *origin*. Ako se staza ne može odrediti na temelju ovog kriterija priđi na sljedeći korak
- Odaberite stazu s najmanjim atributom *MED*. Ako se staza ne može odrediti na temelju ovog kriterija priđi na sljedeći korak
- Odaberite stazu koja je definirana na temelju eBGP

## Primjer odabira staze na temelju atributa



Iz mreže B (usmjeritelja R5, AS3) se do mreže A (AS5) može doći različitim putovima. Prepostavimo da postoje tri moguće staze:

- staza AS3 - AS2 – AS5,
- staza AS3 - AS1 – AS5,
- staza AS3 - AS4 - AS1 – AS5.

Za odabir staze promatraju se atributi putova zapisani u bazi neobrađenih staza i odabire se jedna od njih. Odabrana staza će se zapisati u tablicu usmjeravanja usmjeritelja R5 i izmjenjivati s ostalim usmjeriteljima (NLRI) porukom *update*.

## Primjer odabira staze od AS3 prema AS5



- ◆ Atributi staze 1
  - Origin: IGP-0
  - AS path: AS2 – AS5
  - Local preference: 3
  - MED: 1
- ◆ Atributi staze 2
  - Origin: IGP-0
  - AS path: AS1 – AS5
  - Local preference: 3
  - MED: 2
- ◆ Atributi staze 3
  - Origin: IGP-0
  - AS path: AS4 - AS1 – AS5
  - Local preference: 3
  - MED: 3

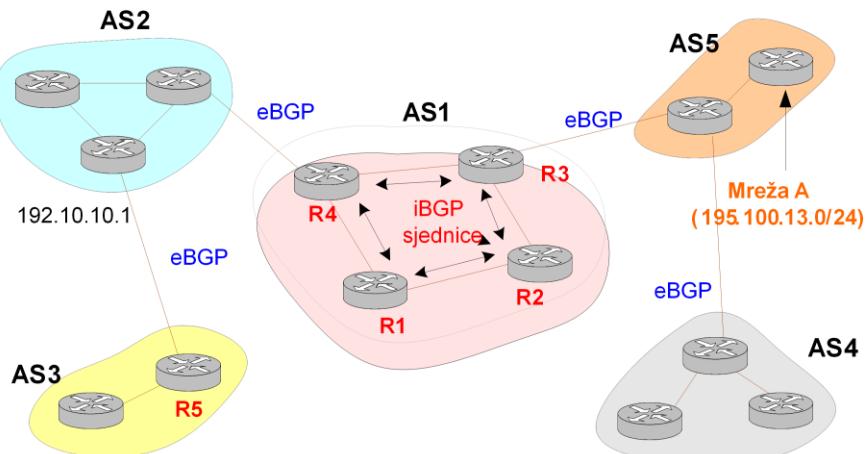
Pretpostavimo da se proces donošenja odluke o putu od više mogućih staza temelji na algoritmu opisanom na slajdu 35 (primjer algoritma/politike usmjeravanja autonomnog sustava). U tom slučaju, prvo se za sve tri moguće staze gleda atribut *local preference*. Kako u ovom slučaju atribut ima istu vrijednost za sve tri staze, prelazi se na sljedeći kriterij, porijeklo staze, odnosno atribut *origin*. Kako su sve ponuđene staze domaćeg porijekla, prelazi se na sljedeći kriterij, duljina staze (AS *path*). Ovim kriterijem otpada staza 3 jer je on najduži. Iz preostale dvije staze gleda se vrijednost atributa *origin*, a kako se opet na temelju njegove vrijednosti ne može odrediti "bolja" staza, promatra se atribut MED po kojem je odabrana staza 1, budući da ima manju vrijednost atributa MED. Dakle staza koja se upisuje u tablicu usmjeravanja usmjeritelja R5 za odredišni AS5 je AS path = AS2 – AS5 sa atributom next hop = R6.

