

SVEUČILIŠTE U ZAGREBU



Diplomski studij

Informacijska i komunikacijska tehnologija:

Telekomunikacije i informatika

Računarstvo:

Programsko inženjerstvo i informacijski sustavi

Računarska znanost

Raspodijeljeni sustavi

11. Modeli za vrednovanje performanci raspodijeljenih sustava

Ak. god. 2020./2021.

Creative Commons





- dijeliti umnožavati, distribuirati i javnosti priopćavati djelo
- prerađivati djelo





- 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, preradu možete distribuirati samo pod licencom koja je ista ili slična ovoj.







U slučaju daljnjeg korištenja ili distribuiranja morate drugima jasno dati do znanja licencne uvjete ovog djela. 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 licence preuzet je s http://creativecommons.org/



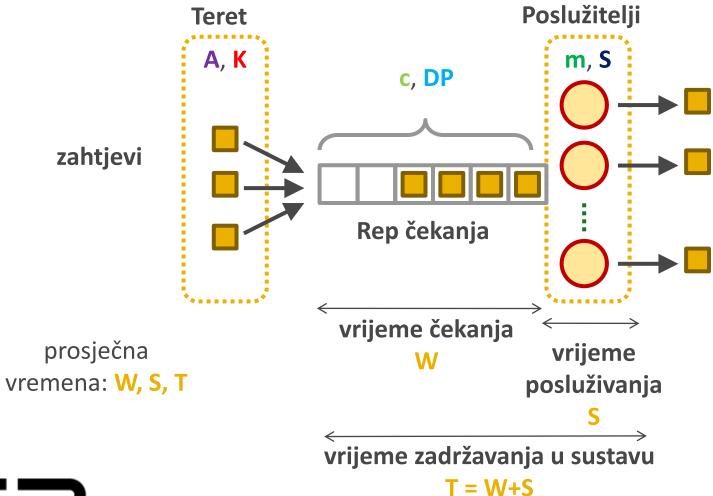
Sadržaj predavanja

- Modeliranje i analiza raspodijeljenih sustava mrežom repova
 - Jednoposlužiteljski sustav
 - Littleov zakon
 - Serijski i paralelni poslužitelji
 - Poslužitelj s povratnom vezom
- Alat PDQ (Pretty Damn Quick)
 - Primjeri uporabe alata PDQ
 - Primjer analize performanci web-aplikacije
- Domaća zadaća



Ponovimo: osnovni pojmovi teorije repova

Proširena notacija: A/S/m/c/K/DP



- A razdioba međudolaznih vremena
- S razdioba vremena posluživanja
- m broj poslužitelja
- c kapacitet sustava
- K veličina ulazne populacije
- DP disciplina posluživanja



Utvrđivanje parametara modela

- Model crne kutije s dva ulazna parametra:
 - Teret sustava karakteriziran prosječnim međudolaznim vremenom $\binom{1}{\lambda}$ i raspodjelom zahtjeva
 - Posluživanje karakterizirano prosječnim vremenom posluživanja bez čekanja
 (5) i raspodjelom vremena posluživanja
- Na temelju dva gornja parametra može se izračunati:
 - Prosječni broj zahtjeva u sustavu N
 - Prosječno vrijeme zadržavanja u sustavu (odziv) T
 - Prosječna iskoristivost sustava (zaposlenost) ρ



Model jednoposlužiteljskog sustava (1)

- Proširena notacija: A/S/m/c/K/DP
- Kendallova notacija: A/S/m npr. M/M/1



- M eksponencijalna raspodjela međudolaznih vremena zahtjeva (Poissonov proces)
- M eksponencijalna raspodjela vremena posluživanja
- jedan poslužitelj
 (c = ∞ kapacitet sustava, K = ∞ zahtjeva, DP = FIFO)



Model jednoposlužiteljskog sustava (2)



Osnovne značajke modela:

- t ukupno vrijeme promatranja rada sustava
- α(t) broj dolazaka zahtjeva u vremenu t
- $\beta(t)$ broj odlazaka zahtjeva u vremenu t
- σ(t) vrijeme kroz koje je poslužitelj zaposlen u vremenu t



Model jednoposlužiteljskog sustava (3)



Izvedene veličine:

$$\begin{array}{lll} \lambda = \alpha(t) \: / \: t & \text{intenzitet dolazaka (z/s)} \\ \delta = \beta(t) \: / \: t & \text{intenzitet odlazaka [propusnost sustava] (z/s)} \\ S = \sigma(t) \: / \: \beta(t) & \text{prosječno vrijeme posluživanja (s/z)} \\ \rho = \sigma(t) \: / \: t & \text{prosječna iskoristivost [zaposlenost / zauzetost / opterećenje] poslužitelja} \end{array}$$



8 od 87

Primjer 1: Posluživanje zahtjeva na disku

- Disk za trajno spremanje podataka ispunjava 50 zahtjeva u sekundi.
 Prosječno vrijeme obrade zahtjeva operacija pisanja i čitanja je 10 ms.
- Kolika je prosječna zaposlenost diska?

- Rješenje:
 - Propusnost sustava $\delta = 50 \text{ z/s}$
 - Prosječno vrijeme obrade zahtjeva S = 10 ms/z
 - Prosječna zaposlenost diska ρ:

$$\rho = \delta \times S = 50 \text{ z/s} \times 0.01 \text{ s/z} = 0.5 (50 \%)$$



9 od 87

Littleov zakon (John Little, 1961.)

- Napomena: sve veličine su prosječne vrijednosti!
- Prosječni broj zahtjeva u sustavu jednak je umnošku intenziteta dolazaka zahtjeva i prosječnog vremena zadržavanja zahtjeva u sustavu:

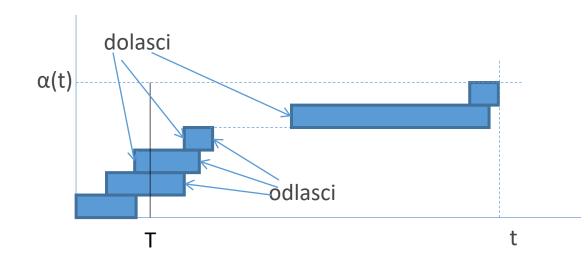
 $N = \lambda \times T$

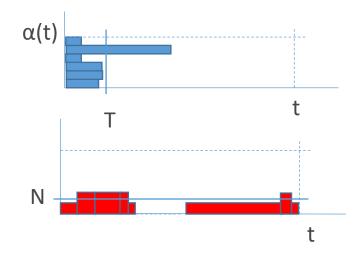
- λ [z/s] Intenzitet dolazaka zahtjeva
- T [s] Prosječno vrijeme zadržavanja zahtjeva u sustavu
- N [z] Prosječni broj zahtjeva u sustavu
- Stabilnost sustava: broj prispjelih zahtjeva u vremenu jednak je broju zahtjeva koji napuštaju sustav ($\lambda = \delta$)

 $N = \delta x T$



Dokaz Littleovog zakona







$$\alpha(t) \times T = N \times t$$

 $\alpha(t)/t \times T = N$
 $T \times \alpha(t)/t = N$

$$T \times \lambda = N$$



Jednoposlužiteljski sustav

- Prosječno vrijeme zadržavanja zahtjeva u sustavu (T)
 - Vrijeme obrade svih neobrađenih zahtjeva u sustavu (N) uvećano za vrijeme obrade novog zahtjeva

$$T = S + W = S + S \times N$$

- U stabilnom stanju, primjenom supstitucije N = $\delta \times T$ (Littleov zakon) T = S + S $\times \delta \times T \rightarrow T$ = S/(1 – $\delta \times S$)
- Primjenom supstitucije $\rho = \delta \times S$ $T = S/(1-\rho)$
- Množenje obje strane sa $\delta \rightarrow [\delta \times T = (\delta \times S)/(1-\rho)]$ $N = \rho/(1-\rho)$ – prosječni broj zahtjeva u sustavu
- Množenje obje strane sa S \rightarrow [S x N = (S x ρ)/(1 ρ)] W = S x N = (S x ρ)/(1 - ρ) – prosječno vrijeme čekanja



Primjer 2: Komunikacijski kanal

- Mjerenjem na pristupnoj točki mreže dobivamo prosječni protok od 125 paketa u sekundi i prosječno vrijeme posluživanja paketa 0,002 sekunde.
 - Što je sve moguće zaključiti o promatranom kanalu?
- Rješenje:

$$\delta = 125 \text{ p/s}, S = 0.002 \text{ s/p}$$

Prosječna zauzetost komunikacijskog sustava (ρ)

$$\rho = \delta \times S = 125 \text{ p/s} \times 0,002 \text{ s/p} = 0,25 (25 \%)$$

