Projektovanje visokopozdanih sistema i mreža

Sadržaj

- Osnovne definicije
- Pouzdanost pojedinačnih komponenti
- Eliminisanje jedinstvenih tačaka otkaza
- First Hop Redundancy protokoli
- Pozdanost složenih sistema

- Literatura:
  - [1] Israel Koren, C. Mani Krishna, "Fault-Tolerant Systems", Elsevier, 2007, ISBN: 978-0-12-088568-8.

Definicije:

- defekt, greška
- pouzdanost
- dostupnost
- MTTF

OSNOVNE DEFINICIJE

Značaj

- U današnje vreme dostupnost sistema je imperativ za poslovanje bilo koje ozbiljne institucije.
- Sve komponente se po pravilu kvare. To je neminovno. Ne postavlja se pitanje da li će se nešto pokvatiri, već KADA će postati neispravno.
- Ključna pitanje na koje odgovor daje naučna oblast koja se bavi sistemima visoke pouzdanosti je:
  - KAKO OD NEPOUZDANIH KOMPONENTI NAPRAVITI POUZDAN SISTEM?
- Na primer, Google ima dostupnost 100%, što ne znači da im se računari ne kvare i h.diskovi ne gube sadržaj. Amazon AWS EC2 garantuje korisnicima minimalnu dostupnost od 99.5%, itd.

Definicije – **defekt, greška, otkaz**

- Defekt (en. defect)
  - predstavlja uzrok kvara
- Greška (en. error)
  - predstavlja pogrešan rezultat
- Kvar/otkaz (en. fault)
  - predstavlja stanje sistema u kom sistem može dati pogrešan izlaz ili prestati da funkcioniše
- Svaki otkaz ne rezultuje nužno greškom za svaki od ulaza.
  - Primer, sabirač sa jednim od izlaznih bitova konstantno na log. 1...
- Otkazi se mogu klasifikovati
  - po vremenu trajanja
    - trajni (en. permanent), privremeni (en. transient), povremeni (en. intermitted)
  - po efektima koje ima
    - otkaz celog sitema (praktično, sisten ne radi)
    - povremene greške u radu

Definicije – **pouzdanost**

- Pouzdanost (en. reliability)
  - je verovatnoća u funkciji vremena, u oznaci *R(t),* koja ima vrednost u trenutku *t* jedanaku verovatnoći da je sistem bez prekida bio ispravan od trenutka t_0 do trenutka *t*.
- Ova mera je pogodna za opis sistema kod kojih i trenutni prekid u radu sistema može imati velike posledice.
- Blisko povezane sa ovom merom su i
  - *srednje vreme do otkaza (Mean Time to Failure)* – MTTF
    - prosečno vreme rada sistema do pojave otkaza
  - *srednje vreme između otkaza (Mean Time Between Failures)* – MTBF
    - *prosečno vreme između dva otkaza – uključuje i vreme potrebno za oporavak (popravku) sistema. Mean Time to Repair* – MTTR

$$MTBF = MTTF + MTTR$$

Definicije – **dostupnost**

- Dostupnost (en. Availability)
  - je funkcija vremena, u oznaci *A(t)*, čija je vrednost jednaka prosečnom vremenu za koje je sistem bio ispravan u vremenskom intervalu [0,*t*].
- Ova mera je pogodna za sisteme gde kontinuirani rad nije teliko važan, kao ukupno vreme za koje je sistem bio dostupan.

- Primeri:
  - Web prezentacija može biti nepouzdana, ali je važno da je visoko dostupna (highly available), zato što su interakcije korisnika kratke.
  - Sa druge strane, sistem za rezervaciju avionskih karata treba da bude visokopouzdan, zato što proces rezervacije traje duže vreme, pa bi bilo kakva greška u toku tog procesa značila gubitak važnih informacija.

Definicije – **dostupnost**

- Dugoročna dostupnost (en. Long-Term Availability)
  - u *A* je vrednost:

$$A = \lim_{t \to \infty} A(t)$$

- A je jednako verovatnoći da je sistem ispravan u bilo kom vremenskom trenutku.
- Dugoročnu dostupnost ima smisla određivati samo kod sistema koji se mogu oporaviti od otkaza.
- Dugoročna dostupnost se može izračunati i kao:

$$A = \frac{MTTF}{MTBF} = \frac{MTTF}{MTTF + MTTR}$$

- Trenutna dostupnost (point availability), Ap(t), je verovatnoća da je sistem dostupan u konkretnom trenutku t.

Definicije – **dostupnost**

- Moguće je imati sistem koji ima malu pouzdanost (koji je nepouzdan), a koji ima visoku dostupnost.
  - Na primer, razmotrimo sistem koji otkaže jednom u sat vremena, ali samo na 1 sekundu.
  - Ovakav sistem ima MTBF = 1h, i kao posledicu veoma malu pouzdanost.
  - S druge strane, dostupnost je velika: A = 3599/3600=0.99972

- Sve ove definicije, naravno, zahtevaju definisanje značenja otkaza sistema i šta se pod tim podrazumeva.
  - Na primer, sijalica ili radi ili ne, provodnik ili provodi ili ne. Ali, procesor sa par hiljada miliona gejtova, kom je jedan od gejtova neispravan i uvek daje logičku 1 bez obzira na ulaze, može da da grešku jednom u, recimo, 25.000 sati korišćenja.
- Primer FDIV greške kod Intel Pentium I procesora:
  - link

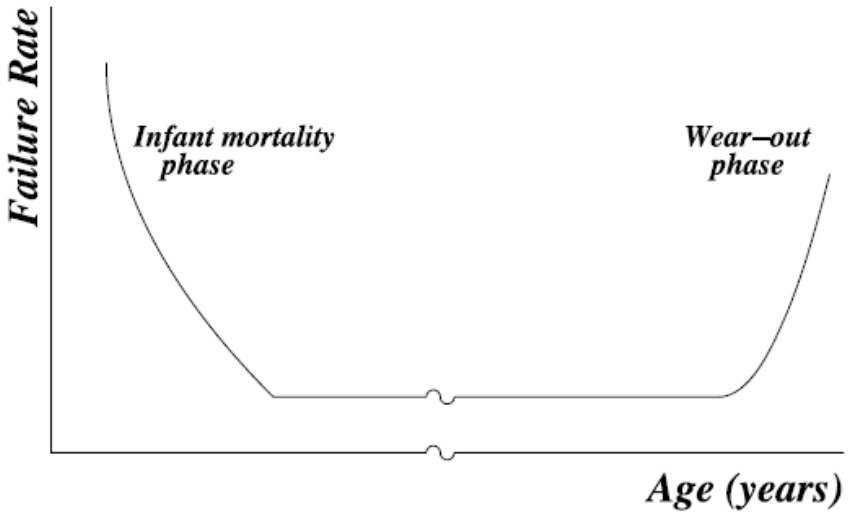
Mere dostupnosti i tipovi sistema

Naziv	Nedostupno (min. po godini)	Dostupnost	Klasa
Unmanaged	52.560 (36 dana)	90%	1
Managed	5.256 (3 dana)	99%	2
Well-managed	526 (9 sati)	99.9%	3
Fault-tolerant	53	99.99%	4
High-availability	5	99.999%	5
Very high-availability	0.5	99.9999%	6
Utra high-availability	0.05 (3 sekunde)	99.99999%	7

POUZDANOST POJEDINAČNIH KOMPONENTI

Definicije – **stopa otkaza**

- Stopa otkaza komponenti je najznačajniji parametar u analizi otkaza pojedinačnih komponenti.
- Predstavlja očekivani broj otkaza pojedinačnih komponenti u jedinici vremena za seriju proizvedenih komponenti.
- Zavisi od starosti komponente, fizičkih karakteristika, temperature u kojoj komponenta radi i dr.



