

Multikast

Predavači:

dr Slavko Gajin
dr Pavle Vuletić

Autori:

prof. dr Zoran Jovanović
dr Pavle Vuletić
dr Slavko Gajin
Marina Vermezović



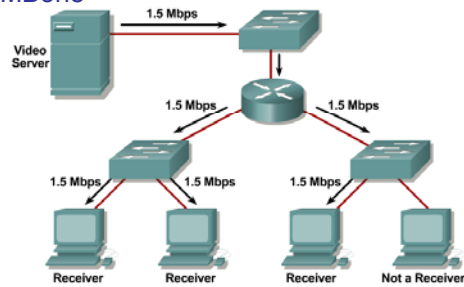
“Multicast could be the poster child for the irrelevance of the networking research community. Few other technologies (quality of service springs to mind) have generated so many research papers while yielding so little real-world deployment.”

Bruce Davies, public review of ACM Sigcomm 2006 accepted paper, “Revisiting IP Multicast” by S. Ratnasamy, A. Ermolinskiy, S. Shenker
<http://www.sigcomm.org/sigcomm2006/discussion/>

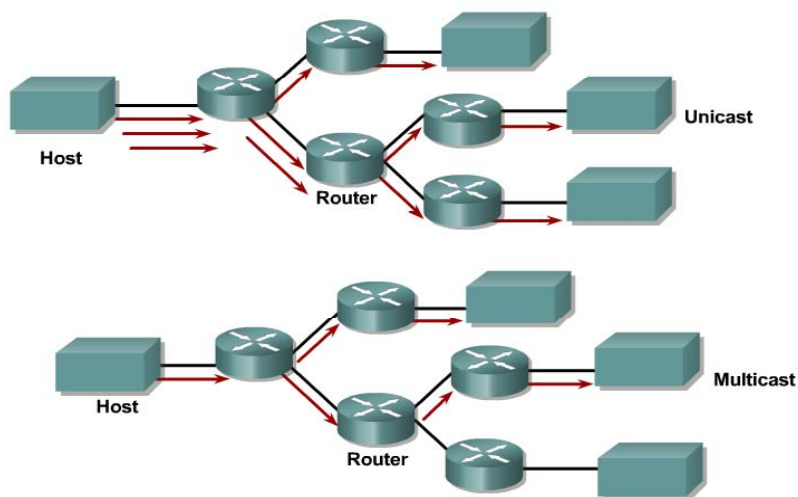


Multikast

- Unikast (*unicast*) – prenos od jednog izvorišta do jednog odredišta
- Brodkast (*broadcast*) – prenos od jednog izvorišta do svih odredišta
- Multikast (*multicast*) – prenos od jednog izvorišta do pojedinih odredišta - jedan IP tok po linku - efikasan prenos IP paketa
- Stephen Deering, doktorska disertacija početkom 90-tih
- Prva mreža - Mbone



Unicast vs. Multicast



Karakteristike

- Multikast se odlikuje sledećim karakteristikama:
 - Prenos multikast paketa do odredištnih multikast adresa (tzv. multikast grupa) sa istom pouzdanošću kao i regularni unicast paketi
 - Omogućava članovima multikast grupe da dinamički (po potrebi) pristupe i napuste grupu
 - Podržava više grupe hostova bez obzira na lokaciju ili broj članova
 - Podržava da pojedinačni hostovi budu članovi jedne ili više grupa
 - Može da prenosi više tokova do iste grupe
 - Može da koristi jednu grupnu adresu za više aplikacija
 - Izvorišni server nema predstavu od broju prijemnih odredišta



Principi

- Multikast saobraćaja se na transportnom nivou prenosi po UDP protokolu – Zašto?
- Ako se posebno ne konfigurise na uređajima, multikast se blokira na ruterima, slično kao i brodkast saobraćaj
- Ruteri između izvorišta i odredišta moraju biti konfigurisani da podržavaju multikast protokol, da bi odredili gde se nalaze odredišta da bi pakete prosledili samo na te interfejsse
- Prijemni host se prijavljuje u grupu registracijom na lokalnom ruteru
- PIM i IGMP – multikast protokoli koji omogućavaju da saobraćaj bude usmeren i izolovan samo na segmente gde se nalaze prijemni hostovi



Mane

- Posledica korišćenja UDP:
 - Nedostatak pouzdanog prenosa
 - Nedostatak kontrole zagušenja
 - Opasnost od dobijanja paketa u pogrešnom redosledu
 - Opasnost od dobijanja duplikata
- Teško se ostvaruje sigurnost



Multikast adrese

28 bits

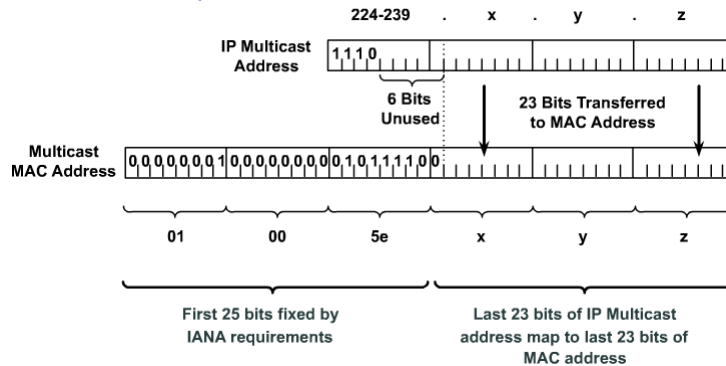
| | | | | | |
|---------|---|---|---|---|--------------------|
| Class D | 1 | 1 | 1 | 0 | Multicast Group ID |
|---------|---|---|---|---|--------------------|

- D klasa: 224.0.0.0 – 239.255.255.255
- Multikast adrese mogu biti dinamički ili statički alocirane
- Dinamičke adrese
 - dodeljuju se na zahtev
 - aplikacija adrese traži i koristi svaki put
 - imaju ograničeno vreme validnosti
- Statičke adrese
 - dodeljuju se jednom, stalno se koriste
- IANA (*Internet Assigned Numbers Authority*) upravlja dodelom adresea
- Permanentne multikast adrese (permanent host groups)
 - slično konceptu dobro poznatih TCP i UDP portova



Mapiranje multikast adresa u MAC

- Odluka iz prvobitne faze razvoja multikasta:
 - samo 23 bita za multikast MAC adrese (IP – 28 bita)
 - opseg 0100.5e00.0000 - 0100.5e7f.ffff
 - Rezultat: mapira se 28 bita IP adrese u 23 bita MAC adrese



Multikast adrese

| Description | Range |
|-----------------------------|------------------------------|
| Reserved link local address | 224.0.0.0 to 224.0.0.255 |
| Globally scoped addresses | 224.0.1.0 to 238.255.255.255 |
| Source specific multicast | 232.0.0.0 to 232.255.255.255 |
| GLOP addresses | 233.0.0.0 to 233.255.255.255 |
| Limited scope addresses | 239.0.0.0 to 239.255.255.255 |



Reserved link local addresses

- Rezervisane adrese za lokalne linkove
- Adrese u opsegu 224.0.0.0 - 224.0.0.255:
- rezervisane za protokole na lokalnom mrežnom segmentu
- **Ruteri ne bi trebali da prosleđuju pakete sa ovim adresama**
- Zato se obično postavlja TTL=1 (*Time to Live*)
- Mrežni protokoli koriste ove adrese za automatsko otkrivanje rutera i za razmenu bitnih podataka:
 - 224.0.0.1 - All local hosts
 - 224.0.0.2 - All local routers
 - 224.0.0.4 - DVMRP
 - 224.0.0.5 – OSPF hello
 - 224.0.0.6 - Designated Router OSPF
 - 224.0.0.9 - RIP2



Globally scoped multicast addresses

- Globalno raspoložive multikast adrese
- 224.0.1.0 - 238.255.255.255
- Mogu se globalno koristiti na Internetu
- Neke od ovih adresa su registrovane kod IANA
 - npr. 224.0.1.1 rezervisana za NTP - *Network Time Protocol*



Source specific multicast addresses

- SSM - *Source Specific Multicast*
- Adrese u opsegu 232.0.0.0 - 232.255.255.255
- SSM je ekstenzija Protocol Independent Multicast (PIM)
- Tradicionalni multikast:
 - prijemnici se prijave za određenu grupu G
- SSM:
 - prijemnici se prijave za određenu grupu G sa izvorištem S



GLOP addresses

- Adrese u opsegu 233.0.0.0 - 233.255.255.255
- Može se koristiti statički od strane organizacija koje imaju registrovan AS broj (autonomni sistem)
- Drugi i treći bajt (oktet) se odnose na AS broj
 - primer: AS 62010
 - 62010 = F23A hex
 - F2 hex = 242
 - 3A hex = 58
 - GLOP adresa = 233.242.58.0/24
 - Ova adresa se koristi u okviru AS 62010 i rutira se globalno kroz Internet Multicast Backbone