## Ažuriranje tablice usmjerenja



- ◆ Vrši se razmjenom poruka *BGP updates*
- ◆ Komunikacija između susjednih BGP usmjeritelja
- ◆ Izmjenjuju informacije o putovima (NLRI)
- ◆ Oglasavaju moguće putove
  
- ◆ Tablica usmjerenja
  - Odredište, put
- ◆ Uz oznaku puta navodi se sljedeći skok
- ◆ Ukoliko dođe do prekida u mreži put se briše

## Primjer BGP tablice usmjeravanja



Dio R5 tablice usmjeravanja

Destination	Next hop	AS path
195.100.13.0/24	192.10.10.1	AS2 AS1 AS5
161.53.19.0/24	192.10.10.1	AS2 AS1 AS5 AS4

## Komunikacija protokolom BGP



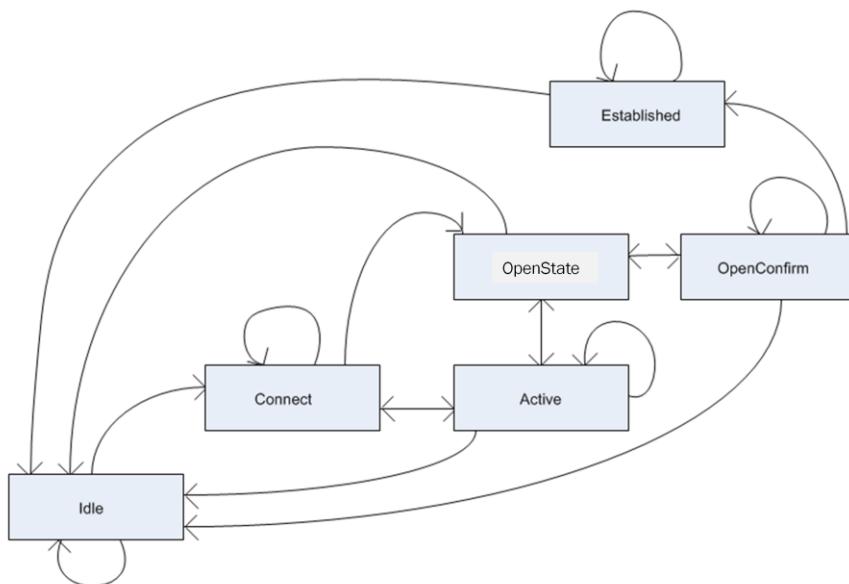
- ◆ Komunikacija BGP usmjeritelja u internetskoj mreži može se predočiti modelom konačnog automata
- ◆ Definirana su šest stanja:
  - *Idle*
  - *Connect*
  - *Active*
  - *OpenState*
  - *OpenConfirm*
  - *Established*

BGP usmjeritelj se inicijalno nalazi u stanju ***Idle***. U ovom stanju usmjeritelj odbija sve dolazne BGP konekcije. Kao odgovor na događaj *Start*, lokalni sustav inicijalizira sve BGP resurse, pokreće brojač *ConnectRetry*, te pokušava uspostaviti TCP vezu s konfiguiranim susjednim usmjeriteljima, i počinje slušati dolazne konekcije od udaljenih BGP usmjeritelja, te prelazi u stanje *Connect*. Ukoliko usmjeritelj otkrije pogrešku, zatvara vezu i prelazi u stanje *Idle*. Kako bi usmjeritelj ponovno izšao iz stanja *Idle*, potrebno je generirati događaj *Start*. Ako se taj događaj generira automatski, uzastopne pogreške mogu uzrokovati uzastopno prelaženje iz stanja *Connect* u stanje *Idle* i obrnuto. Za usmjeritelj koji je došao u stanje *Idle* zbog pogreške, događaj *Start* ne bi se trebao odmah generirati, nego bi se trebalo pričekati određeni period vremena.

Neki od razloga zadržavanja usmjeritelja u stanju *Idle* su:

- zatvoren je port 179;
- zatvoren je slučajno odabrani TCP port veći od 1023;
- adresa usmjeritelja je krivo konfigurirana;
- broj AS-a je krivo konfiguriran.