Izvođenje MTTF

- Pokazaćemo kako se pouzdanost i MTTF mogu izvesti iz date stope otkaza.
- Neka je
  1. komponenta ispravna u trenutku t=0 i ispravna do trenutka otkaza,
  2. T vreme do otkaza
  3. otkaz trajan
- Označimo sa $f(t)$ funkciju gustine verovatnoće i sa $F(t)$ kumulativnu distribuciju.
- Veza između ovih funkcija je

$$f(t) = \frac{\mathrm{d}F(t)}{\mathrm{d}t}, \qquad F(t) = \int_0^t f(\tau)\,\mathrm{d}\tau$$

Gustina verovatnoće

- U teoriji verovatnoće, **funkcija gustine verovatnoće** (engl. *probability density function - PDF*) je funkcija čija se vrednost u datom uzorku (ili tački) može protumačiti kao *relativna verovatnoća* da će vrednost slučajne promenljive biti jednaka tom uzorku.
- Drugim rečima, dok je *apsolutna verovatnoća* da kontinuirana slučajna promenljiva poprimi bilo koju određenu vrednost jednaka 0 (pošto postoji neograničen skup mogućih vrednosti), vrednost funkcije dva različita uzorka mogu se koristiti za izvođenje zaključka.
- Primer: Pretpostavimo da data vrsta bakterija obično živi 4 do 6 sati. Kolika je verovatnoća da bakterija živi *tačno* 5 sati? Odgovor je 0%. Mnogo bakterija živi *oko* 5 sati, ali nema šanse da bilo koja bakterija živi *tačno* u 5.0000000000 ... sati.
- Umesto toga može se postaviti pitanje: Kolika je verovatnoća da bakterija umre između 5 sati i 5,01 sata?

Izvođenje MTTF

- $f(t)$
  - verovatnoća otkaza u datom trenutku, ili, preciznije, verovatnoća otkaza u vremenskom intervalu Δt je $\Delta t \cdot f(t)$
- Za $f(t)$ važi

$$f(t) \geqslant 0 \quad \text{za} \quad t \geqslant 0 \qquad \text{i} \qquad \int_0^\infty f(t)\,\mathrm{d}t = 1$$

- $F(t)$
  - verovatnoća da će sistem otkazati pre, ili u trenutku t. Tj. $F(t) = Prob\{T \leq t\}$
- Poudanost je verovatnoća da će sistem biti ispravan do, ili u trenutku t, pa je:

$$R(t) = \mathrm{Prob}\{T > t\} = 1 - F(t)$$

Izvođenje MTTF

- Uslovna verovatnoća otkaza $\lambda(t)$ je verovatnoća da će komponenta otkazati u narednom periodu Δt, ako je bila ispravna u trenutku t:

$$\lambda(t) = \frac{f(t)}{1 - F(t)}$$

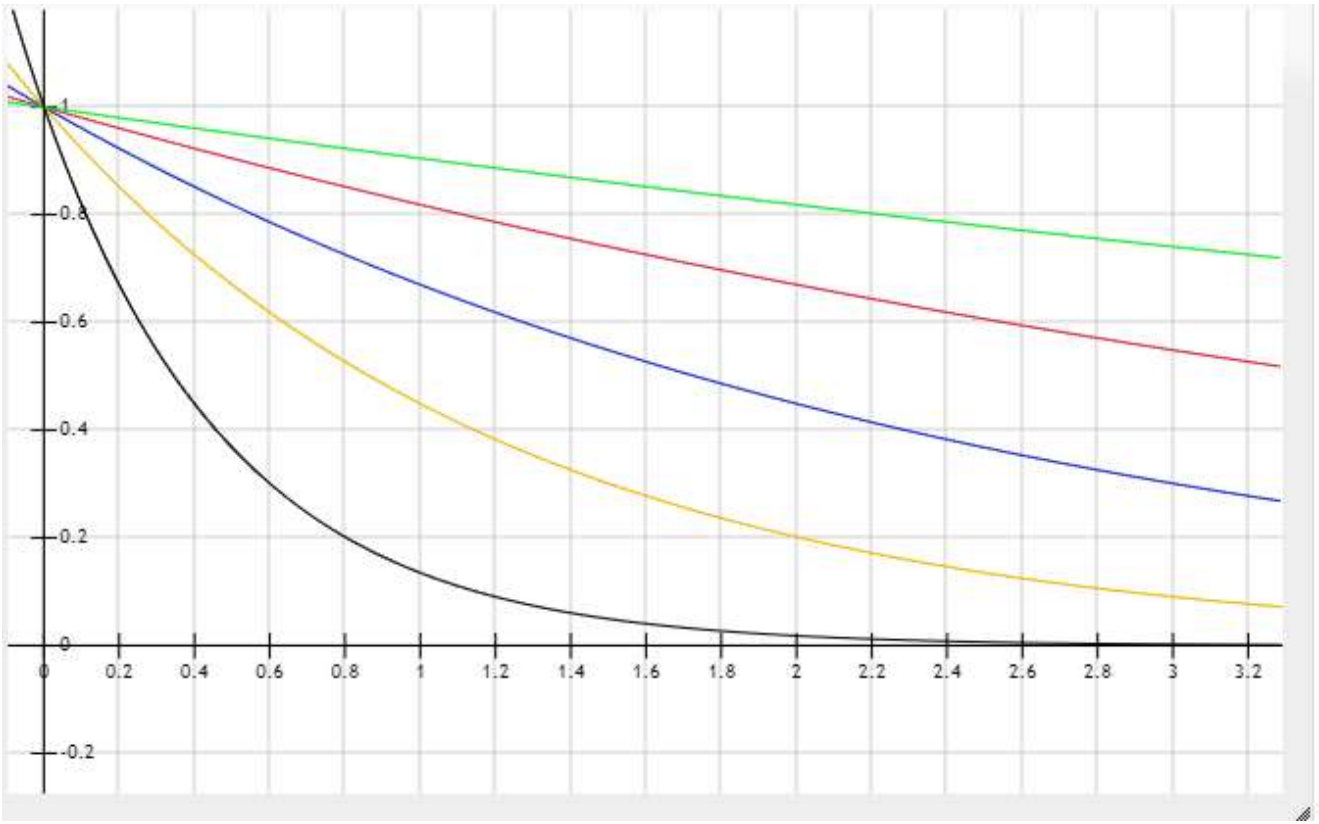
- Kako je $\frac{\mathrm{d}R(t)}{\mathrm{d}t} = -f(t)$, sledi:

$$\lambda(t) = -\frac{1}{R(t)} \frac{\mathrm{d}R(t)}{\mathrm{d}t}$$

- Ukoliko uslovna verovatnoća otkaza ne zavisi od starosti komponente, $\lambda(t) = \lambda$
- Rešenje dif. jednačine $\frac{\mathrm{d}R(t)}{\mathrm{d}t} = -\lambda R(t)$ je $R(t) = \mathrm{e}^{-\lambda t}$

Pouzdanost

$$R(t) = \mathrm{e}^{-\lambda t}$$



Izvođenje MTTF

- Srednje vreme do otkaza MTTF za nepopravljive sisteme je jednako očekivanom vremenu života:

$$MTTF = \int_0^\infty R(t)dt$$

- Za konstantnu stopu otkaza

$$MTTF = \int_0^\infty e^{-\lambda t}dt = \frac{1}{\lambda}$$

ELIMINISANJE SPOF

Jedinstvene tačke otkaza

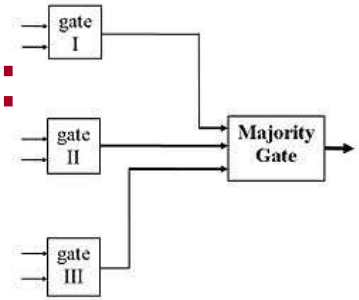
- Jedinstvana tačka otkaza je deo složenog sistema čiji otkaz dovodi do otkaza celokupnog sistema (eng. Single Point of Failure - SPoF).
- Sistem se može učiniti robusnim i pouzdanim eliminisanjem SPoF.
- Sistemi od kojih zavise životi ljudi se obavezno projektuju kao kompletno redundantni sistemi sa dodatnim sistemima za praćenje otkaza pojedinih komponenti (npr. kod aviona).
- Praćenje otkaza je važno da bi se moglo reagovati u slučaju otkaza jedne komponente, ukoliko sistem radi pomoću rezevne komponente (eng. *spare*).

