

Cisco Expo
2008



**Brza konvergencija routing
protokola u mrežama nove
generacije**

**Krešimir Segarić,
CCIE 7691**



**Enable Your Network
Empower Your Business**

20. i 21. ožujka 2008.
Hotel Dubrovnik Palace
Dubrovnik

Welcome to the Human Network.



- 80 djelatnika
- 80M kuna prometa
- Konstantan rast:
- Među prvih pet ponuđača IT rješenja
- 200 partnera i korisnika
- Preko 300 certifikata
- Vendori: Cisco Systems, Microsoft, Nokia, HP, IBM, Arbor, Tandberg, ...



Networking:

- Cisco Gold Partner od 2004.
- 8 CCIE inženjera – najviše u Hrvatskoj
- Presales/postsales tehnička podrška, održavanje opreme, edukacijski i testni (Prometric & VUE) centar

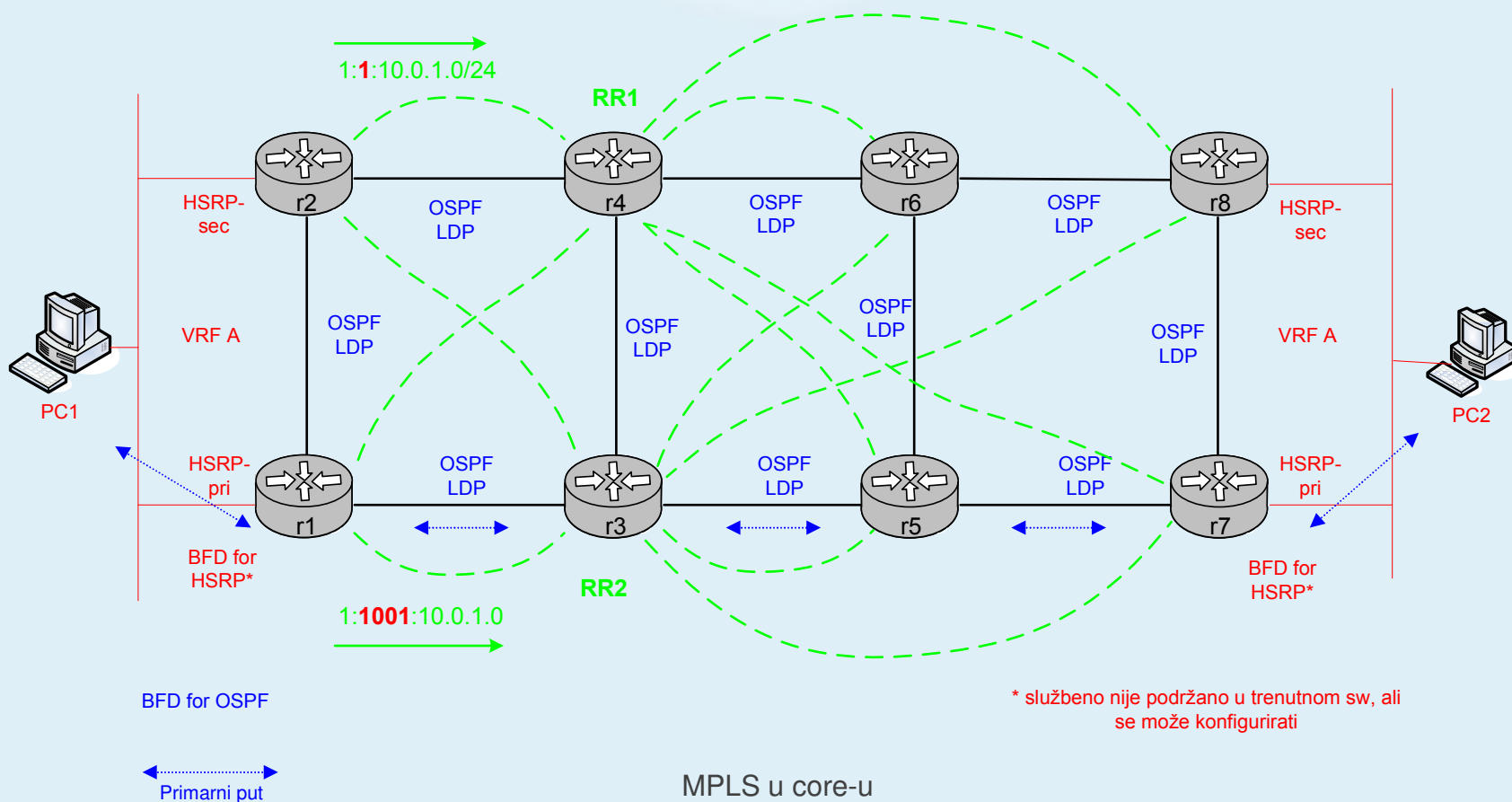
- RECRO-NET: jedini Cisco Learning Partner (CLP) u Hrvatskoj
- instruktori:
 - dugogodišnje predavačko iskustvo (do sada u prethodnim tvrtkama i u RECRO-NET-u školovali više od 2000 polaznika)
 - vrhunski inženjeri : 4 CCIE i 2 CCNP instruktora
- ocjene polaznika: po MTM metrici za instruktore, RECRO-NET se nalazi u prvih 9% CLP-ova s prosječnom ocjenom 4,88 na skali 1-5 (podaci za period 05.03.2007.-05.03.2008.)
- Cisco tečajevi: ICND1/2 (osnove rutiranja/switchinga), BSCI (napredni routing), BCMSN (napredni switching), SNPA (ASA/PIX), QoS, BGP, MPLS, CVOICE, Cisco IP Telephony1/2 , IPv6 (IPVSF / IPVSD),
- Vlastiti tečajevi: IP networking, IP multicast