Limited scope addresses

- Ograničeni opseg adresa
- Adrese u opsegu 239.0.0.0 to 239.255.255.255
- Koriste se u okviru jedne organizacije ili jednog administrativnog domena
- Ne prosleđuju se na Internet
- Na ruterima su obično konfigurisani filteri da ne propuštaju ove adrese van administrativnog domena
- Unutar domena ove adrese mogu biti podeljena na lokalne pod-domene sa lokalnim multikast granicama
- Ova podela (subdivision) se naziva "address scoping" i mogu biti višestruko korišćenje za pod-domene (RFC 2365 - *Administratively Scoped IP Multicast*)



Multikast sesije

- Multikast sesije se oglašavaju pomoću Session Description Protokola (SDP) ili Session Announcement Protokola (SAP).
- SAP periodično oglašava multikast grupe korišćenjem dobro poznate adrese 224.2.127.254 UDP paketima na portu 9875
- SDP daje opis oglašanih multikast grupa. SDP se koristi i za druge primene, za SIP IP telefoniju



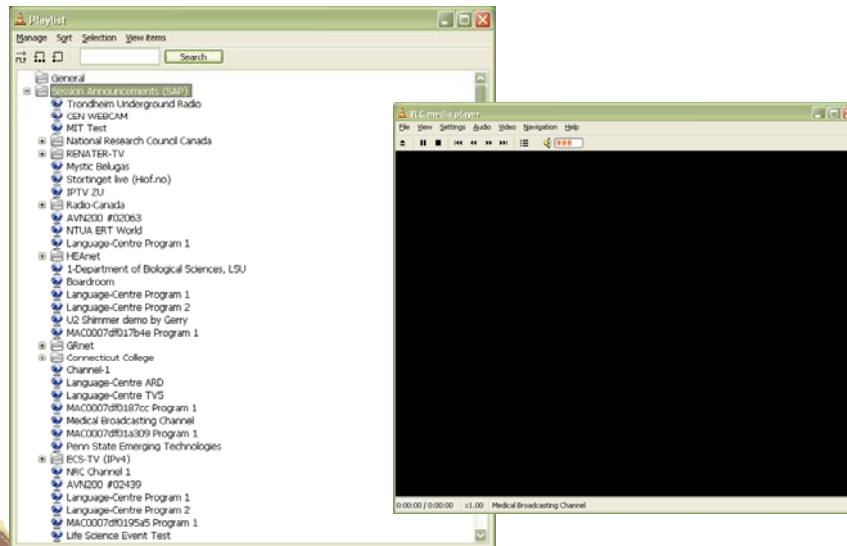
SDP - primer

```
v=0
o=jku2 0 0 IN IP4 213.20.128.35
s=session
c=IN IP4 213.20.128.35
b=CT:1000
t=0 0
m=audio 54742 RTP/AVP 97 111 112 6 0 8 4 5 3 101
a=rtpmap:97 red/8000
a=rtpmap:111 SIREN/16000
a=fmtp:111 bitrate=16000
a=rtpmap:112 G7221/16000
a=fmtp:112 bitrate=24000
a=rtpmap:6 DVI4/16000
a=rtpmap:0 PCMU/8000
a=rtpmap:4 G723/8000
a=rtpmap: 3 GSM/8000
a=rtpmap:101 telephone-event/8000
a=fmtp:101 0-16
```



ETF, Katedra za

17



ETF, Katedra za racunarsku tehniku i informatiku

18

Mbone tools

- WB – white board
- VIC – Video tool
- RAT - Audio tool

The screenshot displays a desktop environment with several windows related to Mbone tools. The central window is a chat application titled "Chat service for the Sunny Virtual Room" which features a whiteboard. On the whiteboard, the mathematical expression $\chi^{(p)} = \sum_k \frac{\hat{h}_{k,p} - \hat{h}_k}{\epsilon_{k,p} - \epsilon_k}$ is written, along with the handwritten note "quite useful" circled in blue. To the right, a window shows a list of participants in the virtual room, including "Robert Stoy, RUS, Stuttgart" and "Gary Stringer (Exeter, UK)". Below the chat window, there is a "Public Sessions" directory listing various sessions like "DUTHnet", "EEIGM - Test Session", and "FAU-TV". Other visible windows include "VLC media player" and "RAI: Sunny Virtual Room".

Internet Group Management Protocol - IGMP

- IGMP se koristi za registraciju pojedinačnih korisnika (primaoci, hostovi) za određene multikast grupe
- Služi da bi ruteri znali da su u određenim granama primaoci kojima treba da se prosleđuje multikast saobraćaj
- IGMP poruke su IP datagrami
 - protocol = 2
 - DST = 224.0.0.2 (All Local Routers)
 - TTL = 1
- Primalac šalje JOIN poruku lokalnom ruteru (jednom ili više)
- Ruteri permanentno osluškiju JOIN poruke
- Ruteri periodično (npr. 60s) šalju IGMP upite (Membership Query) da otkriju koje grupe su aktivne na pojedinačnim segmentima
- Hostovi koji pripadaju multikast grupama, kada prime IGMP upit, odgovaraju ruteru kojim grupama pripadaju
- Hostovi dodaju slučajno izabrano vreme, da ne bi svi istovremeno odgovorili ruteru
- Postoje tri verzije IGMP – 1,2 i 3



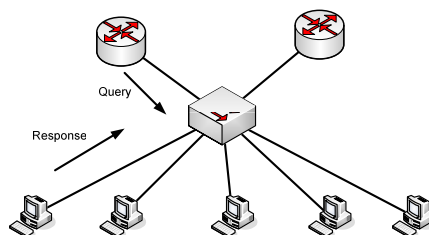
IGMPv1

| | | | |
|---------------|------|--------|----------|
| Version=1 | Type | Unused | Checksum |
| Group address | | | |

- IGMP je enkapsuliran u IP, Protocol number 2
- TTL je uvek 1
- Vrste poruka: Membership Query, Membership report
- U IGMPv1, ruteri periodično šalju Membership query na multikast adresu 224.0.0.1.
- Slušaoci šalju Membership report na adresu multikast grupe kojoj žele da se pridruže.
- Slušaoci tiho napuštaju multikast grupu



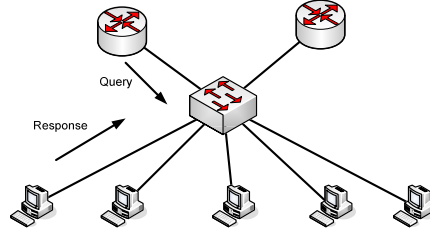
IGMPv1



- IGMP Query/Response proces:
 - Ruteri šalju Query svakih 60s na 224.0.0.1 (All Hosts)
 - Host X, slušalac grupe M.M.M.M šalje Report na M.M.M.M
 - Ako ima drugih slušalaca M.M.M.M oni potiskuju (suppress) svoje Report poruke (neće ih slati)
 - Host Y, slušalac grupe N.N.N.N šalje Report na N.N.N.N
 - Ako ima drugih slušalaca N.N.N.N oni potiskuju svoje Report poruke



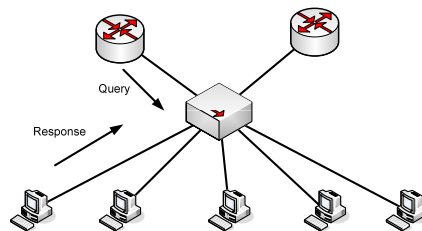
IGMPv1



- Potiskivanje Report poruka:
 - Svaki host po dobijanju Query poruke startuje Report timer (max 10s) i šalje svoj Report posle slučajnog vremena koje je kraće od Report vremena. Ako pre isteka ovog vremena primeti Report poruku, neće poslati svoju.
- IGMP Querier:
 - Ruter koji šalje Query pakete
 - Ne postoji mehanizam izbora
 - Zasniva se na L3 informaciji (designated router)



IGMPv1 – Join & leave procesi



- Join: Ako host želi da sluša neku grupu šalje Unsolicited Report sa adresom grupe M.M.M.M koju želi da prima
- Leave: Ne postoji Leave poruka – ako nema odgovora na Query poruku posle 3 puta Query interval (3 min) smatra se da nema slušalaca date grupe



IGMPv2

| Type | Max resp time | Checksum |
|---------------|---------------|----------|
| Group address | | |

- Postoji proces izbora Querier rutera
- Group specific Query
- Leave group messages
- Maximum response time (MRT) field (10s za General, 1s za Group Specific Query)
- Poruke:
 - General Query (Adresa grupe – sve nule)
 - Group specific Query (group address)
 - V1 membership report
 - V2 membership report
 - Leave Group



IGMPv2 – Izbor Queriera

- Kada se na nekoj mreži uključi ruter koji podržava IGMPv2, šalje General Query na All Hosts
- Kada IGMPv2 ruter primi General Query, poredi IP adresu izvora sa sopstvenom IP adresom
- Ruter sa NAJNIŽOM IP adresom postaje Querier



IGMPv2 Leave proces

- RFC: “A host MAY send a Leave group message when it leaves a group”
- Host šalje Leave poruku na AllRouters multicast adresu (224.0.0.2) sa oznakom grupe koju napušta
- Router prima Leave, i mora da proveriti da li postoje drugi slušaoci date grupe na istom mrežnom segmentu – šalje Group Specific Query
- Ako nema drugih slušalaca, ruter prestaje sa slanjem multicast saobraćaja po datoj grupi posle 2xMRT



IGMPv3 - RFC 3376

- Najvažnija razlika u odnosu na prethodne verzije je mogućnost primanja Source-Specific Multicast saobraćaja. IGMPv3 adds the ability to filter multicasts based on the multicast source.
- Kod IGMPv3, Report poruke se šalju na 224.0.0.22 umesto na 224.0.0.2.