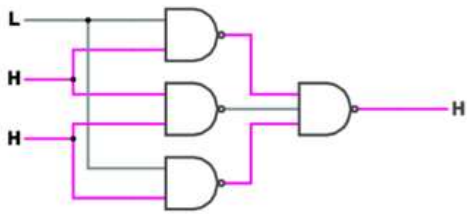
Redundansa

- Redundantni sistemi:
  - Hardverska redundansa
  - Informacina redundansa
  - Sofverska redundansa
  - Redundansa mreža

N-modularna redundansa

- N-modularna redundansa je hardverska tehnika dodavanja (N) komponenti gde svaka komponenta može preuzeti osnovnu funkciju u slučaju otkaza.
- Najpoznatiji sistemi su:
  - Dual-modularna redundansa
    - Pogodna je za multipliciranje celih sistema (npr. računara) tako da ako jedan otkaže, drugi može preuzeti rad. Neophodno je da postoji način za utvrđivanje neispravnosti.
  - Triple-modularna redundansa
    - Pogodna je za hardverske komponente na nivou logičkih komponenata, jer može lako obezbediti „samooporavak" većinskim „glasanjem".
- TMR sistem sa voter-om:





TMR bez SPoF i dinamička redundansa

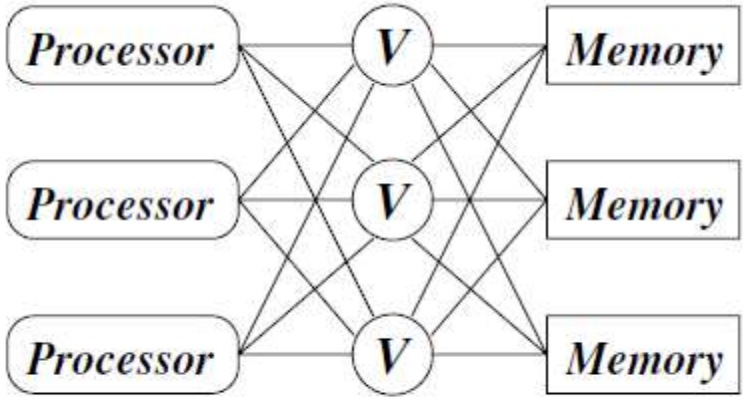


FIGURE 2.9 Triplicated voters in a processor/memory TMR.

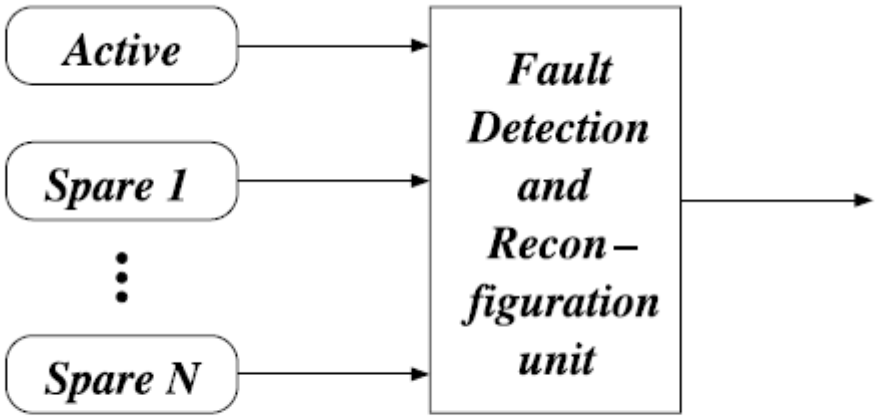
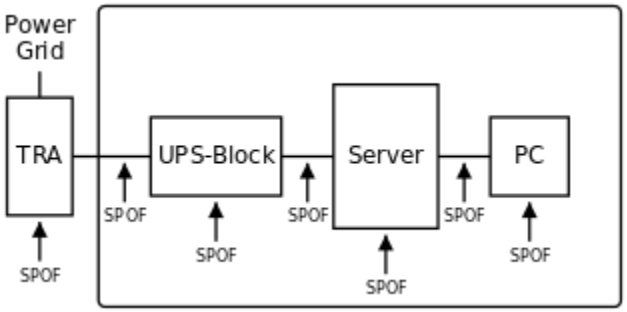
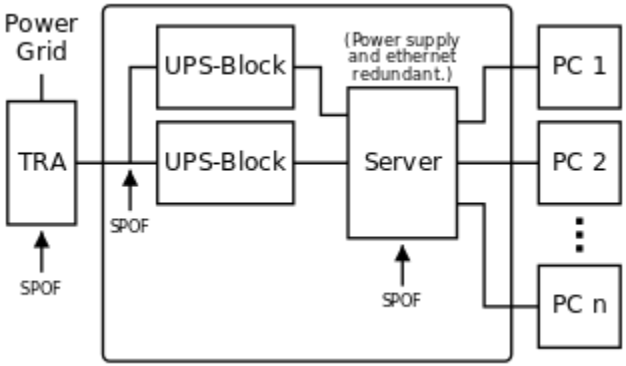


FIGURE 2.10 Dynamic redundancy.

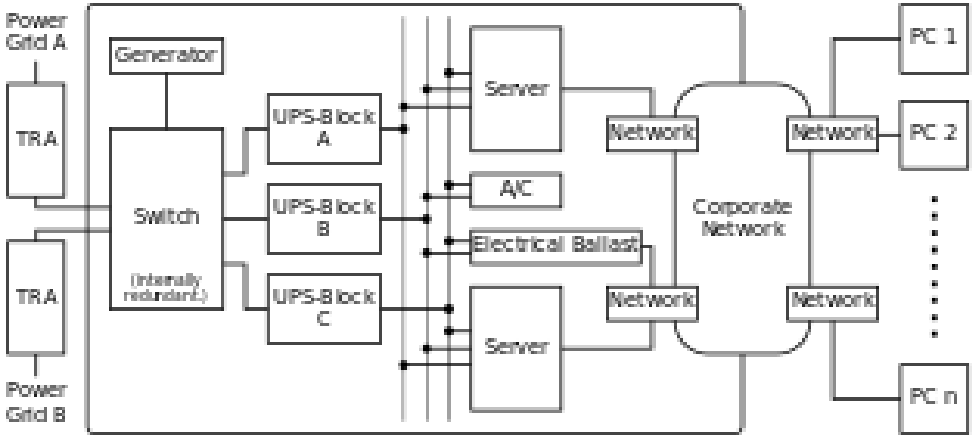
Primer redundantnog sistema



- sistem bez redundanse



- eliminisanje nekih SPoF (dual-napajanje)

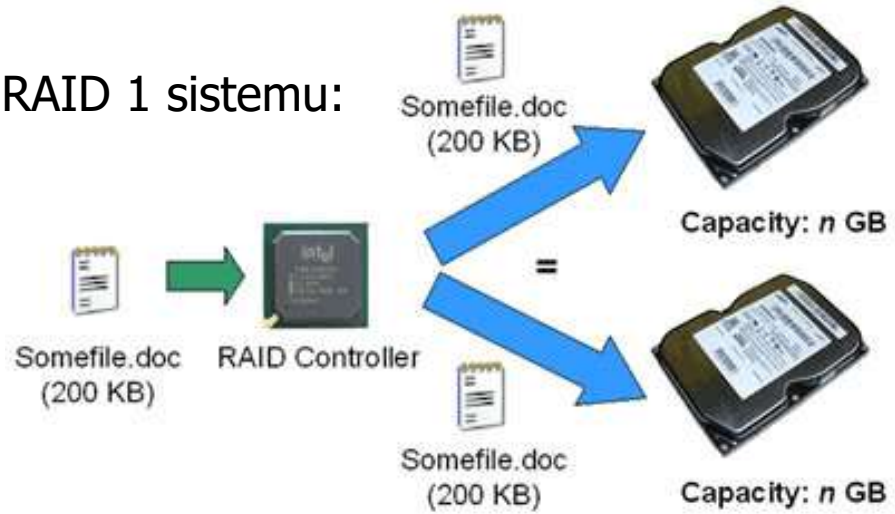


- eliminisanje svih SPoF

Informaciona redundansa

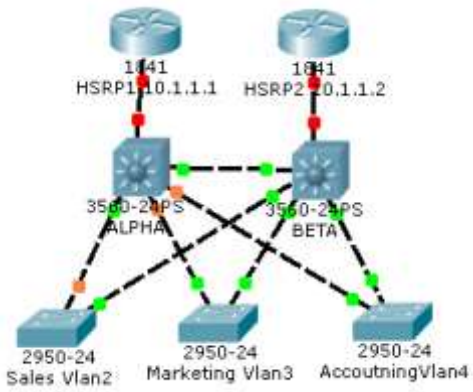
- Specijalni kodovi na osnovu koji se može rekonstruisati pogrešan/pogrešni bitovi
  - Parity check
  - CRC
  - Itd.
- Disk sistemi:
  - RAID 1,2,3,4 i 5 (literatura, str. 79 do 84)