Brza konvergencija
routing protokola u
mrežama nove
generacije

- Redundancija u trenutim mrežama: često implementirana ali ne i optimirana u smislu brze konvergencije
- Povećnjem mogućnosti mreže (npr. implementacija MPLS VPN-ova) raste i vrijeme konvergencije koje sa defaultnim postavkama korištenih protokola može biti i nekoliko desetaka sekundi
- Optimizacija posebno bitna kod integriranih mreža
- Problemi s konvergencijom javljaju se po ispadu linka ili routera ali jednaki ili još i veći problemi mogu se javiti kod vraćanja linka /routera (npr. LDP prije routinga)
- Ovime problem konvergencije može biti posebno izražen u mrežama s nestabilnim linkovima

- Na case study-u razmotrit će se mogućnosti optimizacije parametara protokola kako bi se dobila konvergencija u vremenu od 0-1,5 sekunde (tipično ispod 0,5 s) na ISR routerima za mrežu s MPLS L3 VPN-ovima
- koriste se 2811 routeri s 12.4(15)T3 IOS-om i 2960 switchevi
- vrijeme konvergencije bilo bi ispod 1 s kad bi ovi ruteri momentalno detektirali pad ethernet sučelja koje se nalazi u VRF-u (delay od 1 s)
- pad ethernet sučelja ili routera u core dijelu mreže detektira se BFD-om i konvergencija je uvijek ispod 1 s
- case study ne koristi MPLS TE tunele (pa time ni npr. FRR)
- uz MPLS TE tunele, izbjegli bi se eventualni privremeni loop-ovi kod vraćanja linkova (uz podešene parametre, loop traje bitno manje od 1 s) ali uz dodatnu kompleksnost (uvođnje RSVP-a i targetiranog LDP protokola)

Slika mreže i korišteni protokoli



Optimizacija brzine konvergencije pojedinih protokola

- Defaultni parametri iz vremena kad se u mrežama nalazio isključivo data promet
- Velik broj timera, potrebno podešavanje praktički svih
- Potreban oprez jer pogrešnim podešavanjem (npr. premale vrijednosti određenih timera) konvergencija se može usporiti
- Održavanje veze između susjeda: hello/dead intervali: default na Ethernet sučelju 10/40 sekundi.
- Minimum za “standardni” način održavanja veze putem OSPF hello paketa je dead time od 1 sekunde, pri čemu se onda specificira broj hello paketa koji se šalje u 1 s
- U slučaju da se želi dobiti manje vrijeme, potrebno koristiti BFD mehanizam – bidirectional forwarding detection