Prosječno vrijeme zadržavanja paketa u sustavu (T)

$$T = S/(1 - \rho) = (0,002 \text{ s/p}) / (1 - 0,25) = 0,00266666 \text{ s}$$

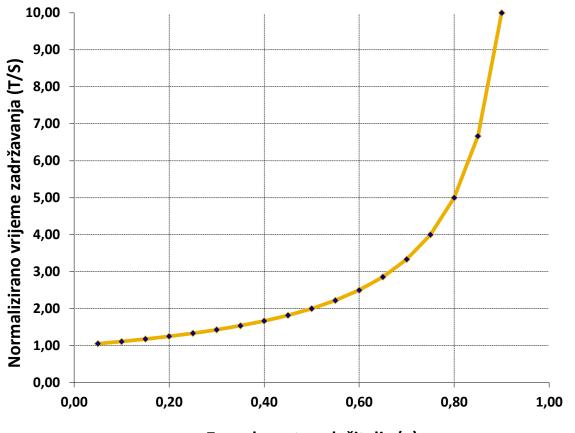
Prosječni broj paketa u sustavu (N)

$$N = \delta \times T = 125 \text{ p/s} \times 0,0026 \text{ s} = 0,333 \text{ p}$$



Odziv sustava s repovima je izrazito nelinearan

Graf: normalizirano vrijeme zadržavanja (T/S) kao funkcija opterećenja (zaposlenosti) poslužitelja ρ



vrijeme zadržavanja / vrijeme posluživanja

 $T/S = 1/(1-\rho)$



Zaposlenost poslužitelja (ρ)

Sadržaj predavanja

- Modeliranje i analiza raspodijeljenih sustava mrežom repova
 - Jednoposlužiteljski sustav
 - Littleov zakon
 - Serijski i paralelni poslužitelji
 - Poslužitelj s povratnom vezom
- Alat PDQ (Pretty Damn Quick)
 - Primjeri uporabe alata PDQ
 - Primjer analize performanci web-aplikacije
- Domaća zadaća



15 od 87

Serijski repovi i poslužitelji

Littleov zakon

$$N = \lambda \times (T_1 + T_2 + T_3)$$

• U stabilnom stanju sustava ($\lambda = \delta$, $\delta_1 = \delta_2 = \delta_3 = \delta$) N = δ x T = δ x (T₁ + T₂ + T₃)

• Uz
$$T_n = S_n / (1 - \delta x S_n)$$

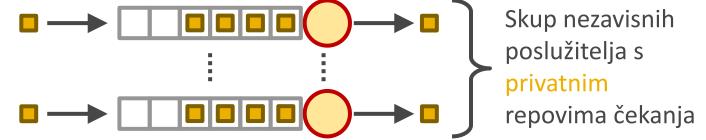
 $N = \delta((S_1/(1 - \delta x S_1)) + (S_2/(1 - \delta x S_2)) + (S_3/(1 - \delta x S_3))$
 $T = T1 + T2 + T3$



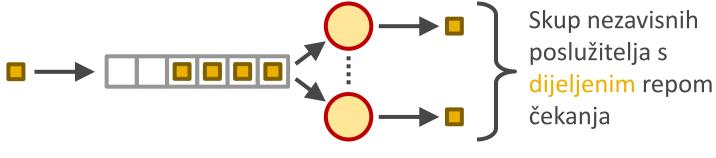


Višestruki paralelni repovi i poslužitelji

- Multiračunalo
 - Model koji se primjenjuje u supermarketima



- Multiprocesor
 - Model koji se primjenjuje u bankama





Sustav s dva paralelna repa i poslužitelja

Vrijeme zadržavanja u sustavu (T):

$$T = S + (S \times (0.5 \times N))$$

- Primjenom Littleovog zakona N = δ x T T = $S/(1 0.5 \times \delta \times S)$
- Primjenom supstitucije $\delta \times S = \rho$ $T = S/(1 - 0.5 \times \rho)$
- Ukupna zaposlenost ρ sustava podijeljena s brojem poslužitelja m određuje faktor iskorištenja ρ' koji predstavlja vjerojatnost da je poslužitelj zaposlen $\rho' = \rho/m$, $T = S/(1 \rho')$
- Sustav s beskonačno mnogo repova i poslužitelja $m \to \infty$; $\rho' \to 0$; $T \to S \leftarrow Nema čekanja na posluživanje!$



Jedan poslužitelj:

 $T = S/(1-\rho)$

Sustav s jednim repom i m paralelnih poslužitelja

• Aproksimativno rješenje:

$$T = S + N \times (S/m) \times \rho'^{(m-1)}, \rho' = \rho/m$$

• Primjenom supstitucija $N = \delta \times T i S = \rho/\delta$

$$T = S + (\delta x T) x (\rho/(\delta x m)) x \rho'^{(m-1)}$$

$$T = S + T x (\rho/m) x \rho'^{(m-1)}$$

$$T = S + T x \rho'^{m}$$

$$T x (1 - \rho'^{m}) = S$$

$$T = S/(1 - \rho'^{m}) \qquad \text{množenjem s } \delta$$

$$\delta x T = \delta x S / (1 - \rho'^{m}) \qquad \text{uz } N = \delta x T i \rho = \delta x S$$

$$N = \rho / (1 - \rho'^{m}) \qquad \text{uz } \rho = m x \rho'$$

$$N = (m x \rho') / (1 - \rho'^{m})$$



Sustav s jednim repom i dva poslužitelja

- Vrijeme zadržavanja u sustavu (T) ovisi o dva čimbenika
 - Broj poslužitelja: dodatni poslužitelj smanjuje vrijeme posluživanja za faktor 0,5
 - Vjerojatnost da je poslužitelj zaposlen: $\rho' = \rho / 2$

$$S(\rho') = (0,5 \times S) \times \rho'$$

 $T = S + N \times S(\rho') = S + 0,5 \times S \times \rho' \times N$

• Primjenom supstitucije $N = \delta \times T$

$$T = S + (0.5 \times S \times \rho' \times \delta \times T)$$

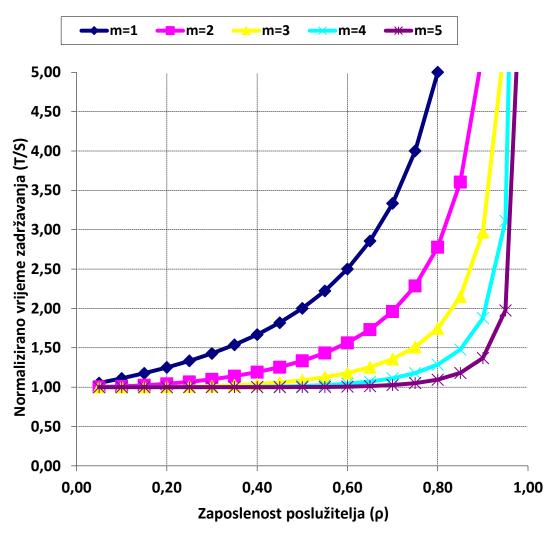
• Primjenom supstitucije $0.5 \times S \times \delta = 0.5 \times \rho = \rho'$

T = S + T x
$$\rho'^2$$

T = S/(1 – ρ'^2) \rightarrow množenjem sa δ i supstitucijom S x δ = 2 x ρ'
N = 2 x ρ' / (1 – ρ'^2)



Usporedba sustava M/M/m - multiprocesor





21

Egzaktno rješenje sustava s m paralelnih poslužitelja

- Erlangova formula
 - Analitičko rješenje za vrijeme zadržavanja T u sustavu s m paralelnih poslužitelja

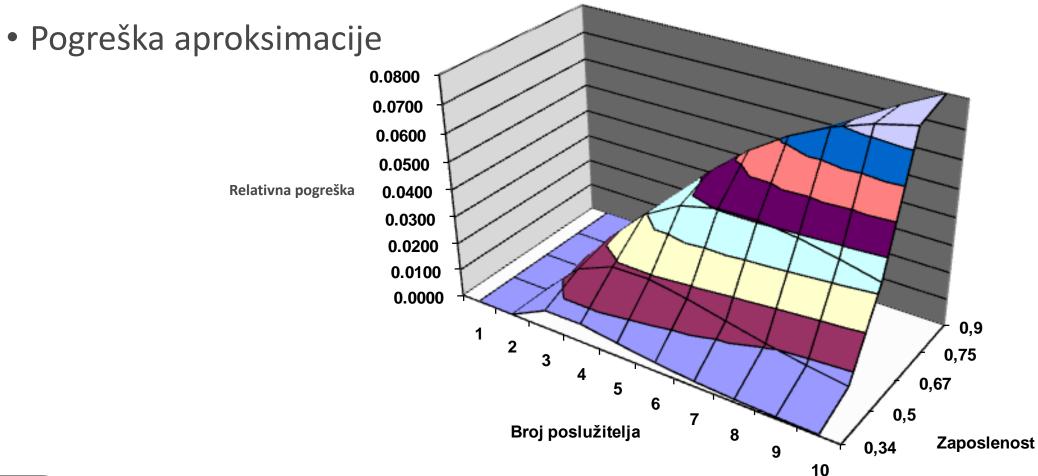
$$T = S * \left[1 + \frac{C(m, \rho')}{m * (1 - \rho')}\right]$$