Diskovi u RAID 1 sistemu:

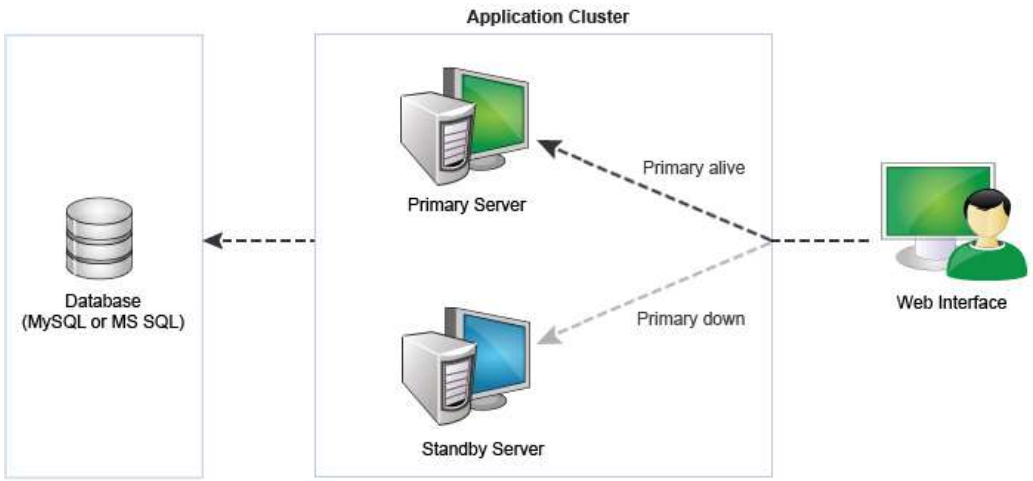


Redundantne mreže

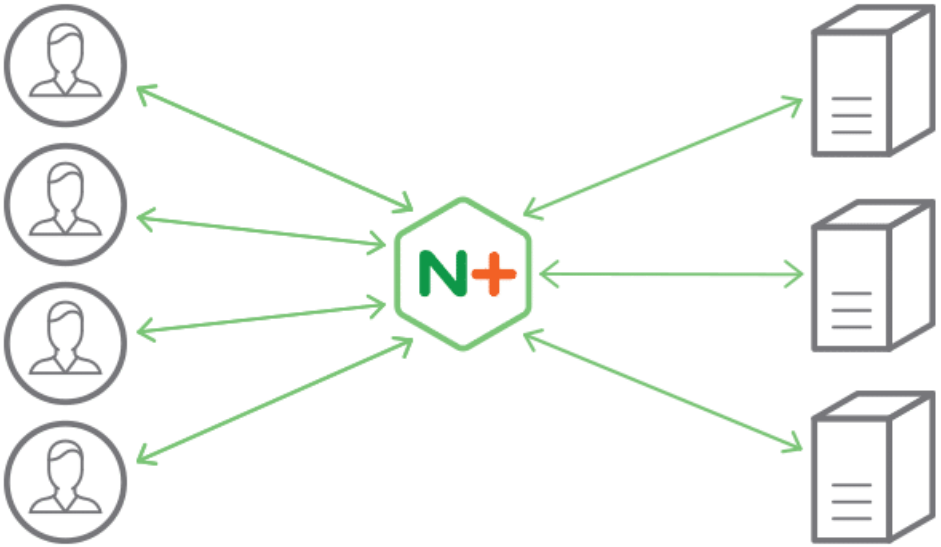
- Redundansa na L1
  - Fizički postavljeni redundantni linkovi i komponente
- Redundansa na L2
  - Omogućena STP protokolom, uz vreme rekonfiguracije koje zavisi od varijante STP-a: PVST+ varijante 50s (zavisno od dijametra i podešavanja), RSTP par 100ms, ako su linkovi p2p
- Redundansa na L3
  - Omogućena ruting protokolom i zavisi od brzinekonvergencije protokola (holddown tajmeri i sl., RIP – 180s)
- Redundansa L4-L7
  - Load balanseri



Load balanseri



- Primeri (link)
  - nginx (L7 load balancer)
  - HA Proxy
  - Neutrino (L4 loadbalancer)
  - …

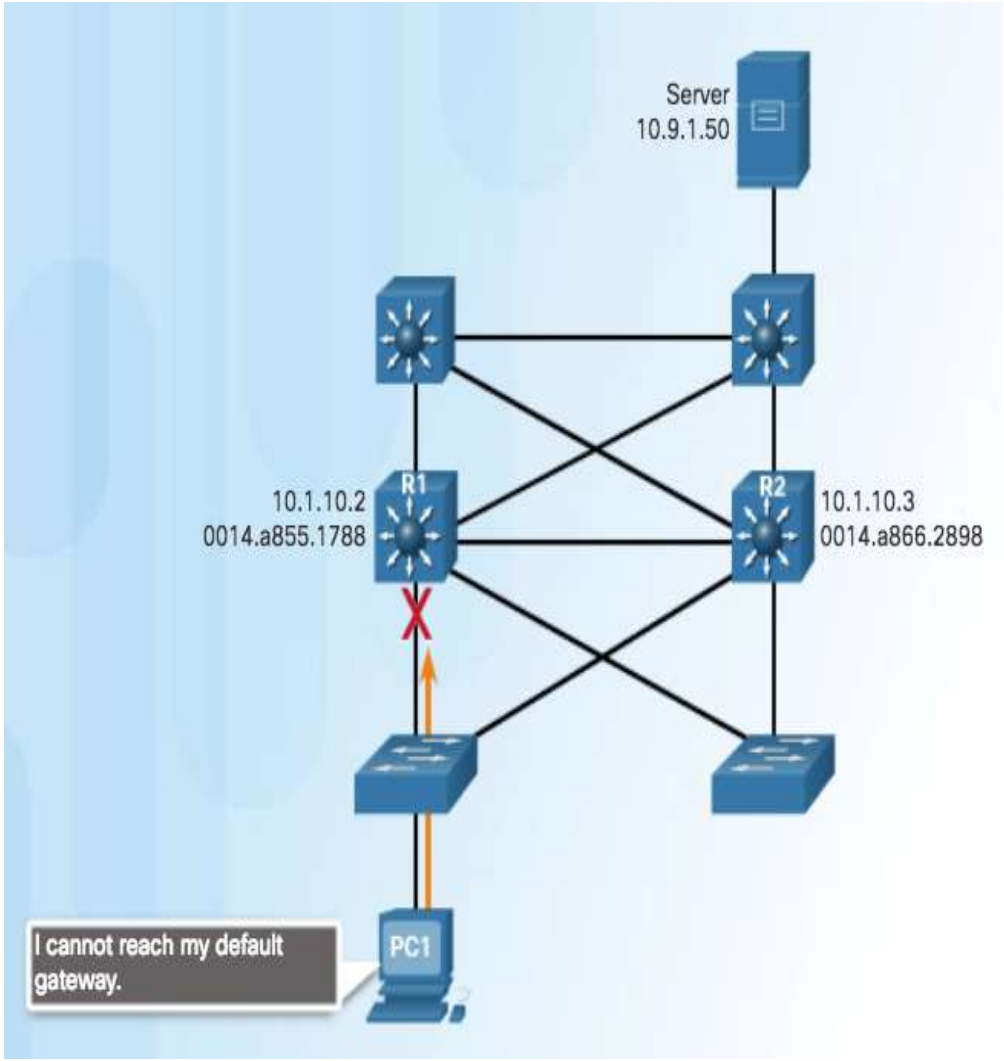


FIRST-HOP REDUNDANCY

Ograničenja gateway-a

Ukoliko na istom VLAN-u imamo dva ili više rutera, neophodan je mehanizam za preusmeravanje saobraćaja u slučaju otkaza