- BFD protokol je općeniti protokol kojeg mogu koristiti razni “klijenti” (npr. OSPF, HSRP)
- Klijent javlja BFD-u sučelje i susjeda kojeg želi “monitorirati”
- BFD ispregovara parametre prozivke sa susjedom i nakon toga šalje pakete susjedu i očekuje odgovor

```
r2(config-subif)# bfd interval ?  
<50-999> Milliseconds
```

```
r2(config-subif)# bfd interval 50 ?  
min_rx Minimum receive interval capability
```

```
r2(config-subif)# bfd interval 50 min_rx ?  
<50-999> Milliseconds
```

```
r2(config-subif)# bfd interval 50 min_rx 50 ?  
multiplier Multiplier value used to compute holddown
```

```
r2(config-subif)# bfd interval 50 min_rx 50 multiplier 3
```

- OSPF koristi BFD za održavanje veza:

```
router ospf 1  
bfd all-interfaces
```

- Kod promjene u mreži, ruteri generiraju novu verziju “iskaznice” (LSA)
- Vrijeme generacije iskaznice moguće podešavati:

```
rr2(config-router)#timers throttle lsa all ?
```

```
<0-600000> Delay to generate first occurrence of LSA in milliseconds
```

```
r2(config-router)#timers throttle lsa all 70 ?
```

```
<1-600000> Minimum delay between originating the same LSA in milliseconds
```

```
r2(config-router)#timers throttle lsa all 70 80 ?
```

```
<1-600000> Maximum delay between originating the same LSA in milliseconds
```

```
r2(config-router)#timers throttle lsa all 70 80 100
```

- Kad se mijenjaju ovi parametri, potrebno je konfigurirati i *timers lsa arrival*, na vrijednost manju od vrijednosti “Minimum delay between originating the same LSA in milliseconds” iz ove komande
- U suprotnom, događaju se retransmisije definirane interface parametrom *ip ospf retransmit-interval (default 1, konfigurira se u sekundama)*
- Default za “Minimum delay..” je 5s a za *timers lsa arrival* 1s.

- Prosljeđivanje iskaznice: definirano ospf parametrom *timers pacing flood* .
- Default 33 ms, može se smanjiti ukoliko susjed može dovoljno brzo primati iskaznice

- SPF rekalkulacija

```
r4(config-router)#timers throttle spf ?
```

```
<1-600000> Delay between receiving a change to SPF calculation in  
milliseconds
```

```
r4(config-router)#timers throttle spf 100 ?
```

```
<1-600000> Delay between first and second SPF calculation in milliseconds
```

```
r4(config-router)#timers throttle spf 100 100 ?
```

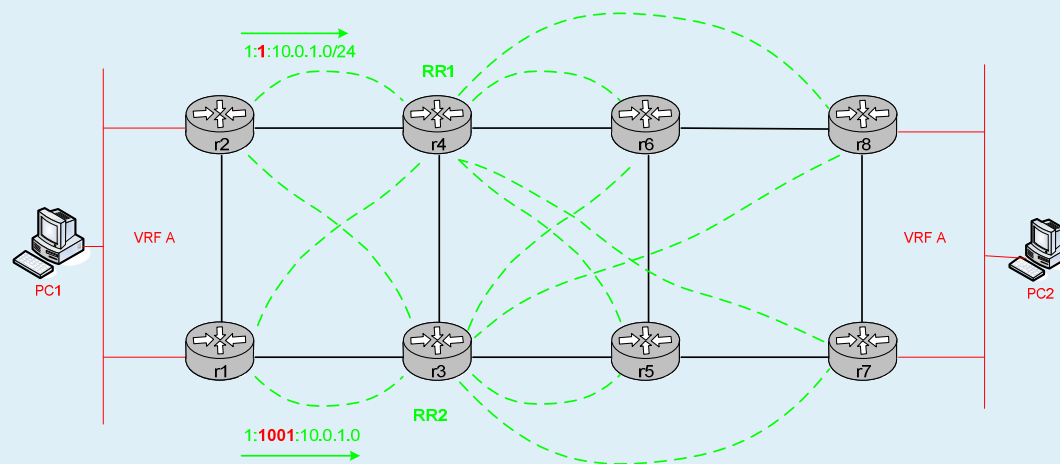
```
<1-600000> Maximum wait time in milliseconds for SPF calculations
```

```
r4(config-router)#timers throttle spf 100 100 250
```