Koeficijent C(m, ρ')

$$C(m, \rho') = \frac{\frac{(N * \rho')^m}{m!}}{(1 - \rho) * \sum_{k=0}^{m-1} \frac{(m * \rho')^k}{k!} + \frac{(m * \rho')^m}{m!}}$$

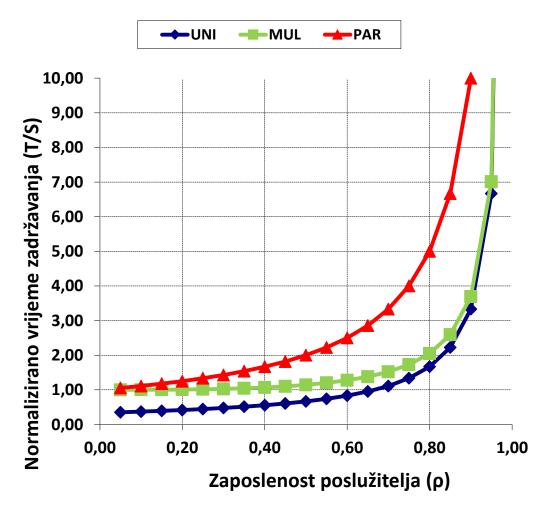


Aproksimacija i egzaktno rješenje sustava





Koji je model bolji: multiračunalo ili multiprocesor?



UNI Uniprocesor (3x procesna moć)

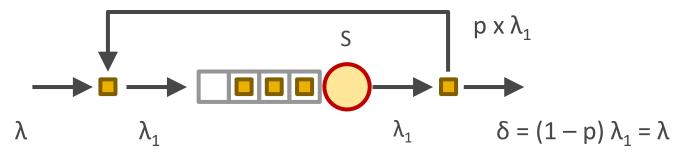
MUL Multiprocesor s 3 procesora

PAR
3 paralelna
računala
(multiračunalo)



Poslužitelji s povratnom vezom

• Dio zahtjeva nakon posluživanja ponovno se vraća u rep za čekanje



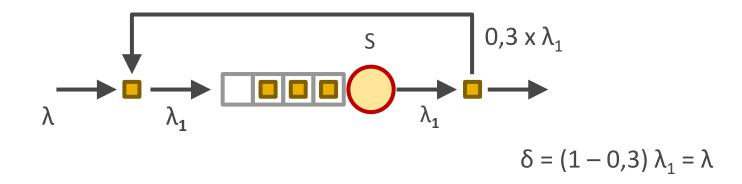
$$\lambda_1 = \lambda + p \times \lambda_1 = \lambda / (1 - p) = \lambda \times v_1$$
 $\rho = \lambda_1 \times S$
 $T_1 = S / (1 - \rho)$
 $v_1 = 1 / (1 - p)$
 $T = T_1 \times v_1 = T_1 / (1 - p)$

iskoristivost sustava vrijeme zadržavanja za jedan prolaz broj prolaza vrijeme zadržavanja u sustavu



Primjer 3: Komunikacijski kanal s pogreškom (1)

- Paketi dolaze u komunikacijski kanal s intenzitetom 0,5 paketa u sekundi i zahtijevaju 0,75 sekundi za obradu. Za 30 % paketa dogodi se pogreška pri prijenosu i takvi paketi se umeću u rep za ponovno slanje.
- Koliko vremena paket prosječno provede u kanalu?





26 od 87

Primjer 3: Komunikacijski kanal s pogreškom (2)

• Rješenje:

- Broj pristiglih paketa u sekundi $\lambda = 0.5 \text{ p/s}$
- Prosječno vrijeme obrade paketa S = 0,75 s/p
- Vjerojatnost pogreške paketa pri prijenosu p = 0,3

.....

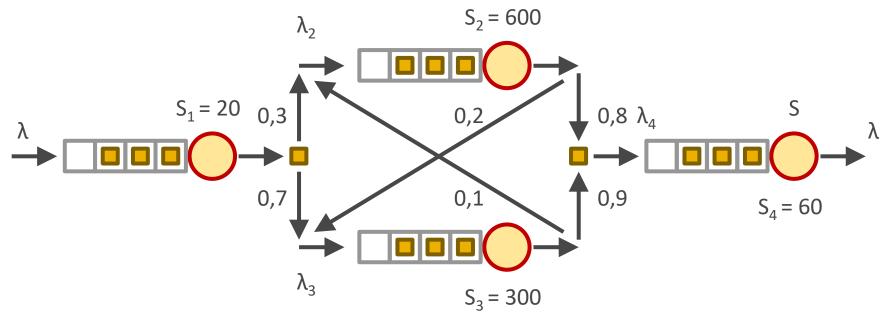
$$\lambda_1 = \lambda / (1 - p) = 0.5 / 0.7 = 0.714 p/s$$

- Prosječna zauzetost kanala (ρ) $\rho = \lambda_1 \times S = 0.714 \text{ p/s} \times 0.75 \text{ s/p} = 0.536 (53.6 \%)$
- Prosječno vrijeme čekanja u repu (W) $W = S \times \rho / (1 \rho) = 0.866 \text{ s/p}$
- Prosječno vrijeme zadržavanja paketa u kanalu (T_1) $T_1 = W + S = 0.866 \text{ s/p} + 0.75 \text{ s/p} = 1.616 \text{ s/p}$
- Prosječno vrijeme u kanalu: $T = T_1/(1-p) = 2.31$



Višestruke povratne veze (1)

- Mreža repova poruka
 - Mrežna struktura proizvoljne složenosti s povratnim granama



• Izračunajte prosječno vrijeme zadržavanja zahtjeva u sustavu!



Višestruke povratne veze (2)

- Što znamo na temelju prethodnih razmatranja?
- U stabilnom stanju sustava

$$\rho_n = \delta_n \times S_n (\delta_n = \lambda_n \text{ za stabilni slučaj})$$

$$\rho_1 = \lambda \times S_1 = 20 \times \lambda$$

$$\rho_2 = 600 \times \lambda_2 = 600 (0.3 \times \lambda + 0.1 \times \lambda_3)$$

$$\rho_3 = 300 \times \lambda_3 = 300 (0.7 \times \lambda + 0.2 \times \lambda_2)$$

$$\rho_4 = 60 \times \lambda$$

• Nakon rješenja za λ_2 i λ_3 i izračunavanja ρ_1 - ρ_4 , izračunavamo N_1 - N_4 iz:

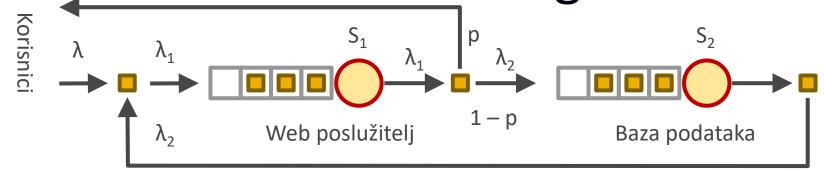
$$N_n = \rho_n / (1 - \rho_n)$$

Vrijeme zadržavanja zahtjeva u sustavu (T):

$$T = N/\lambda = (N_1 + N_2 + N_3 + N_4)/\lambda$$



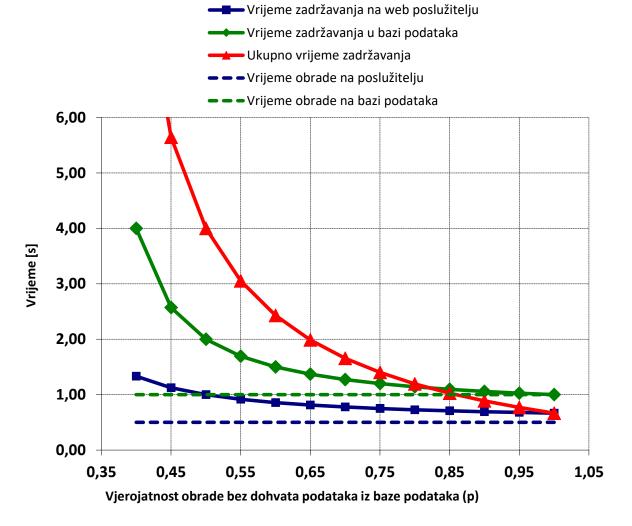
Jednostavan model web-usluge



- Intenzitet dolazaka zahtjeva (λ):
 - $\lambda_1 = \lambda + \lambda_2 = \lambda + (1-p) \lambda_1 = \lambda/p$
 - $\lambda_2 = (1 p) \lambda_1 = ((1 p)/p) \times \lambda$
- Vrijeme zadržavanja zahtjeva u sustavu (T):
 - $\delta_1 = \lambda_1$; $\delta_2 = \lambda_2$
 - $\rho_1 = \delta_1 \times S_1 = \lambda \times S_1/p$; $\rho_2 = ((1-p)/p) \times \lambda \times S_2$
 - $T_1 = S_1/(1 \rho_1)$; $T_2 = S_2/(1 \rho_2)$
 - $T = T_1 \times (1 + (1 p)/p) + T_2 \times (1 p)/p$