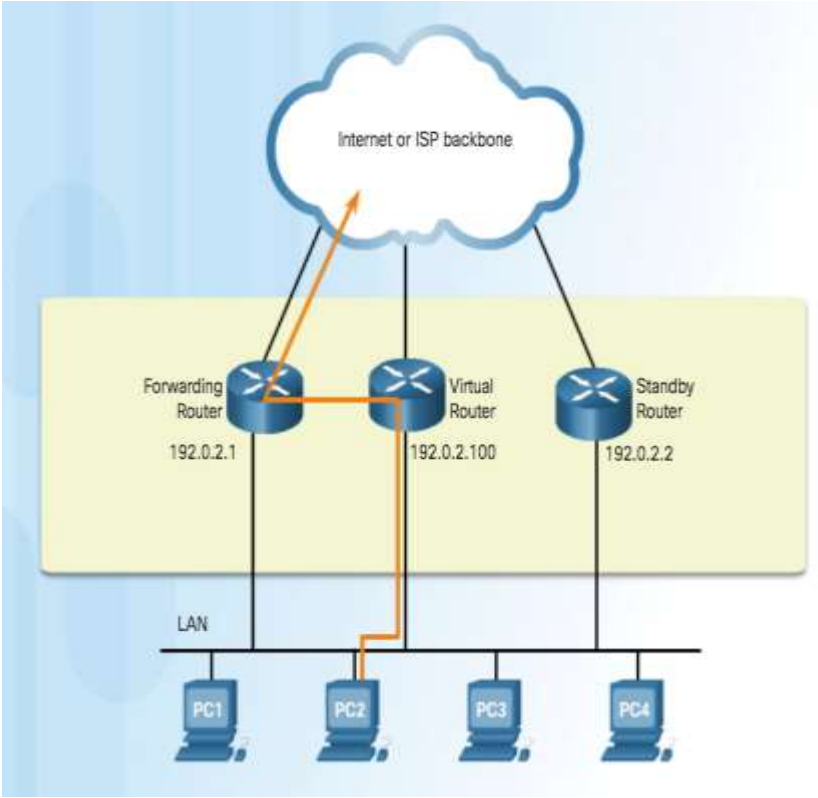
- Na slici su prikazani L3 svičevi koji imaju ulogu gateway-a.
- Na IP mreži svaki host (PC, telefon, i sl.) ima podešavanje za samo jedan gateway-
- Jednostavno, na OS-u ne postoji mogućnost dodavanja redundantnog gateway-a, čak i da fizički postoji na mreži.
- Na slici, R1 je podešen kao gateway za PC1 i rutira pakete koji dolaze od njega. U slučaju otkaza R2 bi mogao da preuzme ulogu R1 jer postoji put, a i po IP adresama su na istom VLAN-u



Router Redundancy

Za prevenciju SPoF može se podesiti „virtuelni ruter“.

- Računari imaju iluziju da je na mreži dostupan samo jedan ruter.
- Dva ili više rutera mogu imati iste IP i MAC adrese, tako da deluju kao jedan (virtuelni) ruter.
- Na celom segmentu hostovi imaju adresu virtuelnog rutera podešenu kao gateway.
- ARP rezolucija vraća MAC adresu virtuelnog routera.
- Fizički uređaj koji prosleđuje pakete na spoljašnjim mrežama (internetu) je transparentan za korisnike.



„Redundancy“ protokoli pružaju mehanizam pomoću koga se određuje koji će ruter imati ulogu „forwarding“ rutera.

Mogućnost da mreža samostalno odredi na osnovu otkaza koji će uređaj raditi kao default gateway se naziva **first-hop redundancy**.

„Failover“ koraci

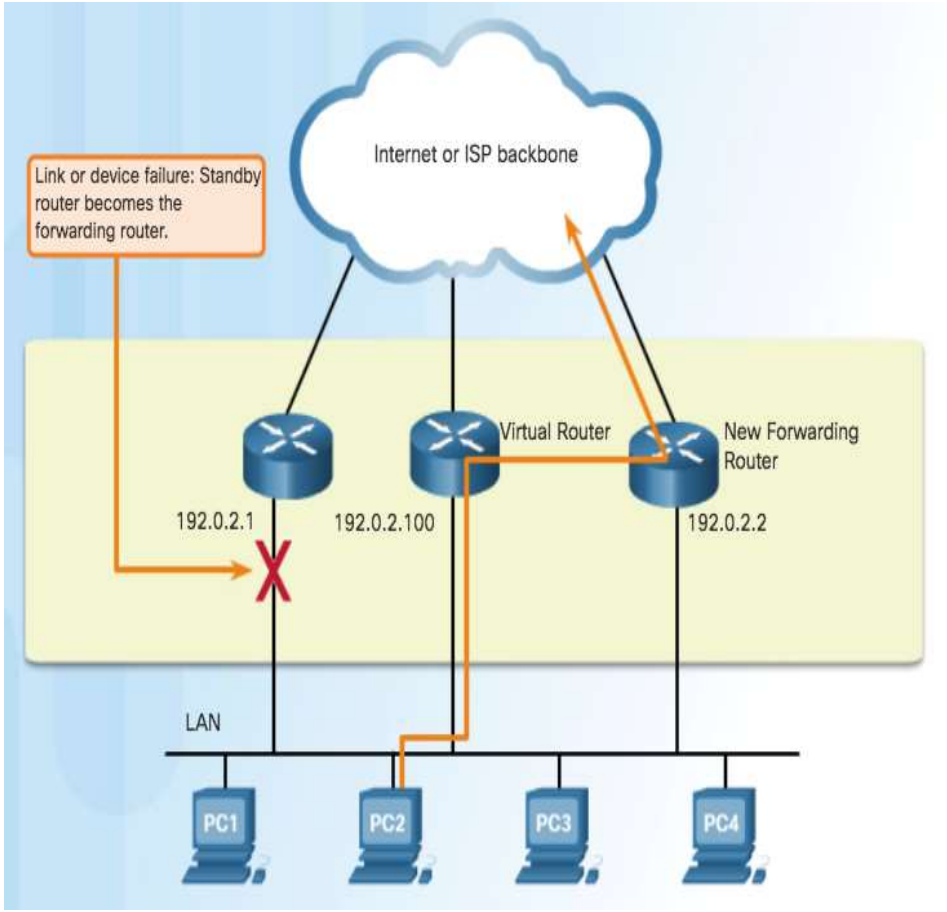
Kada aktivni ruter otkaže, redundancy protokol predaje standby ruteru aktivnu ulogu (active role).

Koraci su sledeći:

1. Standby ruter prestaje da dobija hello poruke od forwarding rutera usled njegovog otkaza.

2. Standby ruter preuzima ulogu forwarding rutera.

3. S obzirom da novi forwarding ruter preuzima i IPv4 i MAC adrese virtuelnog rutera, hostovi ne primećuju prekid u radu mreže.



First Hop Redundancy protokoli

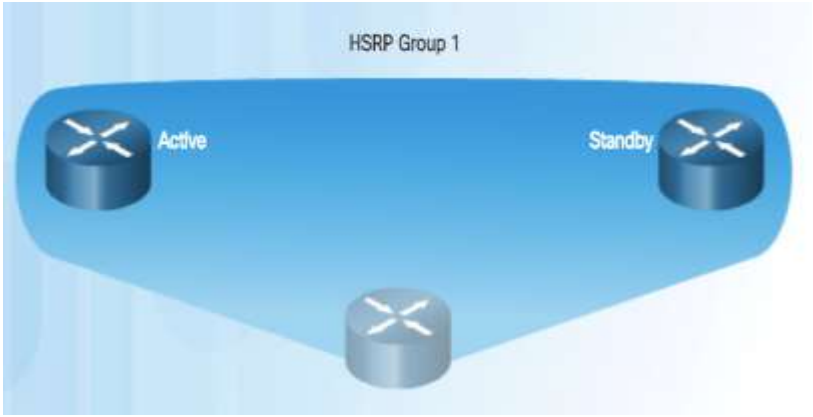
Hot Standby Router Protocol (HSRP) - Cisco-proprietary FHRP dizajniran da omogući transparentni failover gejtveja.

Active device je uređaj preko koga se rutiraju paketi.

Standby device je uređaj koji preuzima ulogu aktivnog u slučaju njegovog otkaza.

Funkcija HSRP protokola je da prati rad HSRP grupe uređaja i brzo preda ulogu forvardovanja paketa u slučaju otkaza.

HSRP for IPv6 - Cisco-proprietary FHRP sa istom funkcionalnošću, ali za IPv6 mreže.



- HSRP defines a group of routers - one active and one standby.
- Virtual IP and MAC addresses are shared between the two routers.
- To verify HSRP state, use the **`show standby`** command.
- HSRP is Cisco proprietary.
- VRRP is a standard protocol.

First Hop Redundancy protokoli (nast.)