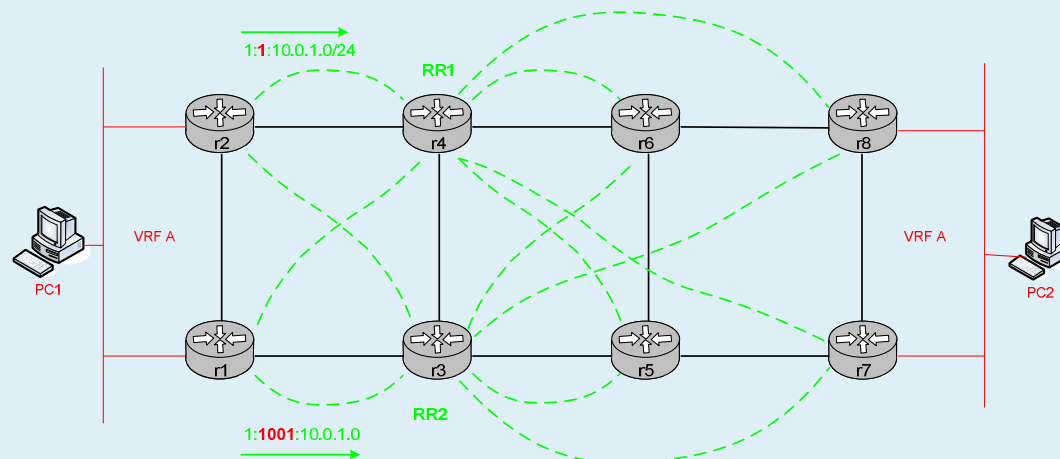
- Defaultne vrijednosti SPF rekalkulacije: 5s /10s /10s

- Kod podešavanja timera treba biti konzistentan i u čitavoj mreži namjestiti iste timere
- Npr, ukoliko bi na jednom ruteru bio namješten spf delay na (bitno) veću vrijednost od one koja je na drugim ruterima, po promjeni topologije njegova routing tablica se duže ne mijenja i mogući su privremeni loop-ovi

- MPBGP nužan za slanje VPNv4 informacije
- zbog skalabilnosti, praktički uvijek se koristi design sa RR
- Potrebni redundanti RR
- Kako bi se izbjegla odluka o best-pathu na RR, koriste se različiti rd-ovi na “primarnom” i “sekundrnom” ruteru uz isti import/export RT
- PE ruteri dobijaju 4 puta do VPN mreže
- BGP bestpth algoritam odabire put koji ima najkraću IGP metriku do next-hopa ako se ništa dodatno ne konfigurira (npr, uz istu cijenu svih linkova, r7 do PC-1 ide preko r5,r3,r1)



- BGP next-hop tracking feature, implementiran u novijim IOS-ima, omogućuje brzu promjenu patha ukoliko next-hop postane nedostupan ili se promijeni metrika. Budući da svi susjedi imaju u MPBGP-u rutu i preko “sekundarnog” rutera, kod ispada primarnog rutra, alternativna ruta se “vrlo brzo” instalira u vrf routing tablicu
- Brzina reakcije definirana komandom “bgp nexthop trigger delay”
- Alternativa ovome rješenju je rušenje BGP susjeda na RR (“neighbor x.x.x. fall-over route-map xxx”) ukoliko se izgubi ruta na njega u routing tablici. Tad RR obavještava ostale susjede da povlači rute koje je prije taj peer oglasio. Budući da svi susjedi imaju u MPBGP-u rutu i preko “sekundarnog” rutera, ruta se instalira momentalno u vrf routing tablicu