Utjecaj parametara na ponašanje web-usluge





Sadržaj predavanja

- Modeliranje i analiza raspodijeljenih sustava mrežom repova
 - Jednoposlužiteljski sustav
 - Littleov zakon
 - Serijski i paralelni poslužitelji
 - Poslužitelj s povratnom vezom
- Alat PDQ (Pretty Damn Quick)
 - Primjeri uporabe alata PDQ
 - Primjer analize performanci web-aplikacije
- Domaća zadaća



Alat Pretty Damn Quick (PDQ)

- Omogućuje izgradnju modela za vrednovanje performanci računalnih sustava
- Modeli se grade primjenom načela teorije repova
- Značajke modela izračunavaju se primjenom analitičkih postupka i algoritama
- Dodatne informacije
 - PDQ: Pretty Damn Quick Performance Analyzer

http://www.perfdynamics.com/Tools/PDQ.html

PDQ User Manual Online

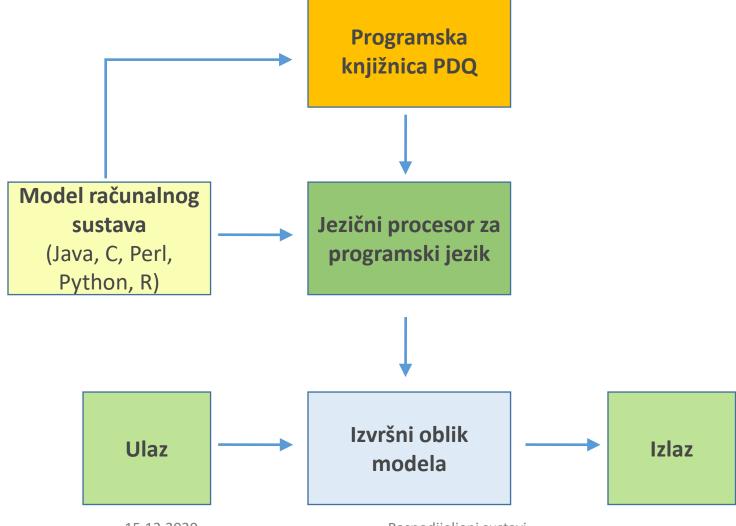
http://www.perfdynamics.com/Tools/PDQman.html

• J. Gunther: Analyzing Computer System Performance With Perl, Springer, 2010.



33 od 87

Alat Pretty Damn Quick (PDQ)





15.12.2020. Raspodijeljeni sustavi

Sadržaj predavanja

- Modeliranje i analiza raspodijeljenih sustava mrežom repova
 - Jednoposlužiteljski sustav
 - Littleov zakon
 - Serijski i paralelni poslužitelji
 - Poslužitelj s povratnom vezom
- Alat PDQ (Pretty Damn Quick)
 - Primjeri uporabe alata PDQ
 - Primjer analize performanci web-aplikacije
- Domaća zadaća



Primjer 1: Posluživanje zahtjeva na disku (1)

Disk za trajno spremanje podataka obrađuje 50 zahtjeva u sekundi.
 Prosječno vrijeme obrade zahtjeva operacija pisanja i čitanja je 10 ms.

Kolika je prosječna zaposlenost diska?





Primjer 1: Posluživanje zahtjeva na disku (2)

- Analitičko rješenje
 - Propusnost sustava

$$\delta = 50 \text{ z/s}$$

Prosječno vrijeme obrade zahtjeva

$$S = 10 \text{ ms/z}$$

Prosječna zaposlenost diska

$$\rho = \delta \times S = 50 \text{ z/s } \times 0.01 \text{ s/z} = 0.5 (50 \%)$$



Primjer 1: Posluživanje zahtjeva na disku (3)

```
main() {
  extern int nodes, streams;
  float L = 50;
  float S = 0,01;
  PDQ Init("Diskovni podsustav");
  nodes = PDQ CreateNode("Posluzitelj", CEN, FCFS);
  streams = PDQ CreateOpen("Operacije", L);
  PDQ_SetDemand("Posluzitelj", "Operacije", S);
  PDQ Solve(CANON);
  PDQ_Report();
```



Primjer 2: Čekanje na posluživanje zahtjeva s diska (1)

Disk iz prethodnog slučaja ima prosječno 1 zahtjev u repu

Koliko je prosječno vrijeme čekanja na obradu zahtjeva?





Primjer 2: Čekanje na posluživanje zahtjeva s diska (2)

- Analitičko rješenje
 - Intenzitet zahtjeva $\lambda = 50 \text{ z/s}$
 - Broj zahtjeva u repu N = 1 z
 - Vrijeme zadržavanja zahtjeva u sustavu

$$T = N/\lambda = (1 z) / (50 z/s) = 20 ms$$

 Vrijeme zadržavanja uključuje vrijeme čekanja u repu (W) i vrijeme obrade zahtjeva (S):

$$T = W + S$$

- Vrijeme čekanja na obradu
- W = T S = 20 ms 10 ms = 10 ms



Primjer 2: Čekanje na posluživanje zahtjeva s diska (3)

```
main() {
  extern int nodes, streams;
  float L = 50;
  float S = 0,01;
  PDQ Init("Diskovni podsustav");
  nodes = PDQ CreateNode("Posluzitelj", CEN, FCFS);
  streams = PDQ CreateOpen("Operacije", L);
  PDQ SetDemand("Posluzitelj", "Operacije", S);
  PDQ Solve(CANON);
  PDQ Report();
```



Primjer 3: Komunikacijski kanal (1)

 Mjerenjem na pristupnoj točki mreže dobivamo prosječni protok od 125 paketa u sekundi i prosječno vrijeme posluživanja 0,002 sekunde.

• Što je sve moguće zaključiti o promatranom kanalu?





Primjer 3: Komunikacijski kanal (2)

- Analitičko rješenje
 - Prosječni protok paketa $\delta = 125 \text{ p/s}$
 - Prosječno vrijeme posluživanja paketa S = 0,002 s/p
 - Prosječna zaposlenost komunikacijskog sustava $\rho = \delta \times S = (125 \text{ p/s}) \times (0,002 \text{ s/p}) = 0,25 (25 \%)$
 - Prosječno vrijeme zadržavanja paketa u sustavu $T = S/(1 \rho) = (0,002 \text{ s/p})/(1 0,25) = 0,0026666 \text{ s}$
 - Prosječni broj paketa u repu $N = \delta \times T = 125 \text{ p/s} \times 0,0026 \text{ s} = 0,333 \text{ p}$



Primjer 3: Komunikacijski kanal (3)

```
main() {
  extern int nodes, streams;
  float L = 125;
  float S = 0,002;
  PDQ Init("Mrezni podsustav");
  nodes = PDQ CreateNode("Posluzitelj", CEN, FCFS);
  streams = PDQ CreateOpen("Operacije", L);
  PDQ SetDemand("Posluzitelj", "Operacije", S);
  PDQ Solve(CANON);
  PDQ Report();
```



Primjer 4: Vrijeme čekanja i broj zahtjeva (1)

• Sustav ima prosječno vrijeme posluživanja 1 sekunda i intenzitet zahtjeva je 0,5 zahtjeva u sekundi.

 Kolika je prosječna vrijednost vremena zadržavanja i prosječna vrijednost broja zahtjeva u sustavu?