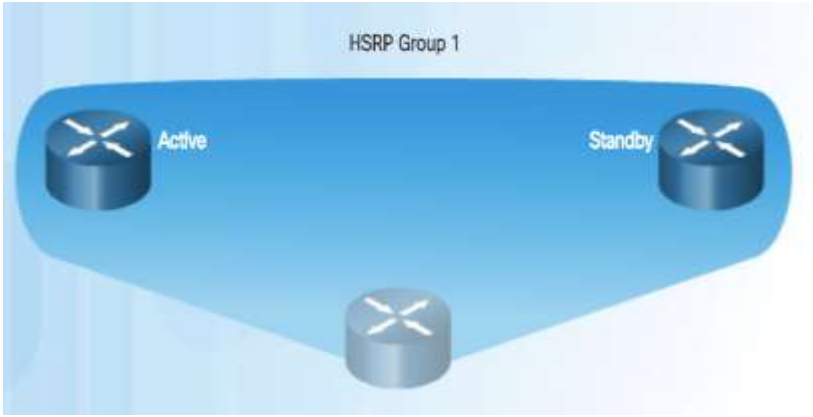
Virtual Router Redundancy Protocol version 2 - Nonproprietary protokol koji dinamički dodeljuje odgovornost jednom ili više virtuelnih rutera VRRP grupi na IPv4 LAN-u.

Jedan ruter se bira za master, dok su ostali backup ruteru, za slučaj otkaza mastera.

VRRPv3 - Podrška za IPv4 i IPv6.

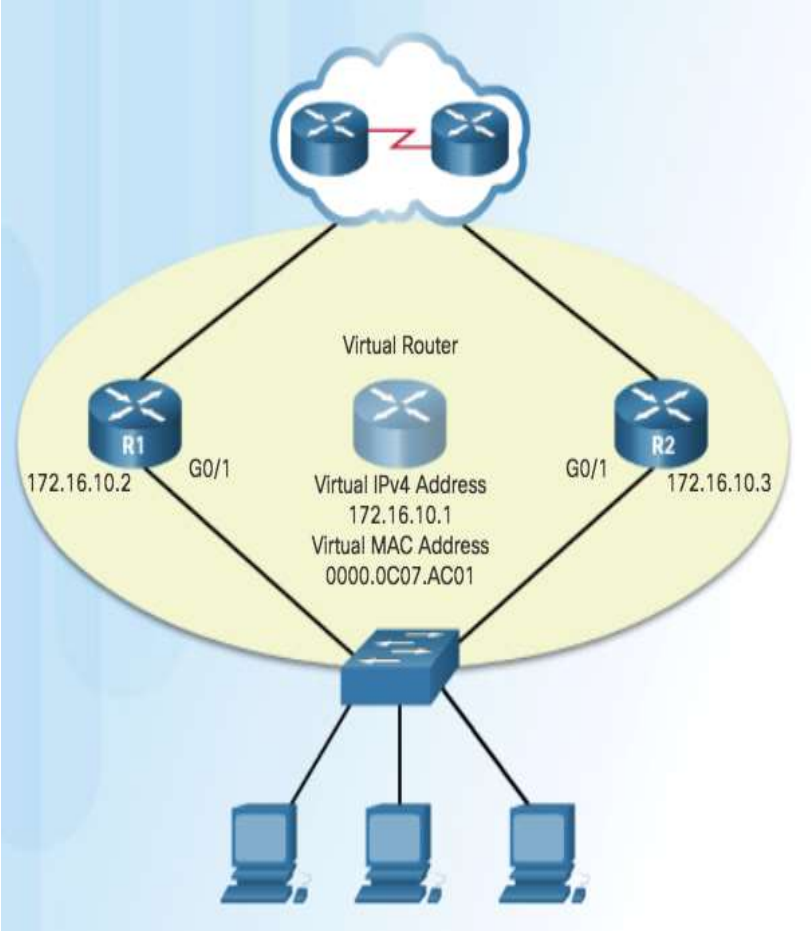
Gateway Load Balancing Protocol (GLBP) - Cisco-proprietary FHRP koji omogućava balansiranje opterećenje između redundantnih rutera.

GLBP for IPv6 - Cisco-proprietary FHRP za IPv6 sa balansiranjem opterećenja.



- HSRP defines a group of routers - one active and one standby.
- Virtual IP and MAC addresses are shared between the two routers.
- To verify HSRP state, use the **`show standby`** command.
- HSRP is Cisco proprietary.
- VRRP is a standard protocol.

HSRP



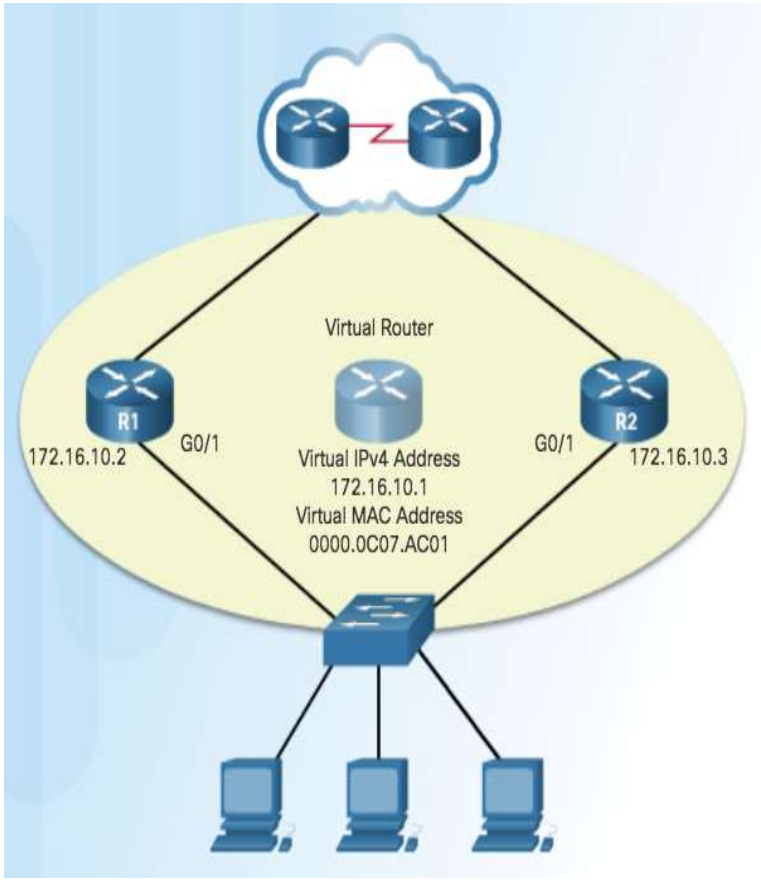
Jedan od rutera je izabran za aktivni ruter i preko njega se rutira saobraćaj.

Drugi ruter će postati standby ruter.

Ako aktivni ruter otkaže, standby ruter će preuzeti njegovu ulogu.

Na hostovima je podešena jedna default gateway adresa, i to VIRTUALNA adresa koji prepoznaju oba rutera.

HSRP verzije



Version	HSRP V1 (Default)	HSRP V2
Group numbers	0 to 255	0 to 4095
Multicast address	224.0.0.2	224.0.0.102 or FF02::66
Virtual MAC address	0000.0C07.AC00 - 0000.0C07.ACFF (last two digits group number)	**IPv4** 0000.0C9F.F000 to 0000.0C9F.FFFF **IPv6** 0005.73A0.0000 - 0005.73A0.0FFF (last three digits group number)
Support for MD5 authentication	No	Yes

HSRP prioritet

active i standby ruteri se određuju na osnovu procesa izbora (election process).

Podrazumevano, ruter sa najvećom IPv4 adresom će biti izabran za aktivni ruter.