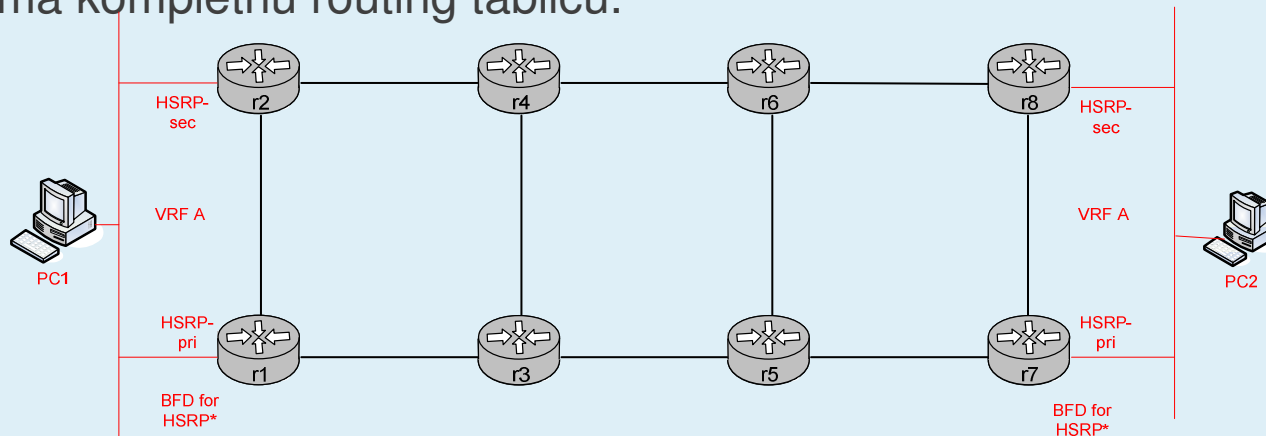


- “Standardni problem” kod MPLS VPN-ova je problem vraćanja linka preko kojeg se ide do BGP next-hopa
- Ukoliko se routing protokol digne prije LDP-a, MPLS paketi koji idu preko linka gube LSP labelu i imaju samo VPNv4 labelu – gubitak VPNv4 paketa
- Zbog brže konvergencije i smanjenja kompleksnosti OSPF topologije, moguća je primjena komande “ip ospf network point-to-multipoint” na ethernet sučelju, čime po dizanju linka OSPF praktički momentalno izmijeni iskaznice sa susjedom. Uz podešene OSPF parametre za brzu konvergenciju, OSPF je tipično 5-10 sekundi “brži” od LDP-a
- U novijim IOS-ima, elegantan način rješavanja ovog problema je:

```
router ospf 1
mpls ldp sync
```

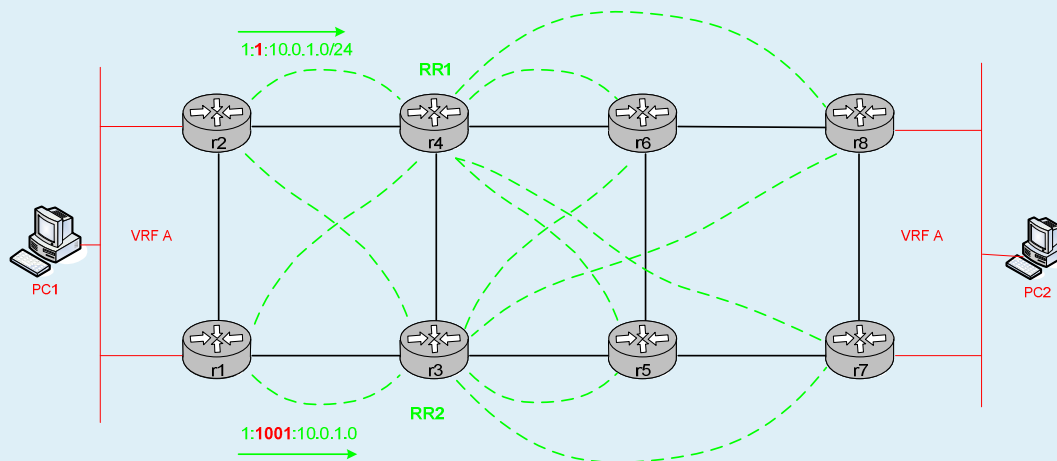
- Ukoliko je LDP transportna adresa nedostupna, OSPF uspostavlja vezu ali sučelje ima max metriku (65535). Inače, OSPF ne uspostavlja vezu dok se LDP veza ne uspostavi