## Model konačnog automata protokola BGP



U stanju **Connect** usmjeritelj čeka uspostavu TCP veze s drugim usmjeriteljem. Ako je TCP veza uspješno uspostavljena, lokalni sustav poništava brojač *ConnectRetry*, dovršava inicijalizaciju, šalje poruku OPEN usmjeritelju s kojim je veza uspostavljena, te prelazi u stanje *OpenState*. Ako TCP veza nije uspostavljena, lokalni sustav resetira brojač *ConnectRetry*, nastavlja čekati konekcije koje može pokrenuti udaljeni BGP peer, i prelazi u stanje *Active*. Ako se dogodi događaj *ConnectRetry timer expired* (istek vremena brojača *ConnectRetry*), lokalni sustav resetira brojač *ConnectRetry*, pokreće uspostavu TCP veze s drugim usmjeriteljem, te nastavlja čekati konekciju koju može pokrenuti udaljeni usmjeritelj, i ostaje u stanju *Connect*.

U stanju **Active** usmjeritelj se nalazi ako TCP veza s *drugim usmjeriteljem* nije uspostavljena u prvom pokušaju. Usmjeritelj ponovno pokušava uspostaviti TCP vezu sa susjednim. Ako je TCP veza uspješno uspostavljena, lokalni sustav poništava brojač *ConnectRetry*, dovršava inicijalizaciju, šalje poruku OPEN usmjeritelju s kojim je veza uspostavljena, postavlja svoj brojač *Hold Timer* na neku veliku vrijednost (predlaže se 4 minute), te prelazi u stanje *OpenState*. Ako se dogodi događaj *ConnectRetry timer expired* (istek vremena brojača *ConnectRetry*), lokalni sustav resetira brojač *ConnectRetry*, pokreće uspostavu TCP veze s drugim usmjeriteljem, te nastavlja čekati konekciju koju može pokrenuti udaljeni usmjeritelj, i prelazi u stanje *Connect*. Ako TCP veza nije uspostavljena, vraća se u stanje *Idle*.

## Stanje OpenState



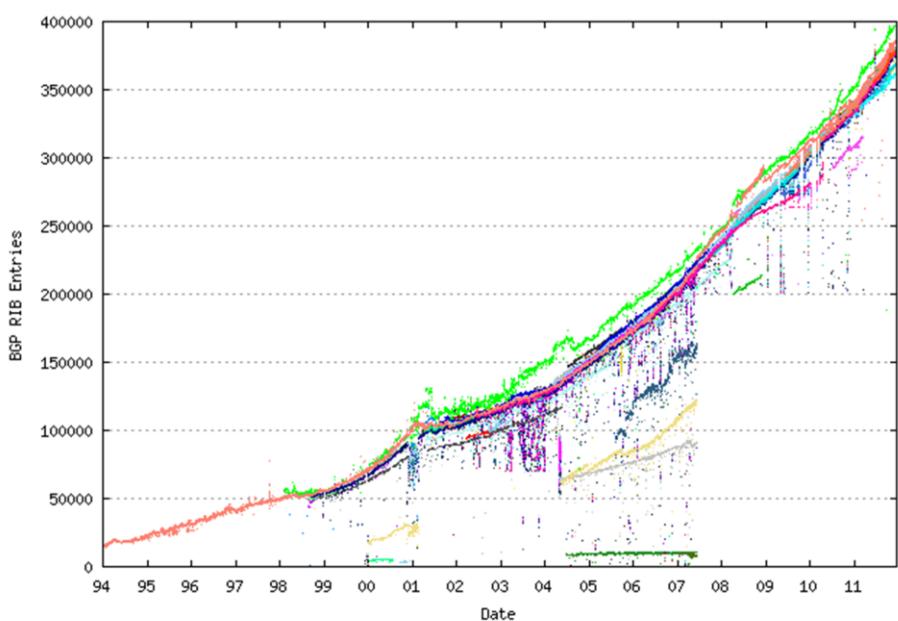
- ◆ Usmjeritelj čeka poruku *open* od drugog usmjeritelja s kojim je uspostavio vezu
- ◆ Nakon primanja poruke *open*, provodi se proces ispitivanja valjanosti poruke
- ◆ Ako otkrije pogrešku u poruci, šalje poruku *notification* i prelazi u stanje *Idle*
- ◆ Ako nema pogreške, šalje poruku *keepalive* kojom potvrđuje da je primio ispravnu poruku *open* te postavlja brojač *hold timer* na dogovorenu vrijednost i prelazi u stanje *OpenConfirm*
- ◆ Ukoliko istekne vrijeme brojača *hold timer*, usmjeritelj šalje poruku *notification* te prelazi u stanje *Idle*