IGMP verzije

| IGMP Version | Attributes and Limitations |
|--------------|---|
| v1 | Single group membership per IGMP message. No group specific query, no leave group message. |
| v2 | Single group membership per IGMP message. Could not specify source host from which the multicast flow is desired. |
| v3lite | Cisco proprietary for SSM applications on hosts without v3 support. |

IGMP u IPv6 = MLD – Multicast Listeners Discovery

MLDv1 ≈ IGMPv2

MLDv2 ≈ IGMPv3



IGMP snooping

- Predefinisano ponašanje za svičeve (Layer 2) je da multikast saobraćaj prosleđuju na svaki port VLAN-a na kome ima primaoca
- IGMP snooping
 - svičevi ne gledaju samo L2 zaglavlje, već i podatke
 - detektuju se IGMP paketi koji pristižu na određene portove
 - multikast saobraćaj se prosleđuje samo na one portove na kojima su detektovane IGMP poruke



Multikast protokoli rutiranja

- Protokoli rutiranja multikast saobraćaja:
 - DVMRP – *Distance Vector Multicast Routing Protocol* (ekstenzija RIP-a)
 - MOSPF – *Multicast Open Shortest Path First* (ekstenzija OSPF-a)
 - CBT – *Core Based Trees v1,v2* (RFC 2189)
 - PIM – *Protocol Independent Multicast* – nezavistan od unicast protokola rutiranja

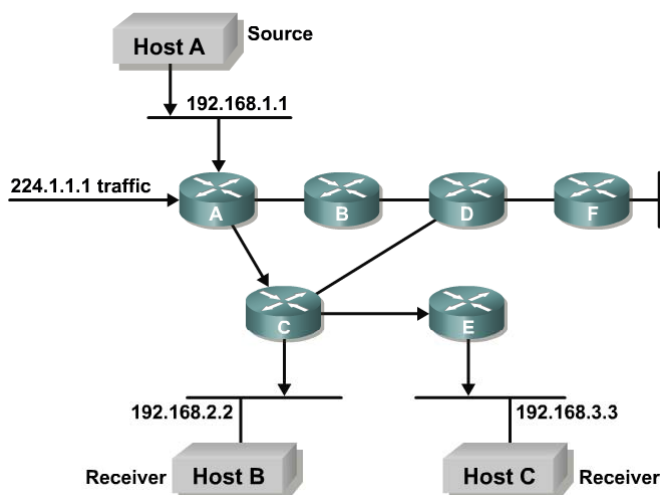


Source Distribution Trees

- Jednostavna forma stabla distribucije multikast saobraćaja
- Izvorište je u korenu stabla, a grane su prema prijemnicima
- Grane se formiraju prema pravilu "najkraćeg puta", pa se često zove i *Shortest Path Tree* - SPT
- Notifikacija:
 - (S, G) – izgovara se "S comma G"
 - S je IP adresa izvorišta
 - G IP adresa multikast grupe kojoj prijemnici pripadaju
 - Primer (192.168.1.1, 224.1.1.1)
- **Unicast IP adrese prijemnika su irelevantne**
- Notacija (S, G) implicira da se za svako izvorište kreira posebno SPT stablo
- Primer:
 - Ako host B, sa adresom 192.168.2.2, takođe emituje saobraćaj ka grupi 224.1.1.1, tada će da postoji novo stablo u oznaci (192.168.2.2, 224.1.1.1)



Source Distribution Trees

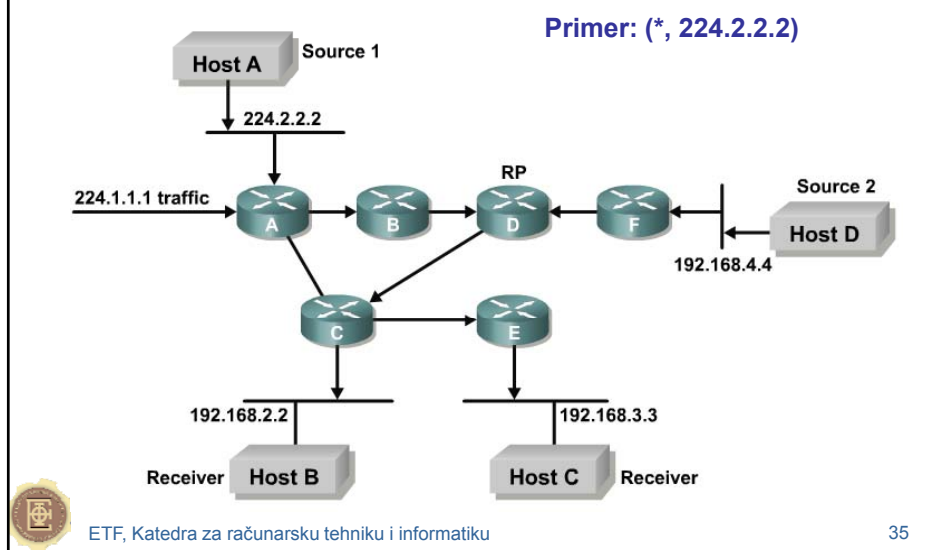


Shared Distribution Tree

- Shared Distribution Tree koristi jedinstveni zajednički koren koji je postavljen u unapred izabranu tačku (ruter)
- Zajednički koren stabla se zove **Rendezvous Point (RP)**
- Princip rada:
 - Sva izvorišta šalju saobraćaj prema RP
 - Saobraćaj se dalje od RP prenosi do svih prijemika
 - Ako se prijemnik nalazi između izvorišta i RP, prijemnik će direktno preuzeti paketa
- Pošto sva izvorišta jedne multikast grupe dele zajedničko stablo, notacija je:
 - (*, G) - izgovara se "star comma G"



Shared distribution tree



Unidirekciona i bidirekciona deljena stabla

- Bidirekciono stablo – multikast saobraćaj ide od izvora ka RP, a zatim od RP ka slušaocima – u oba smera.
- Unidirekciono stablo – multikast saobraćaj se ka RP transportuje ili kao unicast (IP in IP) ili kao SPT multikast tako što se RP učlani u SPT grupu izvora S (koristi se kod PIM), a zatim se kao multikast distribuira do slušalaca



Source vs. Shared trees

- Članovi multikast grupa mogu da pristupe ili napuste grupu u svakom trenutku – Distributivno stablo mora da se dinamički održava
- Prijavljivanje u grupu:
 - novi prijemnik od rutera zahteva prijem multikast saobraćaja, ruter pripadajuću granu dinamički uključuje u distributivno stablo i počinje da prosleđuje multikast saobraćaj u tu granu – prijemnik tada postaje aktivan
- Odjavljivanje iz grupe:
 - aktivni prijemnik ruteru šalje poruku za odjavljivanje (“prune”), na osnovu koje ruter uklanja pripadajuću granu iz distributivnog stabla i ne prestaje sa slanjem multikast saobraćaja u tu granu – prijemnik tada postaje neaktivan
- **Source trees**
 - **Prednost:** optimalno stablo, najkraće putanje od izvorišta
 - **Problem:** skalabilnost – različita stabla za svako izvorište
- **Shared trees**
 - **Prednost:** skalabilnost - minimalna količina podataka u svakom ruteru
 - **Problem:** neoptimalne putanje od izvorišta do prijemnika (preko deljenog RP)



Reverse Path Forwarding - RPF

- Prenos saobraćaja:
 - aspekta posmatranja unicast saobraćaja je “prema odredištu”
 - aspekta posmatranja multikast saobraćaja je “od izvorište”
- Proces utvrđivanja puta “od izvorišta” do posmatranog rutera se naziva *Reverse Path Forwarding* - RPF
- Ruteri kreiraju stablo distribucije multikast saobraćaja (distribution trees) koje kontroliše prenos paketa od izvorišta do svih prijemnika
- Stablo distribucije može biti:
 - source distribution trees
 - shared distribution trees
- Vrsta stabla distribucije zavisi od primenjenog tipa multikast protokola