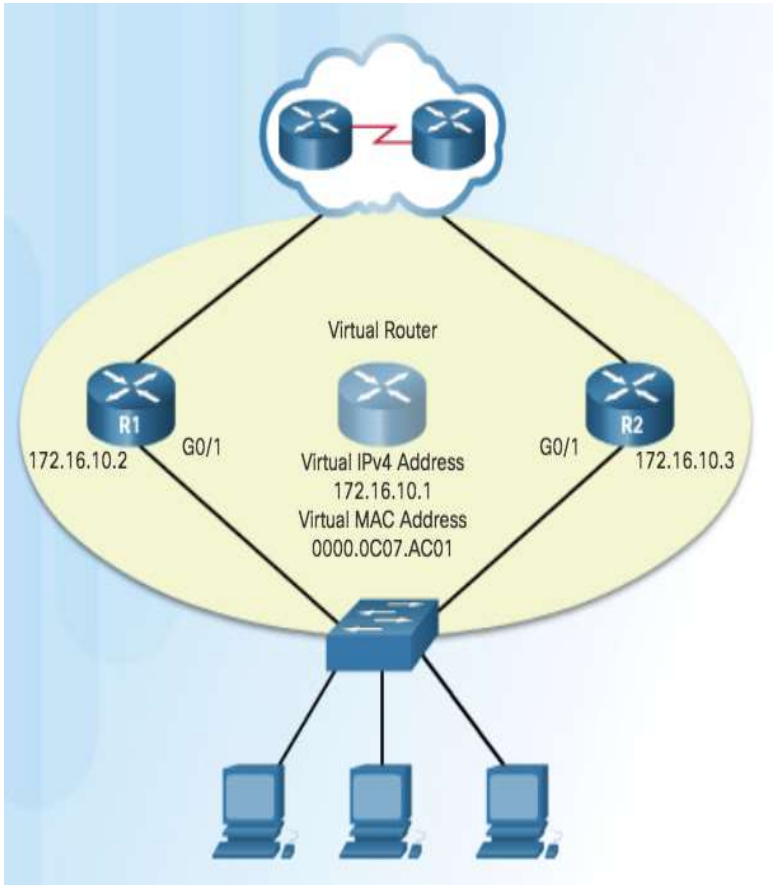
Proces HSRP izbora će koristiti prioritet umesto adrese ako je podešen.

HSRP prioritet

- Default HSRP priority: 100.
- Opseg: 0 to 255
- Veći broj, veći prioritet.
- **standby priority interface** naredba.

HSRP Preemption

- Preemption – mogućnost da se automatski pokrene
- Proces izbora novog rutera



HSRP States and Timers

State	Definition
Initial	This state is entered through a configuration change or when an interface first becomes available.
Learn	The router has not determined the virtual IP address and has not yet seen a hello message from the active router. In this state, the router waits to hear from the active router.
Listen	The router knows the virtual IP address, but the router is neither the active router nor the standby router. It listens for hello messages from those routers.
Speak	The router sends periodic hello messages and actively participates in the election of the active and/or standby router.
Standby	The router is a candidate to become the next active router and sends periodic hello messages.
Active	The router currently forwards packets that are sent to the group virtual MAC address. The router sends periodic hello messages.

- Active i standby HSRP ruteri šalju hello pakete na multikast adresu HSRP-a na svake 3 sekunde po difoltu. Standby ruter će postati aktivan ako ne dobije 3 sukcesivna hello paketa, t.j. nakon 10 sekundi.
- Tajmeri se mogu podešavati.

Naredbe za konfiguraciju HSRP-a

Step 1. Configure HSRP version 2.

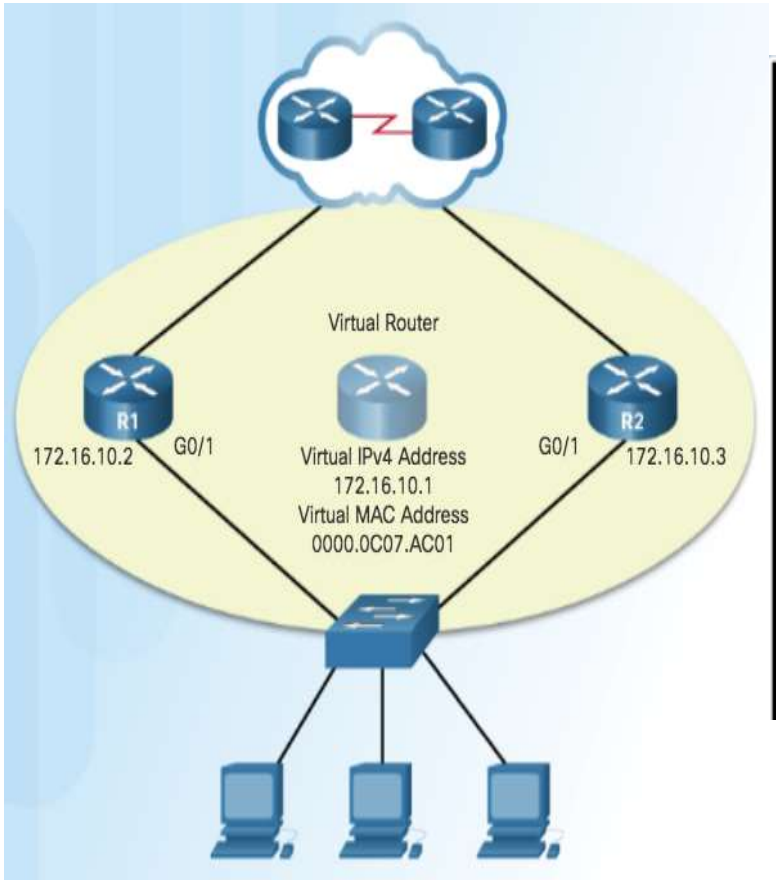
Step 2. Configure the virtual IP address for the group.

Step 3. Configure the priority for the desired active router to be greater than 100.

Step 4. Configure the active router to preempt the standby router in cases where the active router comes online after the standby router.

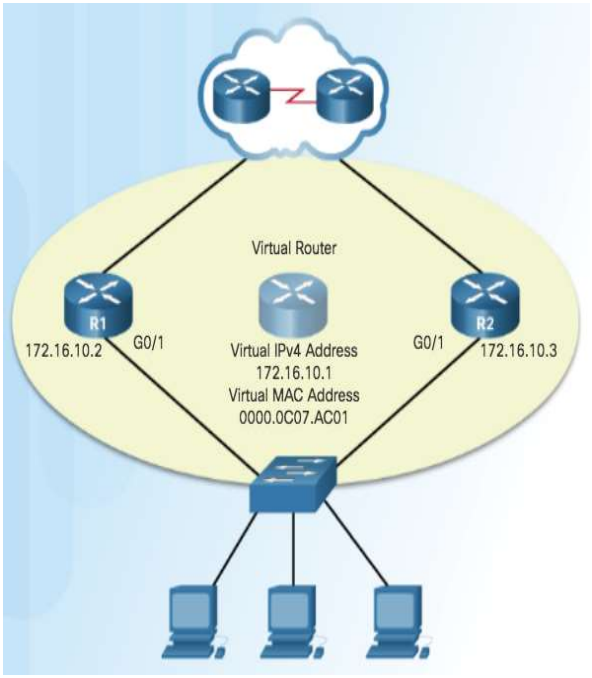
Command	Definition
`Router(config-if)# standby version 2`	Configures HSRP to use version 2. HSRP version 1 is the default.
`Router(config-if)# standby [group-number] ip-address`	Configures the HSRP virtual IP address that will be used by the specified group. If no group is configured, then the virtual IP address is assigned to group 0.
`Router(config-if)# standby [group-number] priority [priority-value]`	Configures the desired active router with a higher priority than default priority of 100. Range is 0 to 255. If no priority is configured or if priority is equal, than the router with the highest IP address has priority.
`Router(config-if)# standby [group-number] preempt`	Configures a router to preempt the currently active router.

Naredbe za konfiguraciju HSRP-a




```
R1(config)# interface g0/1
R1(config-if)# ip address 172.16.10.2 255.255.255.0
R1(config-if)# standby version 2
R1(config-if)# standby 1 ip 172.16.10.1
R1(config-if)# standby 1 priority 150
R1(config-if)# standby 1 preempt
R1(config-if)# no shutdown
!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
R2(config)# interface g0/1
R2(config-if)# ip address 172.16.10.3 255.255.255.0
R2(config-if)# standby version 2
R2(config-if)# standby 1 ip 172.16.10.1
R2(config-if)# no shutdown
```