Ako u stanju **OpenState** usmjeritelj primi od protokola TCP obavijest o prekidu veze, lokalni sustav zatvara BGP vezu, resetira brojač *ConnectTimer*, te nastavlja čekati konekciju koju može pokrenuti udaljeni usmjeritelj, i prelazi u stanje *Active*. Svaki put kad BGP speaker pređe iz stanja *OpenState* u stanje *Idle* zatvara se BGP veza i otpuštaju se svi resursi povezani s tom vezom.

U stanju **OpenConfirm** usmjeritelj čeka poruku *KEEPALIVE* (potvrda usmjeritelja da je primio valjanu poruku *OPEN*) ili *NOTIFICATION* (u slučaju kad usmjeritelj primi poruku *OPEN* s nekakvom pogreškom). Lokalni sustav prelazi u stanje *Established* ako primi poruku *KEEPALIVE*. Ako istekne vrijeme brojača *Hold Timer* prije nego lokalni sustav primi poruku *KEEPALIVE*, tada on šalje peeru poruku *NOTIFICATION* s kodom pogreške *Hold Timer Expired*, i prelazi u stanje *Idle*. Lokalni sustav prelazi u stanje *Idle* ako od peer-a primi poruku *NOTIFICATION*. Ako istekne vrijeme brojača *KeepAlive*, lokalni sustav šalje usmjeritelju poruku *KEEPALIVE* i resetira brojač *KeepAlive*. Ako od protokola TCP primi obavijest o prekidu veze, lokalni sustav prelazi u stanje *Idle*. Svaki put kad usmjeritelj priđe iz stanja *OpenConfirm* u stanje *Idle* zatvara se BGP veza i otpuštaju se svi resursi povezani s tom vezom.

U stanju **Established** usmjeritelj može razmjenjivati poruke *UPDATE*, *NOTIFICATION* i *KEEPALIVE* s usmjeriteljem s kojim je uspostavljena veza. Ako lokalni sustav primi poruku *UPDATE* ili *KEEPALIVE*, resetira vrijednost brojača *Hold Timer*. Lokalni sustav prelazi u stanje *Idle* ako primi poruku *NOTIFICATION*. Ako lokalni sustav primi poruku *UPDATE* iako procesom provjere utvrdi grešku u toj poruci, tada on šalje poruku *NOTIFICATION* i prelazi u stanje *Idle*. Ako od protokola TCP primi obavijest o prekidu veze, lokalni sustav prelazi u stanje *Idle*. Ukoliko istekne vrijeme brojača *Hold Timer*, tada on šalje poruku *NOTIFICATION* s kodom pogreške *Hold Timer Expired*, i prelazi u stanje *Idle*. Ukoliko istekne vrijeme brojača *KeepAlive*, lokalni sustav šalje peeru poruku *KEEPALIVE* i resetira brojač *KeepAlive*. Svaki put kad lokalni sustav pošalje poruku *UPDATE* ili *KEEPALIVE*, resetira svoj brojač *KeepAlive*. Svaki put kad BGP speaker pređe iz stanja *Established* u stanje *Idle* zatvara se BGP veza i otpuštaju se svi resursi povezani s tom vezom, te se brišu sve rute izračunate tijekom trajanja te veze.

## Broj zapisa u BGP tablici usmjerenja



Komunikacijski protokoli

13.12.12.