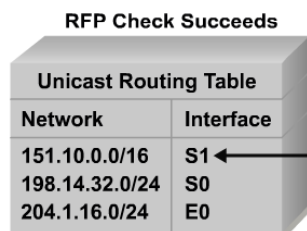
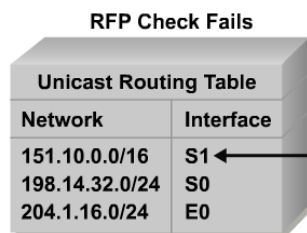


RPF provera

- Služi da se izbegnu petlje u distributivnom stablu
- Izvor emituje saobraćaj prema članovima grupe, na osnovu multikast adrese
- Multikast ruter odlučuje koji je pravac (interfejs) usmeren prema izvorištu (*upstream*), a koji prema primaocima (*downstream*)
- RPF koristi unicast IP adresu izvorišta da odredi na koji interfejs će da pristignu validni multikast paketi
- Kada multikast paketi pristignu, ruter sprovodi RPF proveru:
 - ako su paketi stigli sa *upstream* interfejsa provera je uspešna – paketi se prosleđuju na sve *downstream* interfejse
 - u suprotnom, provera je nesuđna, paketi se odbacuju
- U slučaju više potencijalnih pravaca do odredišta bira se pravac sa najmanjom težinom – na osnovu tabele rutiranja



RPF provera



PIM

- **Protocol Independent Multicast (PIM)** – odlučuje na koje portove da prosleđuje pristigle multikast paketa (packet-forwarding decisions)
- PIM koristi postojeću unicast tabelu rutiranja
- Ne koristi se posebna razmena ruting informacija između rutera
- PIM može da radi u tri moda:
 - **PIM Dense Mode (PIM-DM)**
 - **PIM Sparse Mode (PIM-SM)**
 - **PIM Sparse-Dense Mode**



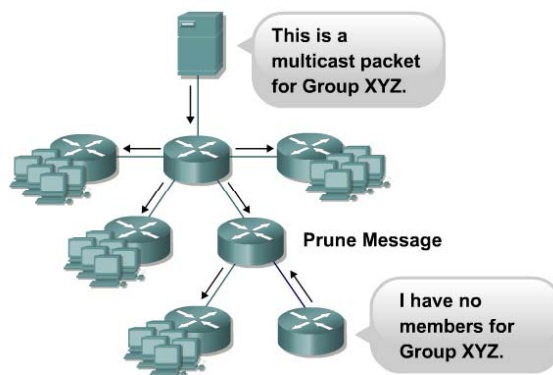
PIM

- PIM ruteri ostvaruju susedske odnose, kako bi se znalo preko kojih interfejsa treba da se šalje multikast saobraćaj
- SPT stablo kod PIM se pravi samo od onih linkova na kojima su PIM susedi ili multikast slušaoci
- Ako postoji više ruta ka source-u na datom ruteru, RPF proveru će proći samo jedna - ona ka Next Hopu sa najvećom adresom
- PIM DR – ruter sa najvećom adresom na datom segmentu – ako se koristi IGMPv1 ovaj ruter će postati Query Router
- Od verzije PIMv2 – postoji i prioritet za izbor DR. Default vrednost je 1



PIM Dense Mode

- Push model – multikast paketi se prosleđuju u sve delove mreže svaka tri minuta
- Ruteri u mreži mogu da otkazu pojedine tokove (grane) ako nemaju registrovane multikast korisnike za pojedinačne grupe
- Flood and Prune

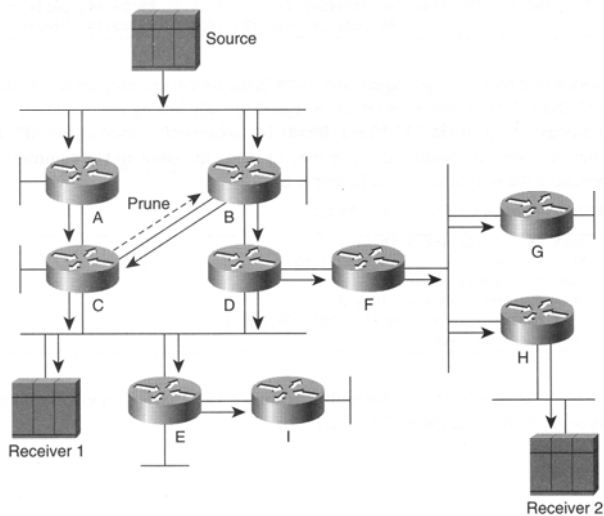


PIM DM Pruning

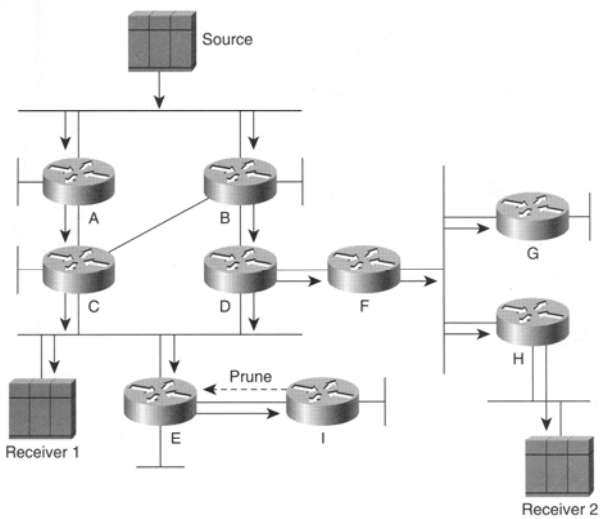
- Ako paket dođe preko ne-RPF point-to-point interfejsa
- List-ruter nema slušalaca date grupe
- Ne-list ruter je dobio Prune poruku preko point-to-point segmenta
- Ne-list ruter je dobio Prune poruku preko LAN segmenta i nije dobio Prune override (Join)



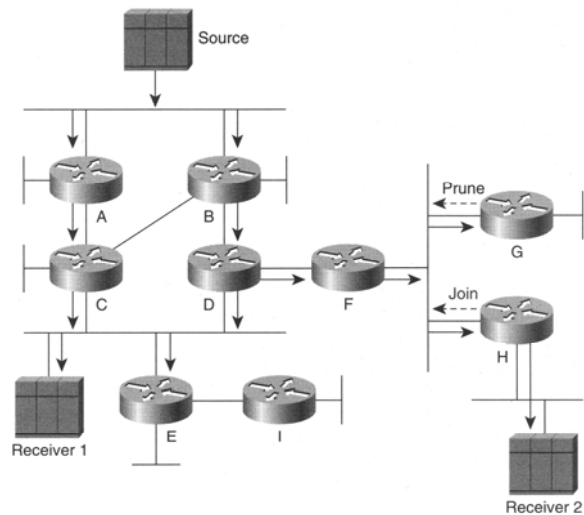
PIM Dense Mode



PIM Dense Mode

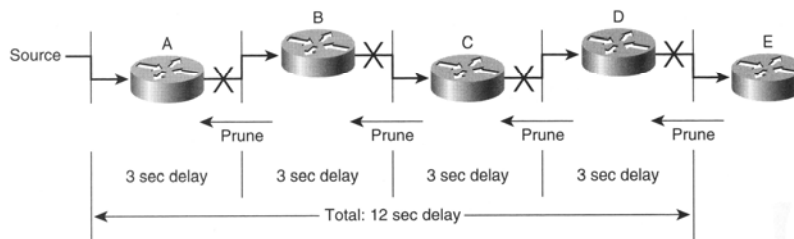


PIM Dense Mode

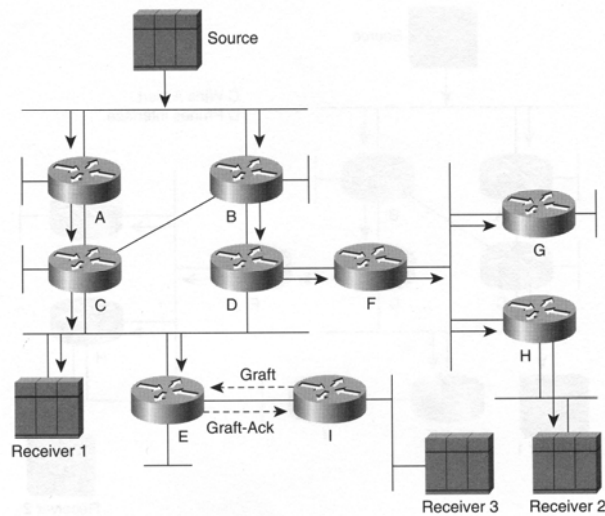


PIM Dense Mode

- Da bi se sačekalo na Prune override, postoji 3s dug Prune timer
- Zbog ovoga postoji akumulacija kašnjenja i relativno sporo se gasi neželjeni multikast tok