Primjer 4: Vrijeme čekanja i broj zahtjeva (2)

- Analitičko rješenje
 - Prosječno vrijeme posluživanja S = 1 s/z
 - Intenzitet zahtjeva $\lambda = 0.5$ z/s
 - Prosječna zaposlenost sustava

$$\rho = S \times \lambda = (1 \text{ s/z}) \times (0.5 \text{ z/s}) = 0.5 (50 \%)$$

Prosječno vrijeme zadržavanja paketa u sustavu

$$T = S / (1 - \rho) = 1 / (1 - 0.5) = 2 s$$

Prosječna vrijednost broja zahtjeva u sustavu

$$N = \rho / (1 - \rho) = 0.5/(1 - 0.5) = 1 z$$



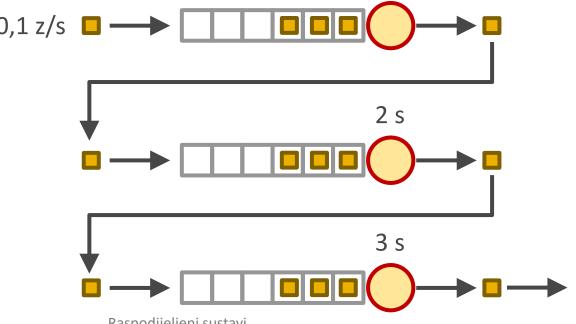
Primjer 4: Vrijeme čekanja i broj zahtjeva (3)

```
main() {
  extern int nodes, streams;
  float L = 0,5;
  float S = 1.0;
  PDQ Init("Posuzitelj s repom");
  nodes = PDQ CreateNode("Posluzitelj", CEN, FCFS);
  streams = PDQ CreateOpen("Zadaci", L);
  PDQ SetDemand("Posluzitelj", "Zadaci", S);
  PDQ_Solve(CANON);
  PDQ Report();
```



Primjer 5: Posluživanje u seriji (1)

- Sustav sadrži 3 serijske procesne jedinice s prosječnim vremenima posluživanja 1 s, 2 s i 3 s.
 - Koliko će biti vrijeme zadržavanja u sustavu uz intenzitet zahtjeva od 0,1 z/s?
 - Koliki će biti prosječni broj zahtjeva u sustavu?





15.12.2020.

Primjer 5: Posluživanje u seriji (2)

- Analitičko rješenje
 - Prosječna vremena posluživanja

$$S1 = 1 \text{ s/z}, S2 = 2 \text{ s/z}, S3 = 3 \text{ s/z}$$

Propusnost sustava

$$\delta = 0.1 \text{ z/s}$$

• Vremena zadržavanja TN = SN $/(1 - \delta \times SN)$

$$T1 = 1.11 \text{ s}, T2 = 2.5 \text{ s}, T3 = 4.29 \text{ s}$$

Prosječni broj zahtjeva u sustavu

N =
$$\delta$$
 x (T1+ T2 + T3) = 0,1 x (1.11 + 2.5 + 4.29) = 0,79 z



Primjer 5: Posluživanje u seriji (3)

```
extern int nodes, streams;
float L = 0,1; float S1 = 1.0; float S2 = 2.0; float S3 = 3.0;
PDQ Init("Serija tri posluzitelja");
streams = PDQ_CreateOpen("Zadaci", L);
nodes = PDQ_CreateNode("Posluzitelj1", CEN, FCFS);
nodes = PDQ_CreateNode("Posluzitelj2", CEN, FCFS);
nodes = PDQ CreateNode("Posluzitelj3", CEN, FCFS);
PDQ SetDemand("Posluzitelj1", "Zadaci", S1);
PDQ_SetDemand("Posluzitelj2", "Zadaci", S2);
PDQ_SetDemand("Posluzitelj3", "Zadaci", S3);
PDQ Solve(CANON);
PDQ Report();
```



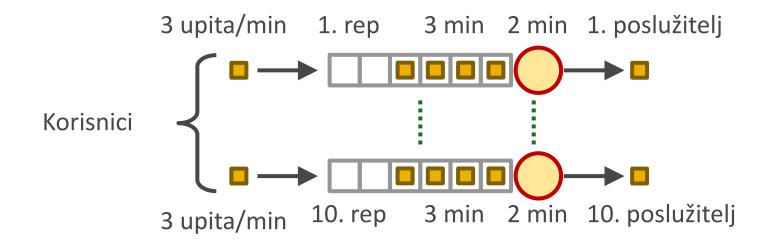
Primjer 6: Aplikacija korisničke podrške (1)

 Web-aplikacija uključuje podršku korisnicima putem usluge chat. Kupci sami odabiru jedan od 10 repova čekanja. Mjerenja pokazuju da zahtjevi - upiti prosječno dolaze 3 u minuti te da svaki kupac prosječno čeka 3 minute u repu i prosječno provodi 2 minute u konverzaciji.

 Koliko bi dodatnih osoba – poslužitelja trebalo zaposliti da se prosječno vrijeme čekanja svede na 1 minutu?



Primjer 6: Aplikacija korisničke podrške (2)





Primjer 6: Aplikacija korisničke podrške (3)

- Analitičko rješenje
 - Prosječno vrijeme posluživanja S = 2 min/z
 - Broj pristiglih zahtjeva u jednom repu $\lambda = 3$ z/min
 - Prosječna zaposlenost sustava

$$\rho = S \lambda = (2 \text{ min/z}) (3 \text{ z/min}) = 6$$

Faktor iskorištenja

$$\rho' = \rho/m = 6/10 = 0.6$$

Prosječno vrijeme zadržavanja korisnika u sustavu

$$T = S / (1 - \rho') = 2 / (1 - 0.6) = 5 min$$

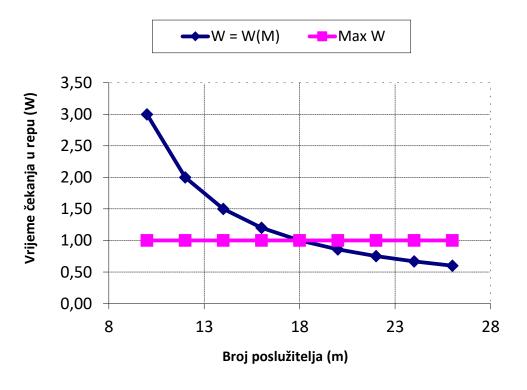
Prosječno vrijeme čekanja u repu

$$W = T - S = 5 min - 2 min = 3 min$$



Primjer 6: Aplikacija korisničke podrške (4)

- Broj poslužitelja
 - Za zadani sustav ne postoji analitičko rješenje. Rješenje se određuje primjenom numeričkih metoda ili primjenom metode pokušaja i promašaja.





Rješenje: potrebno 18 poslužitelja

Primjer 6: Aplikacija korisničke podrške (5)

- Odabrano rješenje
 - Broj poslužitelja (tehničara) m = 18
 - Prosječno vrijeme posluživanja S = 2 min/z
 - Propusnost sustava $\delta = 3$ z/min
 - Prosječna zaposlenost sustava
 - $\rho = \delta * S = (3 \text{ z/min}) * (2 \text{ min/z}) = 6$
 - Faktor iskorištenja $\rho' = \rho/m = 6/18 = 1/3$
 - Prosječno vrijeme zadržavanja korisnika u sustavu
 - $T = S / (1 \rho') = 2 / (1 1/3) = 3 min$
 - Prosječno vrijeme čekanja u repu (W)
 - W = T S = 3 2 = 1 min



Primjer 6: Aplikacija korisničke podrške (6)

```
main() {
    extern int
               nodes;
    extern int
              streams;
    double
           L = 3;
    double
                S = 2;
    char nName[30];
    char cName[30];
    int i;
    int count = 10;
    PDQ_Init("Aplikacija korisnicke podrske");
```

•••



Primjer 6: Aplikacija korisničke podrške (7)

```
for( i=0; i<count; i++ ) {</pre>
  sprintf(nName, "Serv %2d", i);
  sprintf(cName, "Clnt %2d", i);
  nodes = PDQ CreateNode(nName, CEN, FCFS);
  streams = PDQ CreateOpen(cName, L/count);
for( i=0; i<count; i++ ) {</pre>
  sprintf(nName, "Serv %2d", i);
  sprintf(cName, "Clnt %2d", i);
  PDQ SetDemand(nName, cName, S);
PDQ_Solve(CANON);
PDQ Report();
```



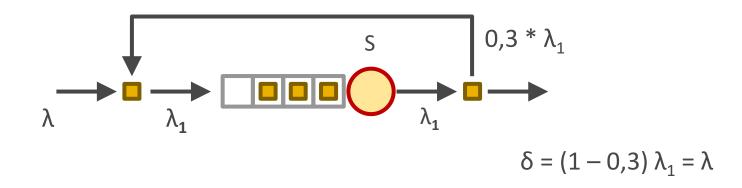
Primjer 7: Komunikacijski kanal s pogreškom (1)

• Paketi dolaze u komunikacijski kanal s intenzitetom 0,5 paketa u sekundi i zahtijevaju 0,75 sekundi za obradu. Za 30 % paketa dogodi se pogreška pri prijenosu i takvi paketi se umeću u rep za ponovno slanje.

Koliko vremena paket prosječno provede u kanalu?