43 od 56

Podaci preuzeti sa <http://bgp.potaroo.net/> dana 29.11.2011.

- ◆ Regionalni internetski registar RIR (Regional Internet Registry) vrši raspodjelu brojeva AS-a i IP adresa dobivenih od IANA (Internet Assigned Number Authority)
- ◆ Aktivno je pet registara s područjem djelovanja:
  - RIPE NCC (RIPE Network Coordination Centre) – Europa, Bliski Istok i središnja Azija
  - ARIN (American Registry for Registry Numbers) – Sjeverna Amerika i dijelovi Kariba
  - APNIC (Asia-Pacific Network Coordination Centre) – Azija i Tihi ocean
  - LACNIC (Latin American and Caribbean Internet Address Registry) – Latinska Amerika i Karibi
  - AfriNIC (African Network Information Centre) - Afrika

CARNet mreža kao i ostali ISP-ovi u RH spadaju pod nadležnost RIPE NCC.

Čelna svjetska organizacija (IANA) koja raspolaže IP adresama sa sjedištem u Los Angeles-u u SAD-u dijeli adrese RIR-ovima, a RIR-ovi dalje adrese dodjeljuju National Internet Registry-ima (NIR) ili izravno ISP-ovima ili LIR-ovima (Local Internet Registry). NIR postoji samo u APNIC-u.

CARNet je LIR kao i ostali pružatelji internetskih usluga koji djeluju u RH (FINA, Global Net Grupa d.o.o., HEP d.d., Metronet telekomunikacije d.d., Optika Kabel TV d.o.o., Portus d.o.o., T-Mobile Hrvatska d.o.o., VIPnet d.o.o., Vodatel d.o.o., Voljatel telekomunikacije d.o.o. – podaci za studeni 2009.).

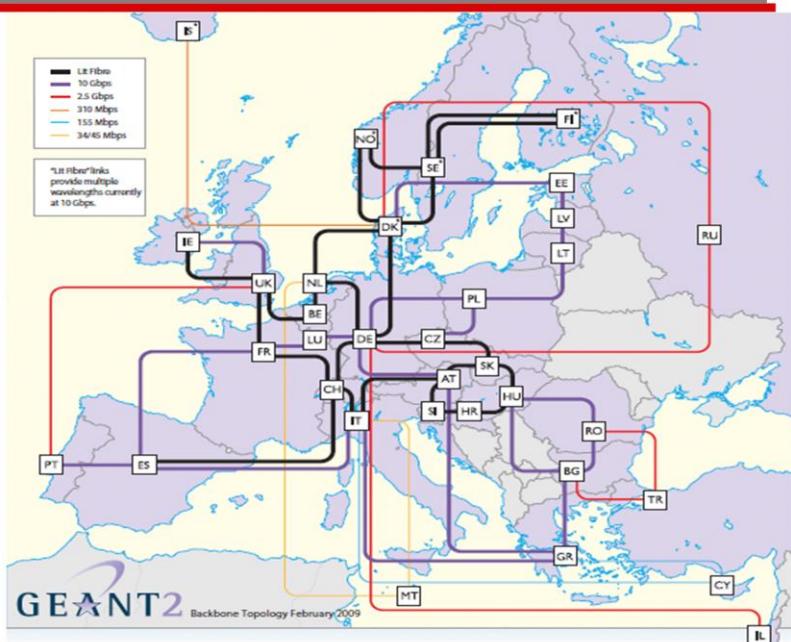
## Hrvatska akademska istraživačka mreža



- ◆ Hrvatska akademska istraživačka mreža (CARNet)
- ◆ Upravlja s 164092 IP adrese
  - 82.132.0.0 – 82.132.127.255 (/17), 32767 adresa
  - 161.53.0.0 – 161.53.255.255 (/16), 65535 adresa
  - 192.84.105.0 – 192.84.105.255 (/24), 255 adresa
  - 193.198.0.0 – 193.198.255.255 (/16), 65535 adresa
- ◆ Ostvaruje vezu s Internetom preko pan-europske istraživačke mreže GEANT brzinom 10 Gbit/s
- ◆ Veza prema drugim ISP-ovima u Hrvatskoj ostvarena je kroz **mjesto razmjene internetskog prometa u Hrvatskoj – CIX** (Croatian Internet eXchange)

<http://www.carnet.hr/mreza?CARNetweb=>

## GEANT mreža



Komunikacijski protokoli

13.12.12.