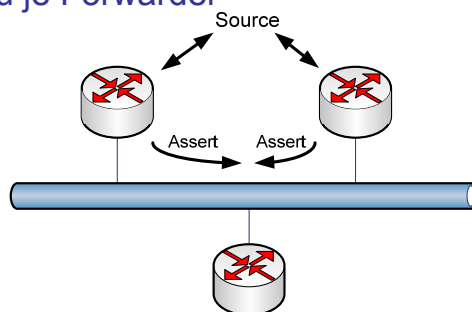


PIM Dense Mode - Graft



PIM DM Assert mehanizam

- PIM DM Assert: Izbor rutera koji će proslediti saobraćaj ka LAN segmentu
- Ruter sa boljom metrikom je Forwarder
- Ako je nerešeno – ruter sa većom IP adresom na segmentu je Forwarder



PIM Dense Mode

- Efikasan metoda kada se podrazumeva:
 - mali broj pošiljaoca, a veliki broj aktivnih primaoca
 - mala udaljenost između pošiljaoca i primaoca
 - veliki intenzitet multikast saobraćaja
- Periodično slanje na sve grane (flooding) i ponovno odjavljivanje, da bi se održalo stablo distribucije



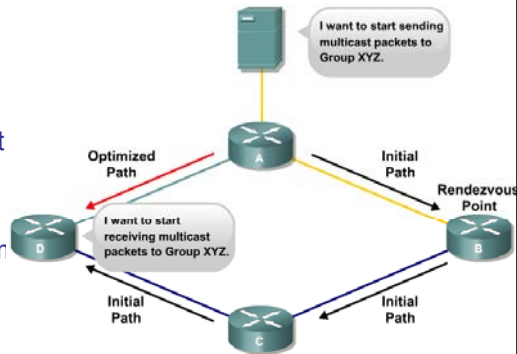
PIM Sparse Mode

- Pull model – primaoci eksplicitno zahtevaju isporuku multikast saobraćaja od rutera
- Saobraćaj se šalje samo na segmente gde ima aktivnih i registrovanih primaoca
- PIM-SM distribuira multikast saobraćaj po deljenom distributivnom stablu (shared tree):
 - pošiljoci šalju pakete do RP (*rendezvous point*)
 - RP prosleđuje pakete do primaoca

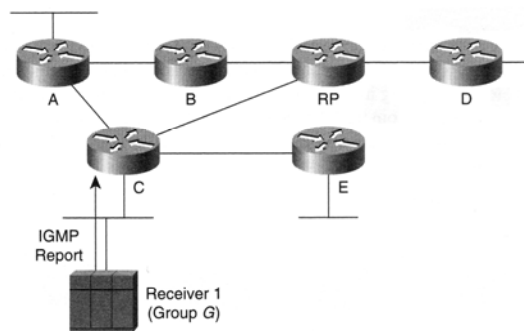


PIM SM

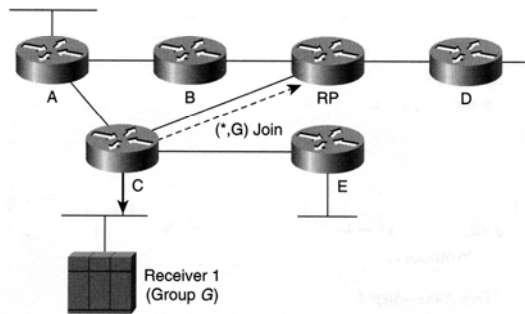
- Pretpostavke
 - relativno malo primaoca, koji su distribuirani po mreži
 - propusni opseg nije široko dostupan
- PIM SM počinje sa praznim distributivnim stablom
- Grane distributivnog stabla se postepeno dodeljuju kao rezultat eksplicitnog prijavljivanja prijemnika
- Kada ruter primeti da ima registrovanih primaoca na svojim interfejsima (povezanim granama) on se registruje u deljenom stablu prema RP



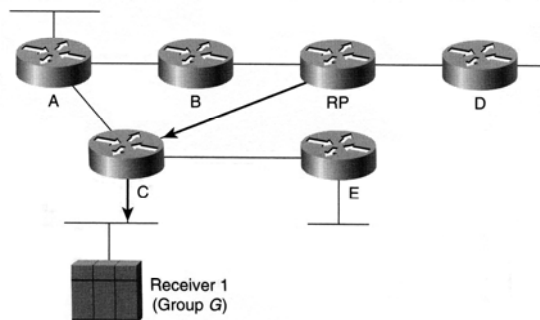
PIM Sparse Mode - 1



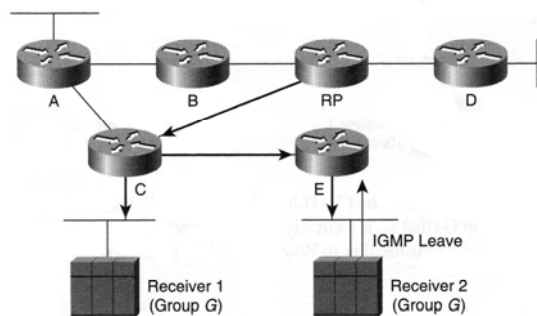
PIM Sparse Mode - 2



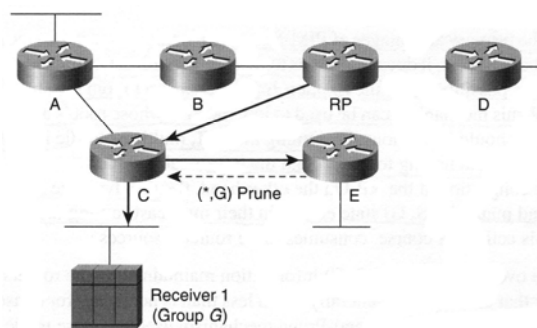
PIM Sparse Mode - 3



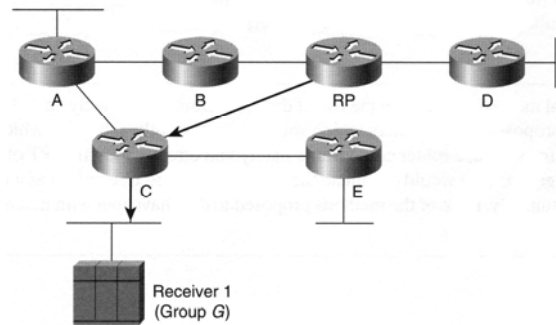
PIM Sparse Mode - 4



PIM Sparse Mode - 5



PIM Sparse Mode - 6

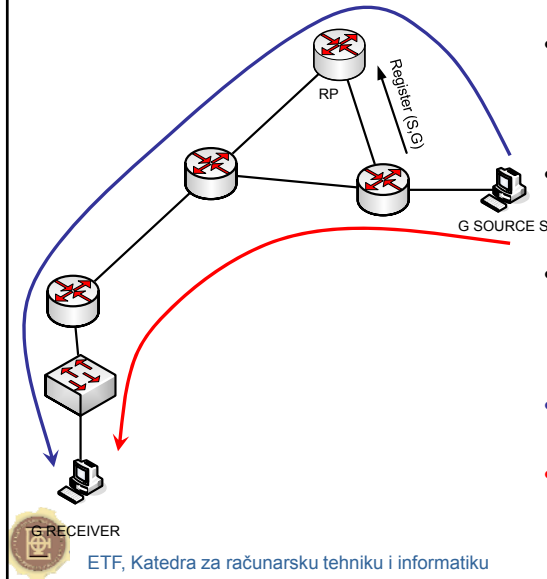


PIM SM – Join/Prune procesi

- Join i Prune su definisani istom porukom.
- Jedna Join/Prune poruka može da nosi informacije o većem broju grupa.
- Ruteri moraju da periodično šalju (1 min) informacije o svim grupama do kojih treba da ide multikast (osvežavanje – State Refresh)
- Ako posle 3 min nema Join poruke za neku grupu, njeno stanje se briše