Primjer 7: Komunikacijski kanal s pogreškom (2)





Primjer 7: Komunikacijski kanal s pogreškom (3)

- Analitičko rješenje
 - Intenzitet paketa $\lambda = 0.5 \text{ p/s}$
 - Prosječno vrijeme obrade paketa S = 0,75 s/p
 - Vjerojatnost pogreške paketa pri prijenosu p = 0,3 $\lambda 1 = \lambda / (1 p) = 0,5 / 0,7 = 0,714 p/s$
 - Prosječna zaposlenost kanala $\rho = \lambda 1 \times S = 0.714 \text{ p/s} \times 0.75 \text{ s/p} = 0.536 (53.6 \%)$
 - Prosječno vrijeme čekanja paketa u repu $W = S \times \rho / (1 \rho) = 0.866 \text{ s/p}$
 - Prosječno vrijeme zadržavanja paketa (1 prolaz) T1 = W + S = 0.866 s/p + 0.75 s/p = 1.616 s/p
 - Prosječno vrijeme zadržavanja u kanalu: T = T1/(1-p) = 2,31 s



Primjer 7: Komunikacijski kanal s pogreškom (4)

```
main() {
 extern int
              nodes, streams;
 float p_{err} = 0,30;
 float L = 0,50;
 float S = 0,75;
 float V = 1.0 / (1.0 - p err);
 PDQ Init("Posluzitelj s repom i povratnom vezom");
 nodes = PDQ CreateNode("Kanal", CEN, FCFS);
 streams = PDQ CreateOpen("Poruka", L);
 PDQ SetVisits("Kanal", "Poruka", V, S);
 PDQ Solve(CANON);
 PDQ Report();
```



Sadržaj predavanja

- Modeliranje i analiza raspodijeljenih sustava mrežom repova
 - Jednoposlužiteljski sustav
 - Littleov zakon
 - Serijski i paralelni poslužitelji
 - Poslužitelj s povratnom vezom
- Alat PDQ (Pretty Damn Quick)
 - Primjeri uporabe alata PDQ
 - Primjer analize performanci web-aplikacije
- Domaća zadaća



Primjer analize raspodijeljene aplikacije

Logička arhitektura raspodijeljene aplikacije

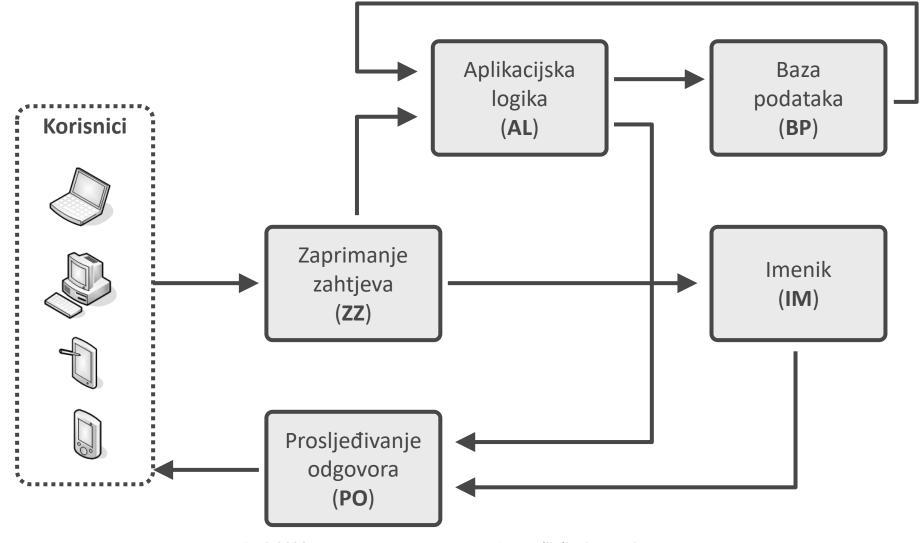
Fizička arhitektura raspodijeljene aplikacije

Model raspodijeljene aplikacije

Vrednovanje značajki performanci aplikacije



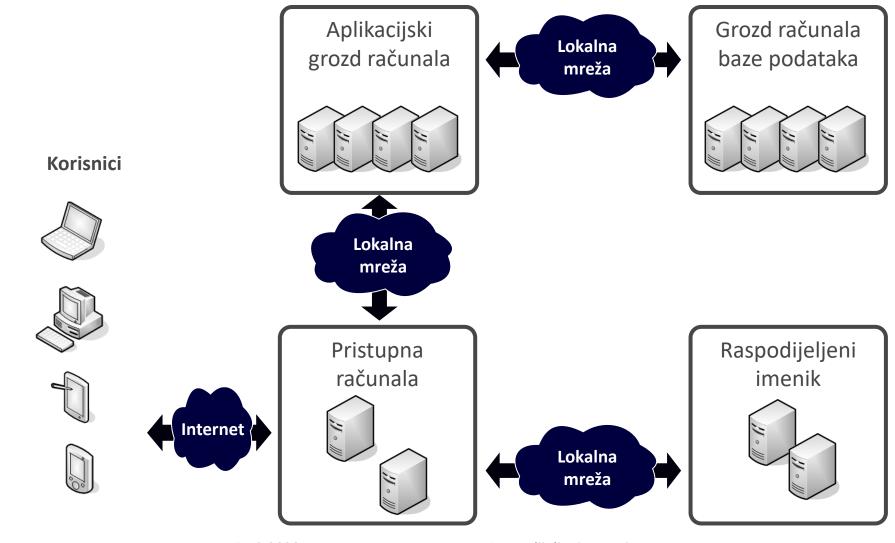
Logička arhitektura aplikacije





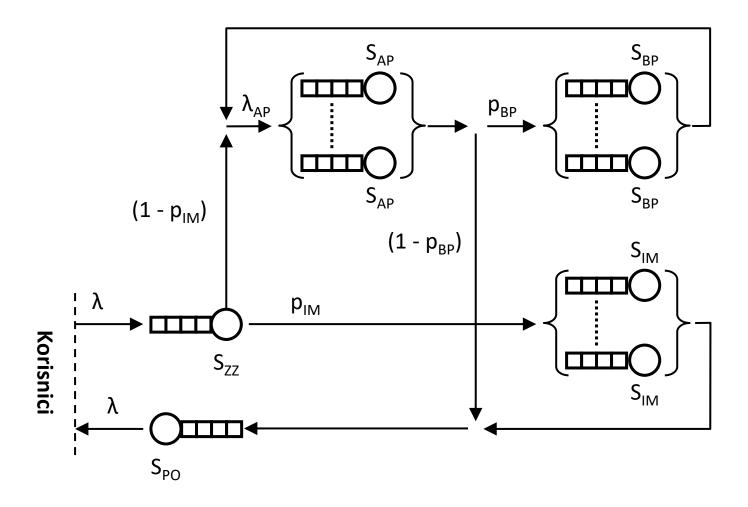
15.12.2020. Raspodijeljeni sustavi

Fizička arhitektura aplikacije





Model aplikacije





Primjer analize raspodijeljene aplikacije (1)

• Intenzitet dolazaka zahtjeva na podsustave IM, AP, BP, ZZ i PO:

$$\begin{split} \lambda_{\text{IM}} &= p_{\text{IM}} \lambda, \, v_{\text{IM}} = p_{\text{IM}} \\ \lambda_{\text{AP}} &= p_{\text{BP}} \lambda_{\text{AP}} + (1 - p_{\text{IM}}) \lambda \Longrightarrow \\ \lambda_{\text{AP}} &= \left[(1 - p_{\text{IM}}) / (1 - p_{\text{BP}}) \right] \lambda \\ v_{\text{AP}} &= \left[(1 - p_{\text{IM}}) / (1 - p_{\text{BP}}) \right] \end{split}$$

$$\lambda_{BP} = p_{BP} \lambda_{AP} \Longrightarrow$$

$$\lambda_{BP} = p_{BP} \left[(1 - p_{IM})/(1 - p_{BP}) \right] \lambda$$

$$\nu_{BP} = p_{BP} \left[(1 - p_{IM})/(1 - p_{BP}) \right]$$

$$\lambda_{ZZ} = \lambda, v_{ZZ} = 1$$

 $\lambda_{PO} = \lambda, v_{PO} = 1$



Primjer analize raspodijeljene aplikacije (2)

Skalirano vrijeme posluživanja:

$$\begin{split} &D_{IM} = v_{IM} \, S_{IM} = p_{IM} \, S_{IM} \\ &D_{AP} = v_{AP} \, S_{AP} = [(1 - p_{IM})/(1 - p_{BP})] \, S_{AP} \\ &D_{BP} = v_{BP} \, S_{BP} = p_{BP} \, [(1 - p_{IM})/(1 - p_{BP})] \, S_{BP} \\ &D_{ZZ} = v_{ZZ} \, S_{ZZ} = 1 \, S_{ZZ} \\ &D_{PO} = v_{PO} \, S_{PO} = 1 \, S_{PO} \end{split}$$

Vrijeme zadržavanja zahtjeva:

$$\begin{array}{ll} T = & v_{IM} \, T_{IM} + v_{AP} \, T_{AP} + v_{BP} \, T_{BP} + v_{ZZ} \, T_{ZZ} + v_{PO} \, T_{PO} \Longrightarrow \\ T = & D_{IM} \, / (1 - \lambda \, D_{IM}) + D_{AP} \, / (1 - \lambda \, D_{AP}) + D_{BP} \, / (1 - \lambda \, D_{BP}) + \\ D_{ZZ} \, / (1 - \lambda \, D_{ZZ}) + D_{PO} \, / (1 - \lambda \, D_{PO}) \\ \text{gdje je } v_x \, T_x = v_x \, S_x \, / (1 - \rho_x) = v_x \, S_x \, / (1 - v_x \, \lambda S_x) = D_x \, / (1 - \lambda \, D_x) \end{array}$$