- HSRP dopušta konvergenciju u vremenu od oko 600 ms na “standardan” način (zbog niskog prioriteta HSRP procesa , konfiguracija dead timera od npr 300 ms nije moguća jer npr. “write mem” uzrokuje “prijelaz” adrese na susjeda)
- s BFD-om, moguće ostvariti konvergenciju od 150 ms
- BFD nije podržan za HSRP u VRF-u, ali se (uz podešavanja) može konfigurirati
- Potrebno podestiti i timere da HSRP ne preuzme adresu čim se sučelje digne odnosno nakon router reloada. U suprotnom, paketi idu prema ruteru koji nema kompletnu routing tablicu.

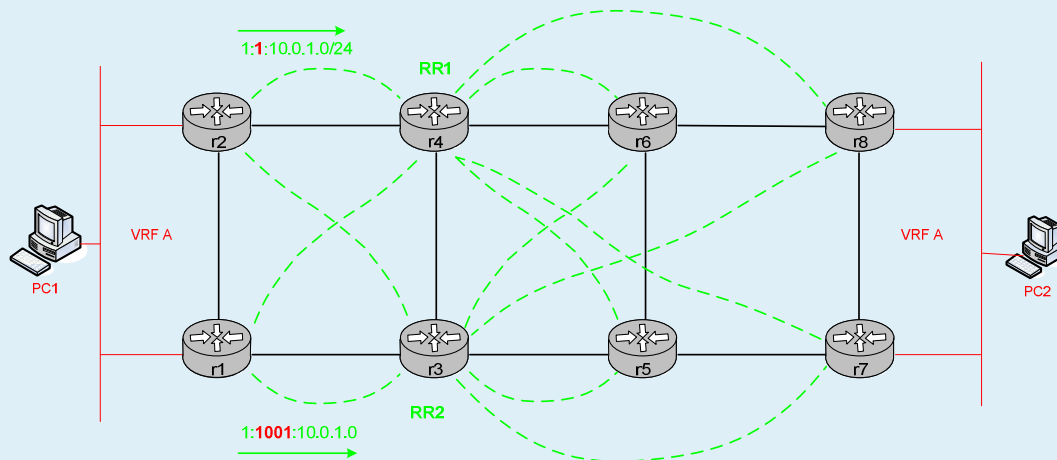


* službeno nije podržano u trenutnom sw, ali se može konfigurirati

- Na ISR routerima 2811 ostaje problem ispada interface-a u VRF-u
- Npr. ispad interfacea na R1 vidi se tek 1 sec nakon što interface padne (mjereno na način da se uspostavi BFD veza r1-r2, r1 i r2 se sinhroniziraju NTP-om, BFD "okrene" HSRP na r2 cca 850 ms prije nego r1 prijavi da je line protocol down. BFD "holdtime" je bio 3*50 ms)
- Problem: r1 oglašava i dalje da je VRF mreža dostupna pa dolazni paketi na tu mrežu idu preko njega (odlazni paketi nisu problem jer idu preko r2)