46 od 56

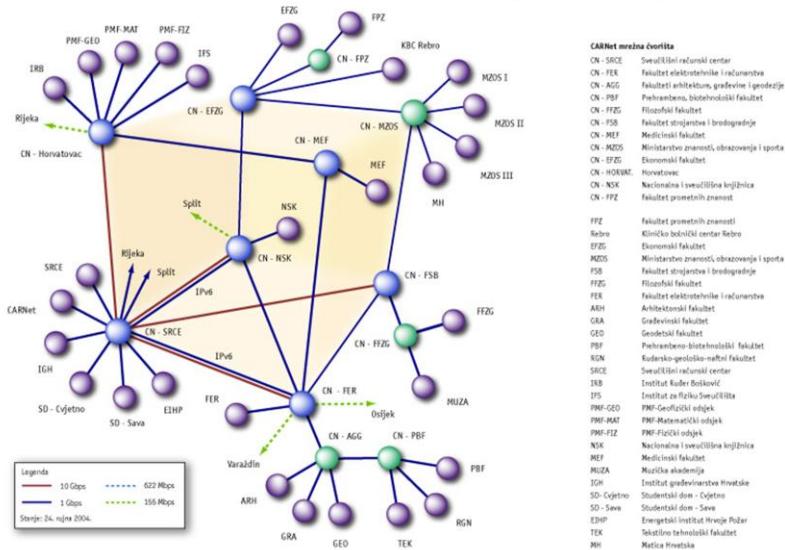
Izvor: <http://www.carnet.hr/>

- ◆ Unutar Hrvatske CARNet povezuje sve veće hrvatske gradove, i to na nekoliko razina različitih tehnologija i pristupnih brzina
- ◆ Okosnica CARNet mreže povezuje veće sveučilišne centre (DU, OS, PU, RI, ST, ZD, ZG) brzinama (od 100 Mbit/s do 1 Gbit/s), dok druge, manje centre povezuje modemskim vezama preko iznajmljenih linija (najčešće brzinom od 2 do 100 Mbit/s).
- ◆ Mreža u samom Zagrebu povezuje veće fakultete i znanstvene ustanove brzinama do 10 Gbit/s

# ZG CARNet mreža



Stanje zagrebačkog dijela CARNet mreže



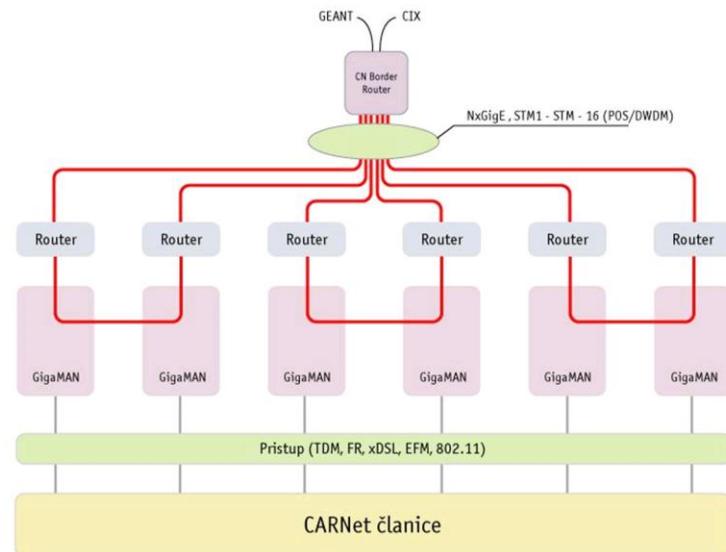
Komunikacijski protokoli

13.12.12.