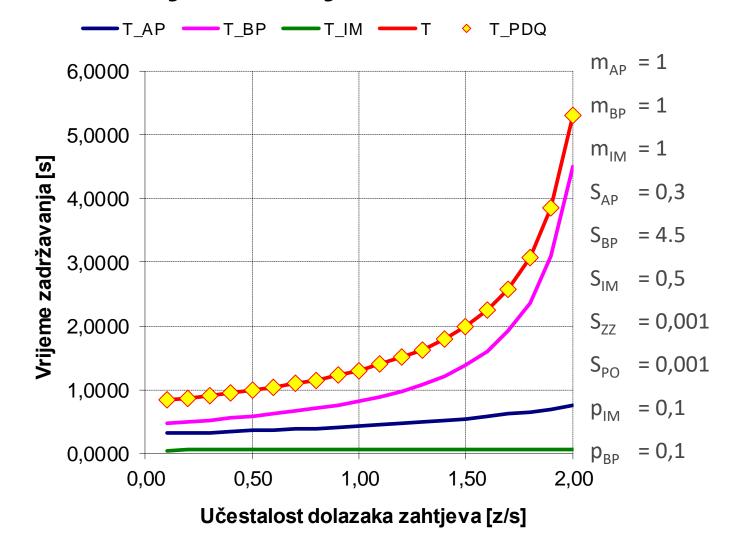
Vrijeme zadržavanja zahtjeva







Ap_Zad.xls





Vrednovanje značajki performansi

Učestalosti pristupa podacima

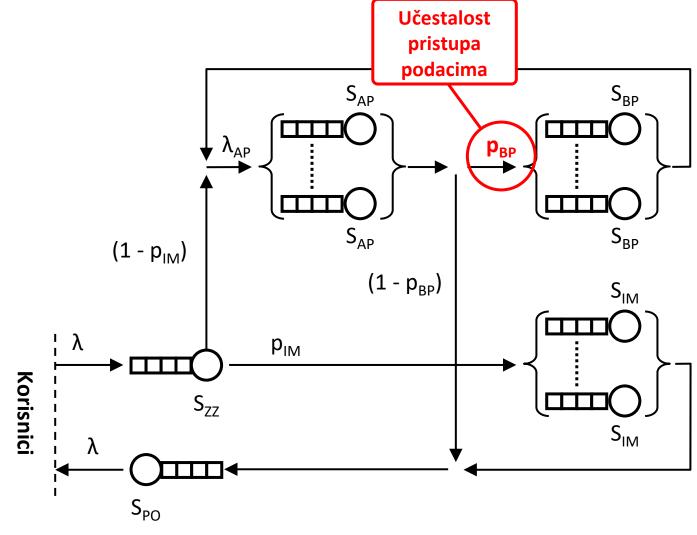
Veličina grozda baze podataka

Promjena organizacije podataka

Promjena stupnja sigurnosti



Učestalost pristupa podacima





Vrijeme zadržavanja zahtjeva

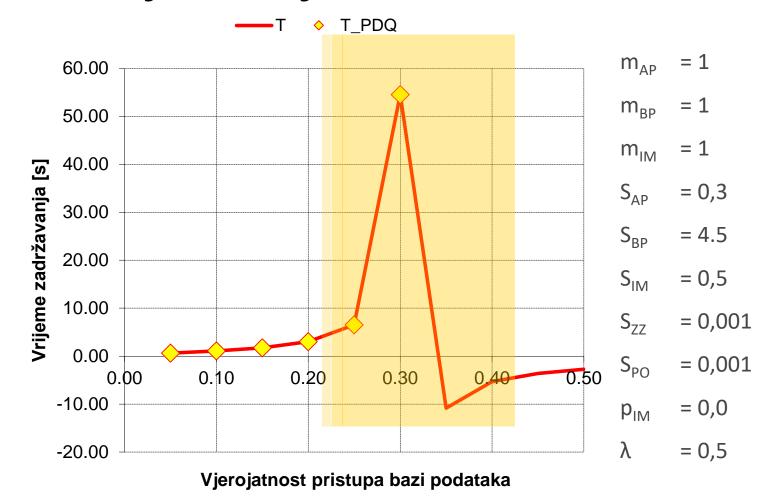


DatabaseA ccess.java



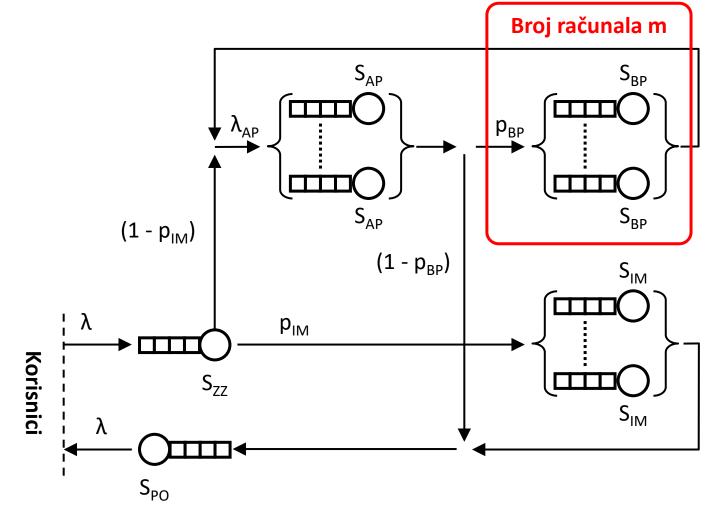


Ap_Prist.xls





Veličina grozda baze podataka





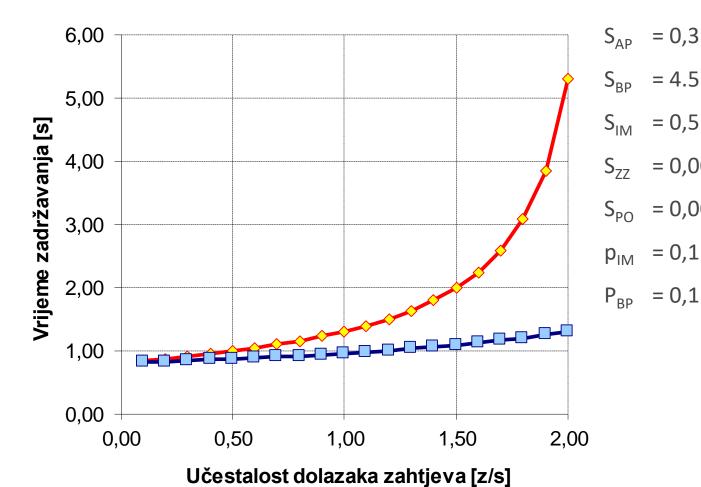
73 od 87

Vrijeme zadržavanja zahtjeva



Response. java

ap-Zad.c





= 0,3

= 4.5

= 0,5

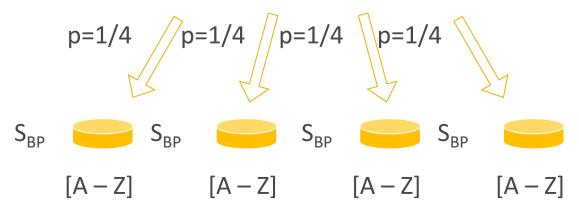
= 0,001

= 0,001

= 0,1

Promjena organizacije podataka (1)

- Replikacija podataka:
 - Skup računala od kojih svako u spremniku sadrži kopiju cijele baze podataka
 - Zahtjevi se raspoređuju na računala s ciljem uravnoteženja opterećenja
 Raspoređivanje zahtjeva

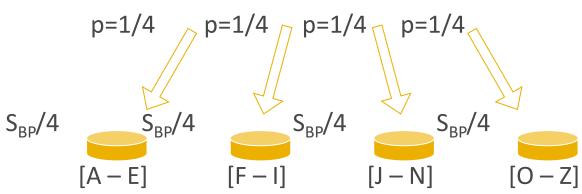




Promjena organizacije podataka (2)

- Segmentacija podataka:
 - Skup računala od kojih svako u spremniku sadrži dio cijele baze podataka
 - Zahtjevi se prosljeđuju prema računalu s traženim zapisima