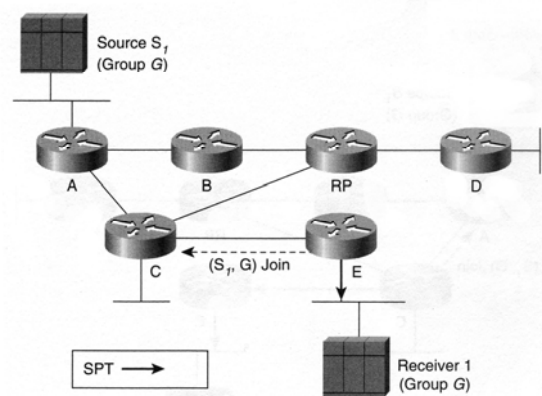
PIM SM – Source registration



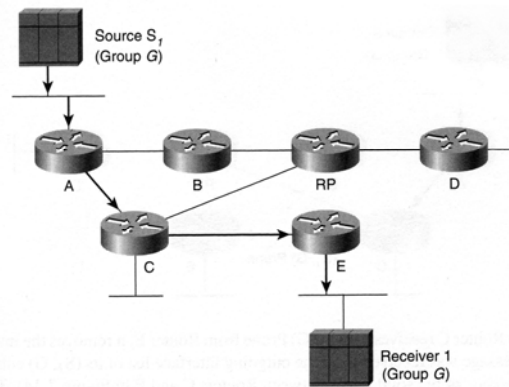
- Saobraćaj izvora se enkapsulira u Register unicast poruke (to radi DR)
- Ako ima slušalaca grupe G, RP šalje Join prema S i kreira SPT (S,G)
- Ako nema slušalaca grupe G, RP unikastom šalje Register-Stop prema S.
- Multicast saobraćaj ide kroz RP
- PIM SM može da prebaci tok na SPT

PIM Sparse Mode – SPT Join

Nakon što dobije prvi multikast paket od izvora S1, ruter E može da traži prebacivanje na SPT

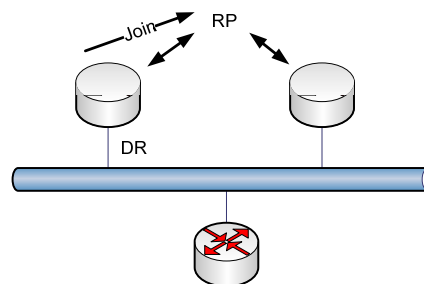


PIM Sparse Mode - SPT



PIM SM DR, RP discovery

- DR šalje Join poruke ako ima više od jednog rutera na LAN segmentu
- RP može da se konfigurira ručno
- RP može da se otkrije automatski kroz:
 - Auto-RP
 - Bootstrap Router BSR
 - Anycast RP with MSDP



Auto-RP

- Proprietary tehnika
- Dva entiteta – RP i RP Mapping Agent
- RP se oglašava RP-Announce porukama na 224.0.1.39. U tim porukama je i skup grupa koje podržava
- RP MA šalje RP Discovery poruke na 224.0.1.40 kojima oglašava RP i koje grupe oni podržavaju
- Problem: Ako se koristi PIM SM, a ruter nije naučio gde je RP, kako će dobiti RP Discovery poruke? – PIM SDM



Bootstrap Router - BSR

- BSR dobija informacije o RP ruterima u domenu i njihovim grupama unikastom (c-RP Advertisement poruke)
- BSR šalje informaciju o mapiranju RP-grupa na 224.0.0.13, a ruteri ih dalje prosleđuju ostalim ruterima u mreži – ne postoji problem kao kod Auto-RP
- Može da bude i više BSR rutera.



Anycast RP with MSDP

- Omogućava redundansu u multikast domenu
- RP se statički konfigurira Anycast adresom
- RP šalje RP Announce, RP MA ih oglašava
- Pošto svaki RP pravi svoje deljeno stablo, a saobraćaj ide od izvora samo ka jednom od RP-ova, RP međusobno razmenjuju MSDP poruke kojima se obaveštavaju o aktivnim izvorima



PIM SM DM

- Kombinovani mod dozvoljava da se pojedinačne multikast grupe odzivaju u DM ili SM modu, u zavisnosti da li je dostupna informacija o RP (rendezvous point):
 - Ako ruter prikupi informaciji o RP za određenu grupu, koristiće se SM,
 - u suprotnom slučaju će se koristiti DM
- PIM SM-DM omogućava da se konfiguriraju grupe koje koriste SM i grupe koje koriste DM
- RP treba da je konfigurisan (za SM)



PIM versions

- PIM verzija 1 je Cisco proprietary
- PIM verzija 2 je standardizovana od strane IEEE i podrazumeva sledeća poboljšanja:
 - može da postoji jedan aktivni RP po grupi, sa više backup RP
 - *bootstrap router* (BSR) omogućava fault-tolerant i automatsko otkrivanje RP. Ruteri dinamički saznaju mapiranje grupa u RP
 - SM i DM su modovi vezani za grupe, a ne za interfejsse kao u V1
 - PIM Join i Prune poruke su fleksibilnije
 - PIM paketi su posebni paketi, koji nisu enkapsulirani u IGMP paketa, kao kod V1



Multicast routing table

```
Switch#show ip mroute [hostname | group_number]
```

- Displays the contents of the IP multicast routing table

```
Switch#show ip mroute

IP Multicast Routing Table
Flags: D - Dense, S - Sparse, C - Connected, L - Local, P - Pruned
       R - RP-bit set, F - Register flag, T - SPT-bit set
Timers: Uptime/Expires
Interface state: Interface, Next-Hop, State/Mode

(*, 224.0.255.3), uptime 5:29:15, RP is 198.92.37.2, flags: SC
  Incoming interface: FastEthernet 1/1, RPF neighbor 10.3.35.1, Dvmrp
  Outgoing interface list:
    Ethernet0, Forward/Sparse, 5:29:15/0:02:57

(198.92.46.1, 224.0.255.3), uptime 5:29:15, expires 0:02:59, flags: C
  Incoming interface: FastEthernet 1/1, RPF neighbor 10.3.35.1
  Outgoing interface list:
    Ethernet0, Forward/Sparse, 5:29:15/0:02:57
```



Multicast routing table

```
Switch#show ip mroute summary

IP Multicast Routing Table
Flags: D - Dense, S - Sparse, C - Connected, L - Local, P - Pruned
      R - RP-bit set, F - Register flag, T - SPT-bit set, J - Join
SPT
Timers: Uptime/Expires
Interface state: Interface, Next-Hop, State/Mode

(*, 224.255.255.255), 2d16h/00:02:30, RP 171.69.10.13, flags: SJPC
(*, 224.2.127.253), 00:58:18/00:02:00, RP 171.69.10.13, flags: SJC
(*, 224.1.127.255), 00:58:21/00:02:03, RP 171.69.10.13, flags: SJC

(*, 224.2.127.254), 2d16h/00:00:00, RP 171.69.10.13, flags: SJCL
(128.9.160.67/32, 224.2.127.254), 00:02:46/00:00:12, flags: CLJT
(129.48.244.217/32, 224.2.127.254), 00:02:15/00:00:40, flags: CLJT
(130.207.8.33/32, 224.2.127.254), 00:00:25/00:02:32, flags: CLJT
(131.243.2.62/32, 224.2.127.254), 00:00:51/00:02:03, flags: CLJT
(140.173.8.3/32, 224.2.127.254), 00:00:26/00:02:33, flags: CLJT
(171.69.60.189/32, 224.2.127.254), 00:03:47/00:00:46, flags: CLJT
```



Multicast routing table

```
Switch#show ip mroute active

Active IP Multicast Sources - sending >= 4 kbps

Group: 224.2.127.254, (sdr.cisco.com)
  Source: 146.137.28.69 (mbone.ipd.anl.gov)
    Rate: 1 pps/4 kbps(1sec), 4 kbps(last 1 secs), 4 kbps(life avg)

Group: 224.2.201.241, ACM 97
  Source: 130.129.52.160 (webcast3-e1.acm97.interop.net)
    Rate: 9 pps/93 kbps(1sec), 145 kbps(last 20 secs),
    85 kbps(life avg)

Group: 224.2.207.215, ACM 97
  Source: 130.129.52.160 (webcast3-e1.acm97.interop.net)
    Rate: 3 pps/31 kbps(1sec), 63 kbps(last 19 secs),
    65 kbps(life avg)
```



Multicast routing table

```
Switch#show ip mroute count

IP Multicast Statistics - Group count: 8,
Average sources per group: 9.87
Counts: Pkt Count/Pkts per second/Avg Pkt Size/Kilobits per second

Group: 224.255.255.255, Source count: 0, Group pkt count: 0
RP-tree: 0/0/0/0

Group: 224.2.127.253, Source count: 0, Group pkt count: 0
RP-tree: 0/0/0/0

Group: 224.1.127.255, Source count: 0, Group pkt count: 0
RP-tree: 0/0/0/0

Group: 224.2.127.254, Source count: 9, Group pkt count: 14
RP-tree: 0/0/0/0
Source: 128.2.6.9/32, 2/0/796/0
Source: 128.32.131.87/32, 1/0/616/0
Source: 128.125.51.58/32, 1/0/412/0
Source: 130.207.8.33/32, 1/0/936/0
Source: 131.243.2.62/32, 1/0/750/0
```



Multicast rutiranje između domena

- Moraju biti ispunjeni sledeći uslovi:
 1. RP u domenima u kojima postoje slušaoci, moraju znati adrese aktivnih izvora u drugim domenima.
 2. Svi ruteri na putanji od izvora do odredišta moraju imati adresu izvora u svojoj ruting tabeli koju koriste za RPF proveru.
- MSDP (Multicast Source Discovery protocol) ispunjava prvi uslov.
- Drugi uslov obično ne predstavlja problem jer većina mreža ima any-any ip povezanost za unicast saobraćaj. Za RPF proveru se može koristiti unicast routing tabela, ili posebna multicast ruting tabela koja mora imati ulaze za svaki potencijalni multicast izvor.
- MBGP ima ulogu da popunjava multicast ruting tabelu.