48 od 56

Izvor: <http://www.carnet.hr/>

# CARNet mreža



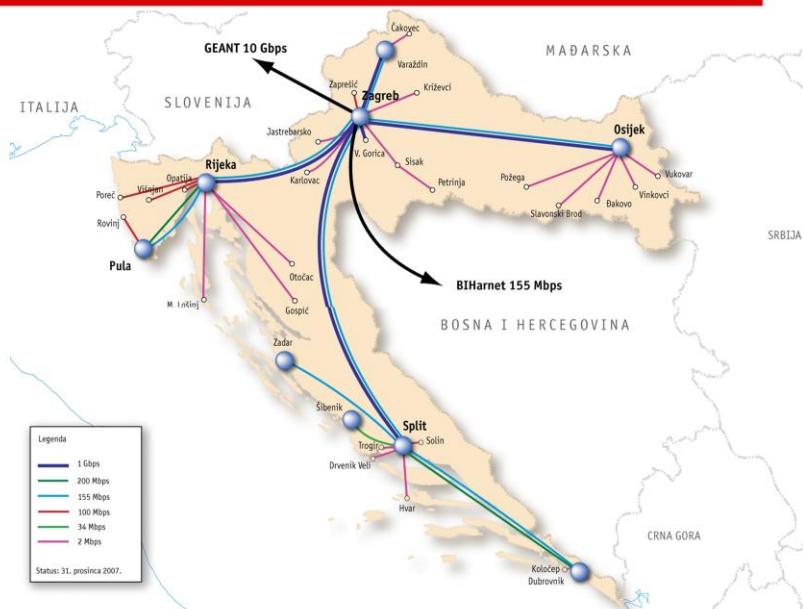
Komunikacijski protokoli

13.12.12.

49 od 56

Izvor: <http://www.carnet.hr/>

## CARNet mreža (2)



Komunikacijski protokoli

13.12.12.

50 od 56

Izvor: <http://www.carnet.hr/>

- ◆ Croatian Internet eXchange (CIX) je hrvatsko nacionalno središte za razmjenu internetskog prometa
- ◆ Otvoren je za sve ISP-ove u RH, kako za komercijalne tako i nekomercijalne, odnosno privatne mreže
- ◆ Uspostavom izravnih komunikacijskih kanala među hrvatskim ISP-ovima postiže se velika ušteda u razmjeni podataka među hrvatskim internetskim korisnicima.
  - izravno međusobno povezivanje ISP-ova smanjuje nepotrebni promet kroz treće mreže
- ◆ CIX članice dogovaraju međusobni *peering* za izmjenu prometa
- ◆ Protokol usmjeravanja je BGPv4, a baza za dokumentiranje *CIX peeringa* je RIPE baza podataka

Izvor: <http://www.cix.srce.unizg.hr/>

## Matrica “peeringa” IPv4



## Komunikacijski protokoli

13.12.12.

52 od 56

Izvor: [http://www.cix.srce.unizg.hr/cix\\_peering\\_matrica.html](http://www.cix.srce.unizg.hr/cix_peering_matrica.html)

## Matrica “peeringa” IPv6



## Komunikacijski protokoli

13.12.12

53 od 56

Izvor: [http://www.cix.srce.unizg.hr/cix\\_peering\\_matrica.html](http://www.cix.srce.unizg.hr/cix_peering_matrica.html)