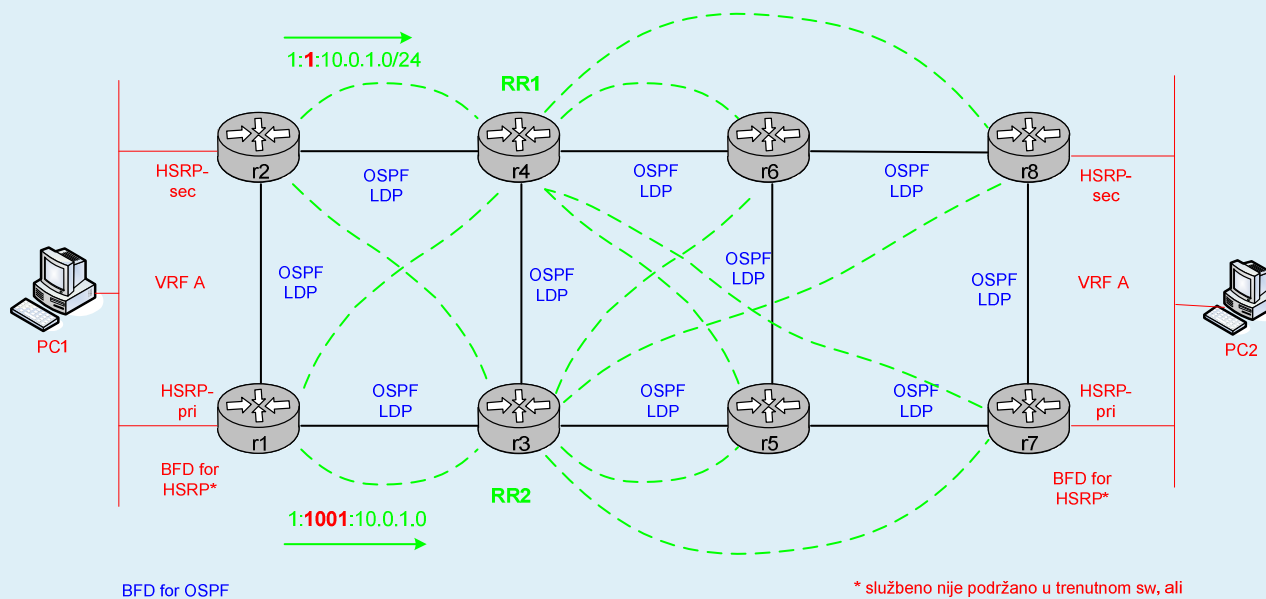
- Nakon što se promijeni ruta, potrebno je promijeniti i CEF informaciju
- Implementacija CEF-a na ISR-u različita od implementacije CEF-a na npr 7600 i IOS-om 12.2(33)SRC
- Implementacija na ISR-u ovisi o CEF background scanneru, ne na SRB / SRC na 7600
- Za rekurzivne VPNv4 rute CEF se na ISR-u update-ira u background procesu ako se mijenja metrika do next-hopa
- Može dovesti do dodatnog delaya (tipično ispod 1s)



Mjerenje: konstantan ping PC1-PC2, paketi svakih 100 ms

Element mreže	Ispad (s)	Vraćanje (s)
r1 - ruter	0,3	0,3
r1 – r3 link	0,4	0 - 0,1
r1 – link prema VRF-u	1,1*	0 - 0,1
r3 - router (RR)	0,3	0 - 0,1
r3 – r5 link	0,3	0 - 0,1
r5 – router	0,3	0 - 0,1

* problem detekcije ispada linka na r1



- Podešavanjem paramtera protokola na Cisco IOS-u na ISR routerima može se postići brza konvergencija u MPLS/VPN okruženju
- Potrebno podesiti praktički sve protokole
- Potrebno dobro poznavanje svih protokola i međusobne ovisnosti
- Case study baziran na ISR ruterima - u “realnim” mrežama, P i PE funkcionalnost trebaju obavljati hardware bazirani ruteri s razdvojenim data i kontrolnim ravninama

- RECRO-NET za svojom tehničkom ekspertizom (8 CCIE inženjera) daje podršku u implementaciji / održavanju / troubleshootingu svih kompleksnih mreža te pruža vrhunsku obuku mrežnih inženjera u našem edukacijskom centru