MSDP

- Omogućava da *RP*-ovi razmenjuju informacije o aktivnim izvorima
- Alternative koje su postojale pre *MSDP*-a :
 - Centralizovani *RP* koga dele svi *ISP*-ovi (skalabilnost, *SPoF*, ko upravlja? ...)
 - Povezivanje *RP*-a preko multiaccess interfejsa na tačku za razmenu informacija o multikast izvorima. Ovaj interfejs bi bio konfigurisan u *PIM-DM*, tako da svi *RP* flood-uju informacije o aktivnim izvorima. Ovakav *RP* bi morao da bude smešten na ivici domena, uz ograničenje da jedan domen može da ima samo jedan *RP*.
(suboptimalno rutiranje, problem skalabilnosti lokalno i globalno..)
- Prednosti *MSDP*-a:
 - Razbija multikast distributivna stabla. Deljeno stablo je lokalno za domen tako da *Join* poruke nikada ne moraju da napuštaju domen.
 - **Ne postoji centralizovani RP**, već svaki domen ima svoj *RP*. Na taj način se povećava otpornost na otkaze i izbegava *SPoF*.
 - **Povećava sigurnost** jer se izvori ne moraju oglašavati van domena.



MSDP - način rada

- *MSDP* ruteri formiraju peer odnose preko TCP konekcije (slično kao *BGP*).
- *MSDP* sesije mogu biti multihop.
- *MSDP* peer-ovi mogu biti iz istog ili iz različitih *PIM-SM* domena. Ako su iz istog domena, omogućavaju postojanje više *RP* unutar domena, čime se postiže redundantnost i load-balancing.
- Čvor sa *MSDP* funkcionalnošću ne mora da bude *RP*. Ovakav dizajn je koristan u domenima koji nemaju multikast izvore i slušaoce, već samo pružaju prenos informacija o multikast izvorima do drugih domena.



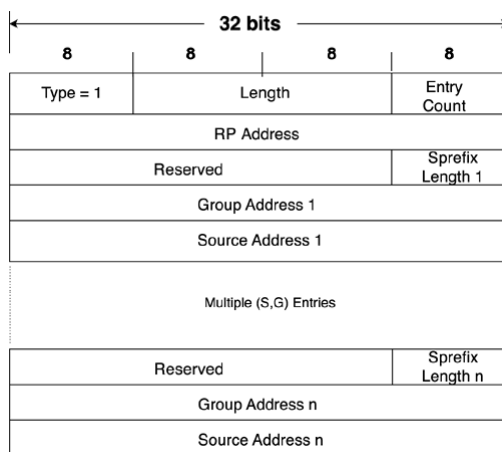
MSDP - način rada

- Kada *RP* sa *MSDP* funkcionalnošću dobije *PIM Register* poruku, on generiše *Source-Active (SA)* poruku za (S,G) par, i prosleđuje je svim svojim *MSPD* peer-ovima.
- SA poruka sadrži:
 - adresu izvora
 - adresu grupe
 - adresu RP-a koji je generisao SA poruku
- Dodatno, podake koji su enkapsulirani u prvu *Register* poruku, RP takođe kopira u SA poruku.
- Kasnije *Register* poruke za isti (S,G) par iniciraju slanje nove SA poruke samo po isteku *SA hold-down* tajmera koji je po defaultu podešen na 30 sekundi.
- *RP* za izvore iz svog domena nastavlja periodično da šalje SA poruke sa intervalom od 60s, dok god je izvor aktivan.



Format SA poruke

- type-length values (TLVs) format poruke



- **Length** – ukupna dužina poruke
- **Entry Count** – broj (S,G) parova u telu poruke
- **RP Address** – adresa RP-a u koji je inicirao stvaranje poruke
- **Sprefix** – dužina prefiksa adrese izvora
- **Group Address** – adresa grupe
- **Source Address** – adresa izvora

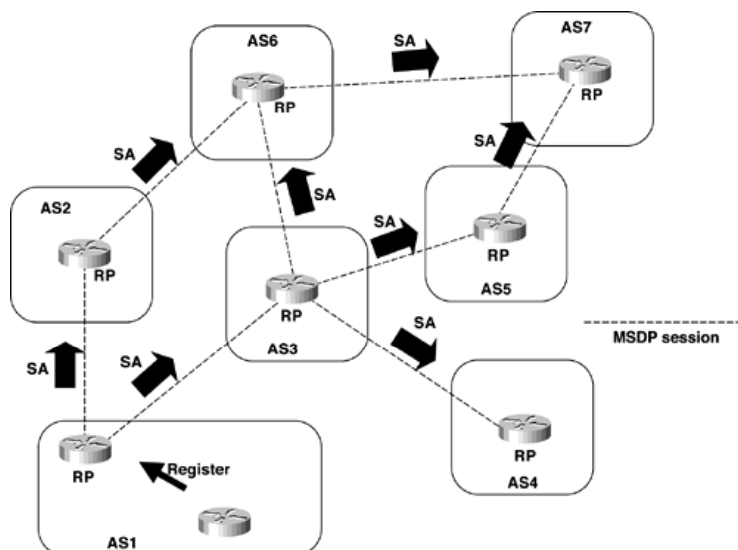


MSDP - način rada

- Kada ruter primi SA poruku, on proverava da li je stigla na interfejs koji prolazi RPF proveru. **Ako nije, SA poruka se odbacuje.** U suprotnom SA poruka se prosleđuje svim MSDP susedima, osim susedu od koga je poruka dobijena.
- RPF se računa **prema adresi RP** enkapsuliranoj u SA poruci. To je zapravo adresa MSDP peer-a koji je generator SA poruke.
- Ovaj proces se zove **peer-RPF flooding**.
- **peer-RPF flooding** garantuje da će SA poruka stići do svih MSDP čvorova, pri čemu se izbegavaju petlje u prosleđivanju.



MSDP - način rada

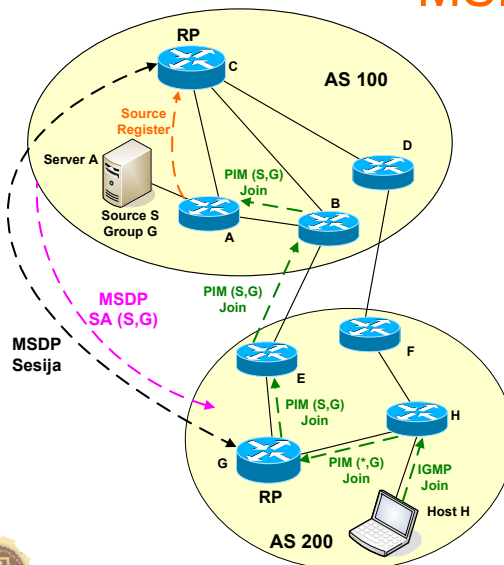


MSDP - način rada

- U slučaju da je ruter koji je primio SA poruku u isto vreme i RP, on obavlja i dodatnu obradu poruke.
- Ako da ima zainteresovanih slušalaca za tu grupu počinje da prosleđuje podatke koji su enkapsulirani u SA poruku niz RPT.
- U isto vreme šalje (S,G) Join poruku prema svom RPF susedu za taj izvor, i gradi **SPT prema izvoru**.



MSDP - način rada



- Server A je izvor multikast saobraćaja za grupu G.
- Host H se prijavljuje kao slušalac za multikast grupu G, tako što šalje **IGMP Join** poruku svom *first-hop* ruteru (ruter H).
- Ruter H gradi **RPT** prema **RP**-u iz svog domena tako što šalje **PIM Join** poruku prema **RP**-u.
- **RP** iz AS 200 i **RP** iz AS 100 su **MSDP peer**-ovi i razmenjuju informacije o aktivnim izvorima u svom domenu.
- Na taj način ruter G zna adresu izvora koji šalje na grupu G.
- Slanjem **PIM (S,G) Join** poruka, formira **SPT** prema izvoru S.