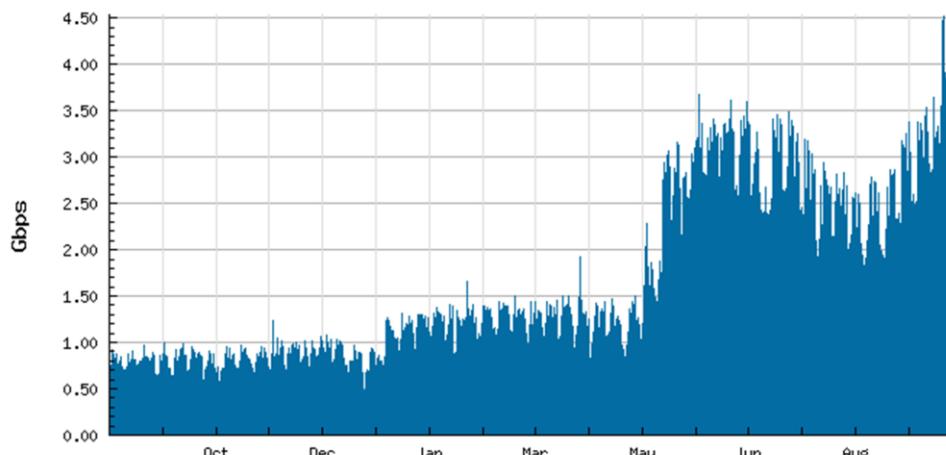
## Promet na CIX mreži



UKUPNI PROMET CIX mreže

Datum: od 12.09.2010 10:06 do 12.09.2011 10:06

Ukupno : MAX: 4.53 MIN: 0.1 AVG: 1.11 Gbps



Komunikacijski protokoli

13.12.12.

54 od 56

Izvor: <http://www.cix.srce.unizg.hr/>

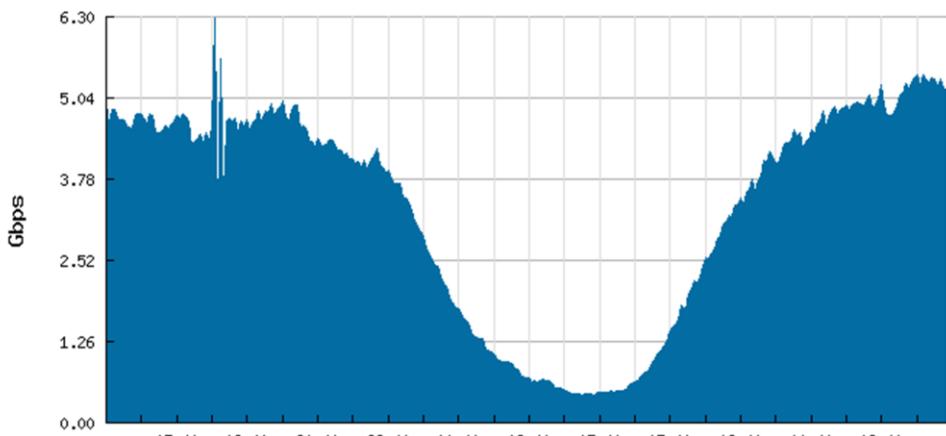
## Dnevni promet na CIX mreži



UKUPNI PROMET CIX mreže

Datum: od 28.11.2011 15:35 do 29.11.2011 15:35

Ukupno : MAX: 6.3 MIN: 0.43 AVG: 3.32 Gbps



Komunikacijski protokoli

13.12.12.

55 od 56

Izvor: [http://www.cix.srce.unizg.hr/cix\\_ukupni\\_promet.html](http://www.cix.srce.unizg.hr/cix_ukupni_promet.html)

## Analiza BGP prometa unutar RIPE NCC



### RIPE Region Analysis Summary

---

Prefixes being announced by RIPE Region ASes:	86545
Total RIPE prefixes after maximum aggregation:	49132
RIPE Deaggregation factor:	1.76
Prefixes being announced from the RIPE address blocks:	79935
Unique aggregates announced from the RIPE address blocks:	52845
RIPE Region origin ASes present in the Internet Routing Table:	15797
RIPE Prefixes per ASN:	5.06
RIPE Region origin ASes announcing only one prefix:	7877
RIPE Region transit ASes present in the Internet Routing Table:	2504
Average RIPE Region AS path length visible:	4.6
Max RIPE Region AS path length visible:	33
Number of RIPE region 32-bit ASNs visible in the Routing Table:	912
Number of RIPE addresses announced to Internet:	478890496
Equivalent to 28 /8s, 139 /16s and 74 /24s	
Percentage of available RIPE address space announced:	77.1

Podaci preuzeti dana 9.07.2011.