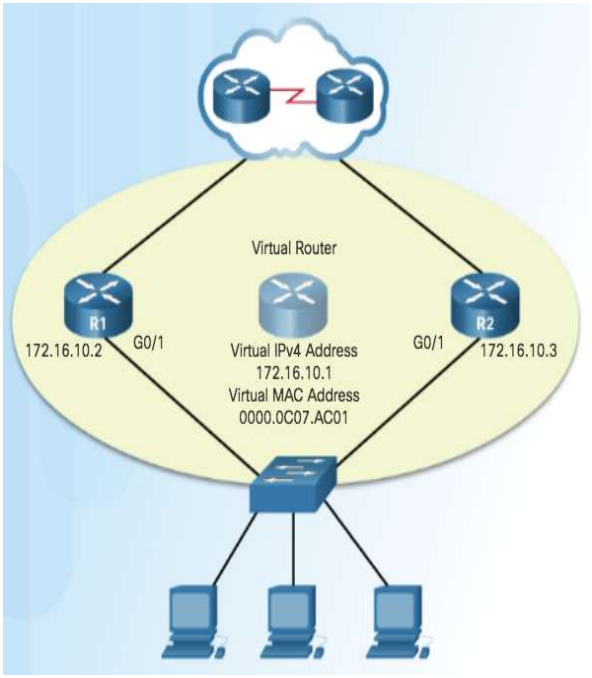
Naredbe za konfiguraciju HSRP-a




```
R1# show standby
GigabitEthernet0/1 - Group 1 (version 2)
  State is Active
    5 state changes, last state change 01:02:18
  Virtual IP address is 172.16.10.1
  Active virtual MAC address is 0000.0c9f.f001
    Local virtual MAC address is 0000.0c9f.f001 (v2 default)
  Hello time 3 sec, hold time 10 sec
    Next hello sent in 1.120 secs
  Preemption enabled
  Active router is local
  Standby router is 172.16.10.3, priority 100 (expires in 9.392 sec)
  Priority 150 (configured 150)
  Group name is "hsrp-Gi0/1-1" (default)
R1#
```

```
R2# show standby
GigabitEthernet0/1 - Group 1 (version 2)
  State is Standby
    5 state changes, last state change 01:03:59
  Virtual IP address is 172.16.10.1
  Active virtual MAC address is 0000.0c9f.f001
    Local virtual MAC address is 0000.0c9f.f001 (v2 default)
  Hello time 3 sec, hold time 10 sec
    Next hello sent in 0.944 secs
  Preemption disabled
  Active router is 172.16.10.2, priority 150 (expires in 8.160 sec)
    MAC address is fc99.4775.c3e1
  Standby router is local
  Priority 100 (default 100)
  Group name is "hsrp-Gi0/1-1" (default)
R2#
```

Naredbe za konfiguraciju HSRP-a




```
R1# show standby brief
                     P indicates configured to preempt.
                     |
Interface   Grp  Pri P State   Active          Standby         Virtual IP
Gi0/1       1    150 P Active  local           172.16.10.3     172.16.10.1
R1#
```

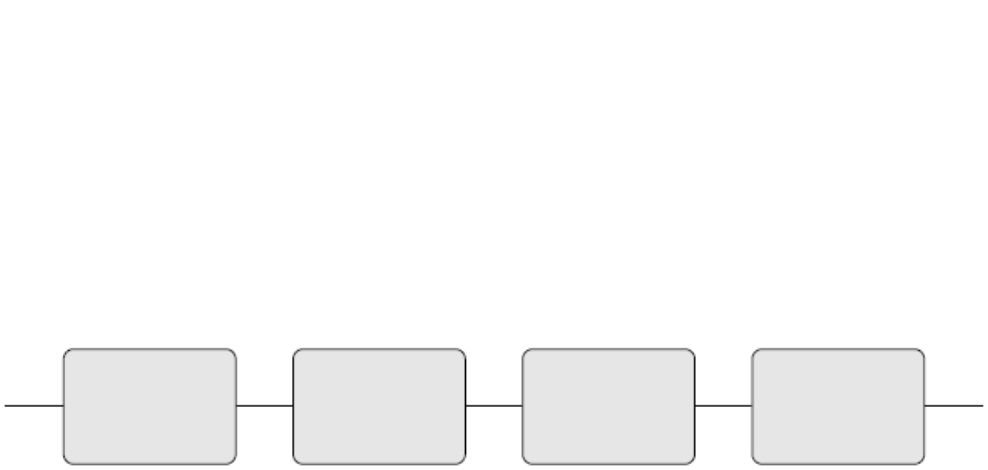
```
R2# show standby brief
                     P indicates configured to preempt.
                     |
Interface   Grp  Pri P State   Active          Standby         Virtual IP
Gi0/1       1    100   Standby 172.16.10.2     local           172.16.10.1
R2#
```

Kakoničke i složene strukture

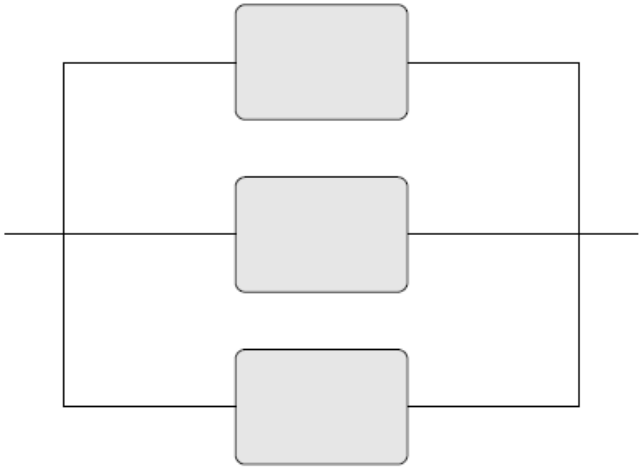
POUZDANOST SLOŽENIH SISTEMA

Serijska i paralelna veza

- Svaki blok je jedna kompomenta čiji uticaj na otkaz celog sistema raznatramo
- Što više tačaka otkaza uzmemo u razmatranje – imamo precizniji model



(a) Series system



(b) Parallel system

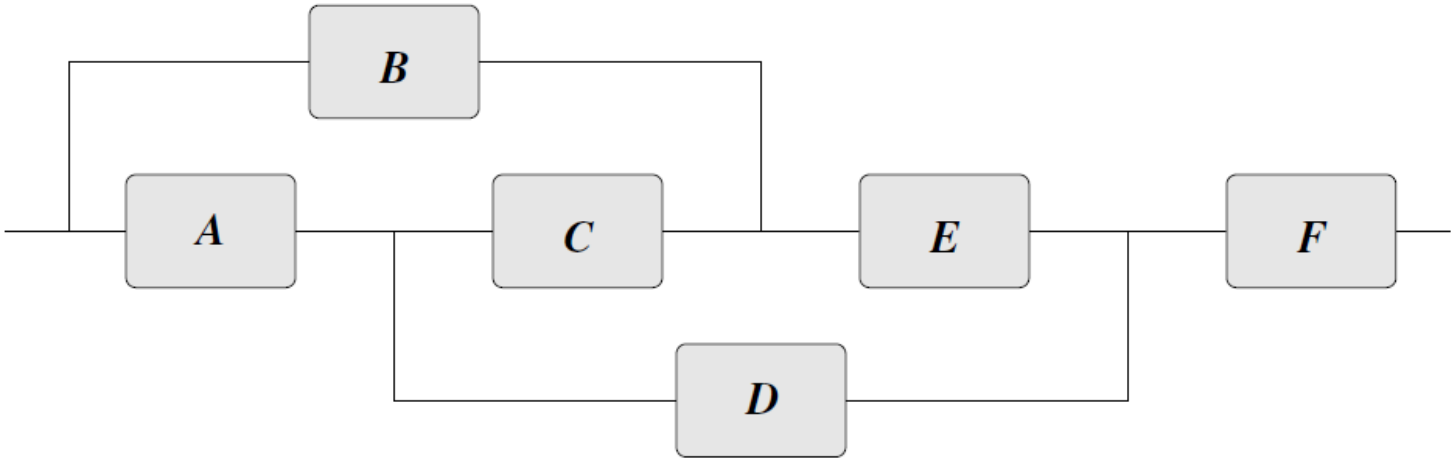
$$R_s(t) = \prod_{i=1}^{N} R_i(t)$$

$$R_p(t) = 1 - \prod_{i=1}^{N} \left(1 - R_i(t)\right)$$

(iste jednačine važe i za određivanje dostupnosti sistema, umesto R(t) je A)

Složene strukture

- Nemaju svi sistemi jasne serijsko-paralelne veze.



- Postoji više načinina za određivanje pouzdanosti/dostupnosti ovakve strukture.
- Jedan od načina je razvoj oko pojedinačnog modula:

$$R_{\text{system}} = R_i \cdot \text{Prob}\{\text{System works}|i \text{ is fault-free}\} + (1 - R_i) \cdot \text{Prob}\{\text{System works}|i \text{ is faulty}\}$$