MSDP - način rada

- Ranije verzije MSDP nisu zahtevale da MSDP peer-ovi održavaju globalno source stanje (ruteri nisu radili keširanje SA poruka).
- Kada takav ruter primi *Join* poruku za grupu koja ne postoji u mroute, on šalje **SA Request poruku** svojim MSDP peer-ovima, kojom traži adrese izvora koji emituju saobraćaj na tu multikast grupu.
- MSDP peer-ovi koji rade keširanje SA poruka, na *SA-Request* poruku odgovaraju *SA-Response* porukom koja sadrži listu SA ulaza koji odgovaraju traženoj grupi.
- Keširanje SA poruka smanjuje kašnjenje pri odgovaranju na zahteve za priključenje grupi. U praksi se uvek radi keširanje SA poruka, a na nekim uređajima ni ne postoji mogućnost da se ova opcija isključi.
- Mogu da se keširaju i odbijene SA poruke. Na taj način se olakšava *troubleshooting*.
- Nedostatak keširanja SA poruka je što ruter ima informacije za sve izvore na Internetu. Ovo dovodi u pitanje skalabilnost MSDP-a, i predstavlja glavni nedostatak ovog protokola.



Uspostavljanje MSDP sesije

- MSDP peer sa višom IP adresom, sluša na TCP portu 639 i čeka zahtev za uspostavljanje TCP konekcije, koji inicira peer sa nižom IP adresom.
- Stanja MSDP peer state mašine:
 - **DISABLED**: MSDP peer nije konfigurisan
 - **INACTIVE**: MSDP peer je konfigurisan, ali nije u LISTENING ili CONNECT stanju
 - **CONNECT**: Aktivan peer pokušava da inicira TCP sesiju
 - **LISTEN**: Pasivan peer je konfigurisan i sluša po TCP portu 639
 - **ESTABLISHED**: TCP sesija je uspostavljena



Uspostavljanje MSDP sesije

- Prelazak između stanja za pasivan peer, u regularnom toku:
 1. DISABLED
 2. INACTIVE
 3. LISTEN
 4. ESTABLISHED
- Za aktivan peer, normalan prelazak između stanja:
 1. DISABLED
 2. INACTIVE
 3. CONNECT
 4. ESTABLISHED
- Aktivan peer menja stanja između INACTIVE i CONNECT sve dok se ne uspostavi TCP konekcija sa pasivnim peer-om. Svaki put kada pređe iz CONNECT u INACTIVE stanje čeka 30s pre nego što ponovo pokuša da uspostavi TCP sesiju.

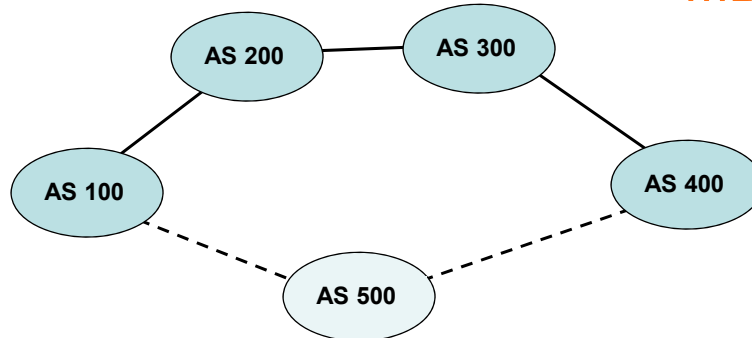


MBGP

- *MBGP* – *Multiprotocol BGP*, nije poseban protokol već ekstenzija *BGP*-ja
- Popunjavanje posebnu multikast tabelu koja se koristi za *RPF* proveru pri prosleđivanju multikast paketa, i odlučivanju da li treba prihvatiti određenu *SA* poruku.
- Korišćenje *MBGP* je **neophodno** u slučaju da **multikast i unicast topologije nisu iste**. Tada next hop za unicast ne mora biti isti kao i za multikast rute.
- Moguće je koristiti *PIM-SM* i *MSDP* bez *MBGP* za multikast rutiranje, sve dok je next hop za svaki prefiks isti za prosleđivanje unikasta i za multikast *RPF*.
- Zbog mogućih promena unicast/multikast topologije, preporučuje se korišćenje *MBGP* od početka.
- *MBGP* se **ne koristi** za pravljenje multikast distributivnih stabala, to je uloga multikast protokola rutiranja, kao što je *PIM-SM*! On ne radi sa multikast adresama. **Uloga *MBGP* je da popunjava posebnu multikast routing tabelu koja se koristi za *RPF* proveru od strane multikast protokola.**



MBGP



- AS 500 nema pušten multikast
- Bez *MBGP*, ruteri u AS100 bi izabrali putanju preko AS500 da stignu do multikast izvora u AS400. Rezultat bi bila crna rupa u multikast rutiranju.
- Sa *MBGP* ruteri iz AS100 bi koristili putanju kroz AS200 i AS300 za multikast saobraćaj (kao rezultat RPF porvere), a AS500 za prosleđivanje unikast paketa.



ETF, Katedra za računarsku tehniku i informatiku

87

MBGP

- Definisani u RFC 2283. Definiše ekstenzije za postojeći BGP protokol, tako da se ne razmenjuju samo IPv4 prefiksi.
- Može da nosi različite vrste ruta :
 - IPv4 prefikse za unikast rutiranje
 - IPv4 prefikse za multikast *RPF* proveru
 - IPv6 prefikse za unikast rutiranje
 - IPv6 prefikse za multikast *RPF* proveru
- Održavaju se posebne *BGP* tabele, prema tome koje se vrste ruting informacije razmenjuju:
 - *Unicast Routing Information Base (U-RIB)* – sadrži prefikse za unikast prosleđivanje
 - *Multicast Routing Information Base (M-RIB)* – sadrži prefikse koji se koriste za *RPF* proveru



ETF, Katedra za računarsku tehniku i informatiku

88

MBGP

BGP Update from Peer
MP_REACH_NLRI: 192.192.2/24
AFI: 1, Sub-AFI: 1 (unicast)
AS_PATH: 300 200
MED:
Next-Hop: 192.168.200.2

Unicast RIB

| Network | Next-Hop | Path |
|-------------------|---------------|-----------|
| *>i160.10.1.0/24 | 192.20.2.2 | i |
| *>i160.10.3.0/24 | 192.20.2.2 | i |
| *>i192.192.2.0/24 | 192.168.200.2 | 300 200 i |

Multicast RIB

| Network | Next-Hop | Path |
|------------------|------------|------|
| *>i160.10.1.0/24 | 192.20.2.2 | i |
| *>i160.10.3.0/24 | 192.20.2.2 | i |

- Skladištenje pristiglih NLRI informacija zavisi od vrednost AFI/SAFI atributa
 - Samo unicast IPV4 ruting tabela



MBGP

BGP Update from Peer
MP_REACH_NLRI: 192.192.2/24
AFI: 1, Sub-AFI: 2 (multicast)
AS_PATH: 300 200
MED:
Next-Hop: 192.168.200.2

Unicast RIB

| Network | Next-Hop | Path |
|------------------|------------|------|
| *>i160.10.1.0/24 | 192.20.2.2 | i |
| *>i160.10.3.0/24 | 192.20.2.2 | i |

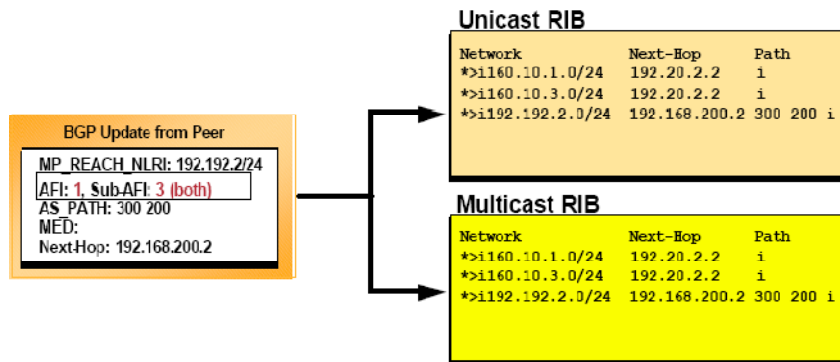
Multicast RIB

| Network | Next-Hop | Path |
|-------------------|---------------|-----------|
| *>i160.10.1.0/24 | 192.20.2.2 | i |
| *>i160.10.3.0/24 | 192.20.2.2 | i |
| *>i192.192.2.0/24 | 192.168.200.2 | 300 200 i |

- Skladištenje pristiglih NLRI informacija zavisi od vrednost AFI/SAFI atributa
 - Samo multikast IPV4 ruting tabela



MBGP



- Skladištenje pristiglih NLRI informacija zavisi od vrednost AFI/SAFI atributa
 - Unikast i multikast IPv4 ruting tabela



Literatura

- Developing IP Multicast Network
Volume 1
Beau Williamson,
Cisco Press 2000