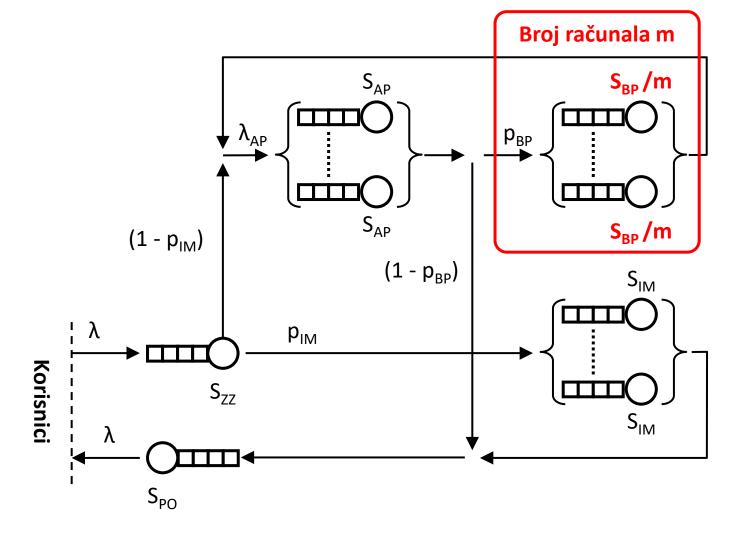
Prosljeđivanje zahtjeva



- Pretpostavke
 - Uniformna raspodjela zahtjeva na zapise
 - Linearna složenost obrade zahtjeva o količini zapisa



Promjena organizacije podataka (3)

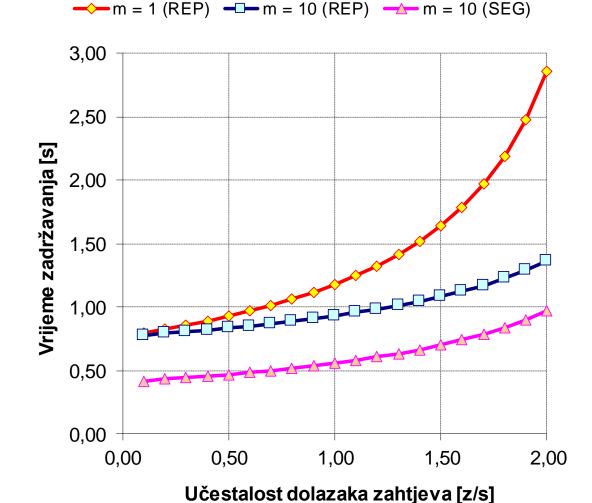




Vrijeme zadržavanja zahtjeva



Organizati on.java



$$m_{AP} = 1$$

$$m_{BP} = 1$$

$$m_{IM} = 1$$

$$S_{AB} = 0.3$$

$$S_{pp} = 2.5$$

$$S_{IM} = 0.5$$

$$S_{77} = 0.001$$

$$S_{PO} = 0.001$$

$$p_{IM} = 0.1$$

$$p_{BP} = 0.15$$





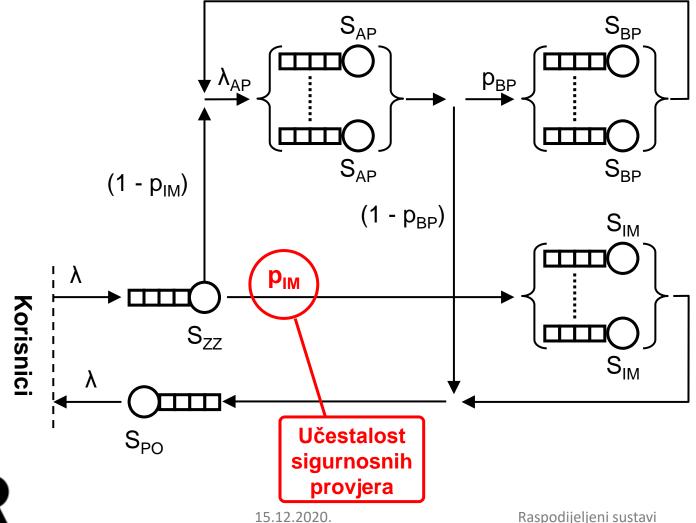
Ap_Pod.xls

Utjecaj stupnja sigurnosti

- Imenik aplikacije sadrži informacije o korisnicima
 - Korisnički identiteti
 - Korisnička prava pristupa
- Sigurnosna značka
 - Određuje sigurnosne postavke korisnika aplikacije
 - Značka se dohvaća iz imenika
- Životni vijek sigurnosne značke
 - Ograničeni broj pristupa
 - Zadano vrijeme korištenja
 - Ostali sigurnosni modeli

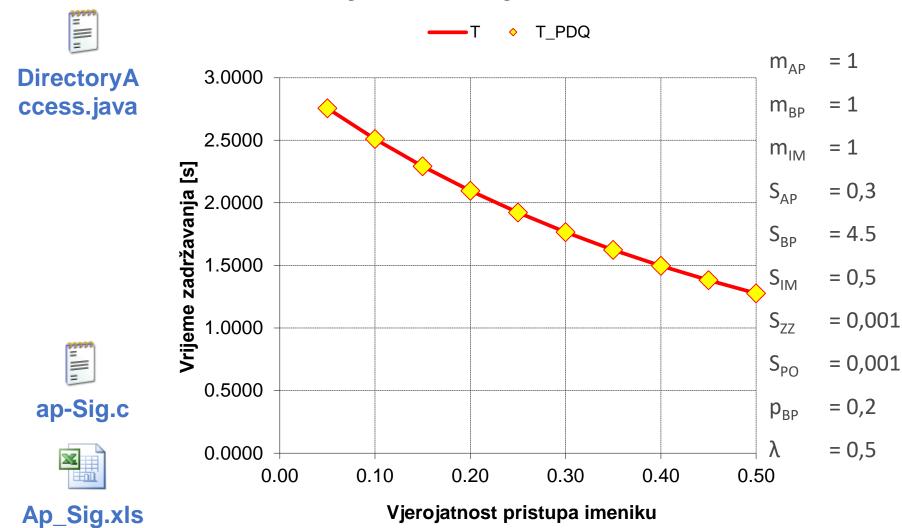


Promjena stupnja sigurnosti





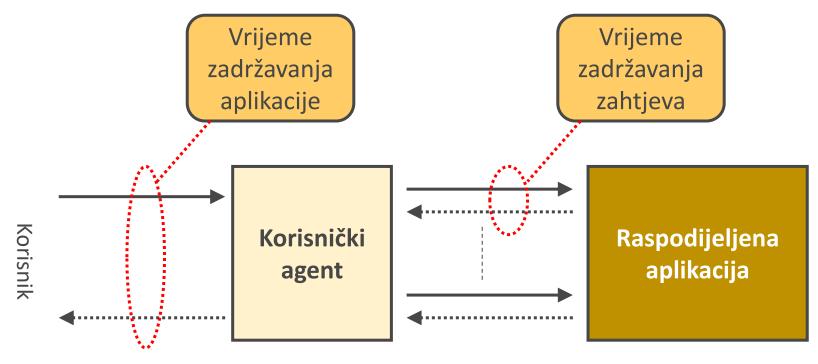
Vrijeme zadržavanja zahtjeva





Utjecaj stupnja sigurnosti

- Zašto vrijeme zadržavanja opada?
 - Modelirano je vrijeme zadržavanja zahtjeva ali ne i ukupno vrijeme zadržavanja aplikacije koje doživljava korisnik





Sadržaj predavanja

- Modeliranje i analiza raspodijeljenih sustava mrežom repova
 - Jednoposlužiteljski sustav
 - Littleov zakon
 - Serijski i paralelni poslužitelji
 - Poslužitelj s povratnom vezom
- Alat PDQ (Pretty Damn Quick)
 - Primjeri uporabe alata PDQ
 - Primjer analize performanci web-aplikacije
- Domaća zadaća



Domaća zadaća

- Zadatak: Oblikovati zadanu raspodijeljenu aplikaciju i provesti analizu performanci ostvarene aplikacije
 - 1) Definirati logičku i fizičku arhitekturu aplikacije
 - 2) Izgraditi model aplikacije primjenom teorije repova
 - Odrediti analitičko rješenje funkcije zadržavanja zahtjeva u aplikaciji $T = f(\lambda)$
 - 3) Izgraditi model aplikacije za alat PDQ
 - Primjenom izgrađenog modela odrediti vrijednosti funkcije zadržavanja zahtjeva $T = f(\lambda)$ u nekoliko točaka
 - 4) Usporediti i obrazložiti dobivene rezultate



Literatura, ...

- Sadržaj ovog predavanja nastao je na temelju:
- N. J. Gunther: "The practical performance analyst", Mcgraw Hill i Authors Choice Press, 1998 i 2000. (poglavlja 2 i 3)
- D. A. Menasce, V.A.F. Almeida: "Capacity planning for web services", Prentice Hall, 2002 (poglavlja 1 i 5)
- D. A. Menasce, V.A.F.Almeida, "Capacity Planning: An Essential Tool for Managing Web Services", IT professional, Vol. 4, No. 4, 2002., pp. 33-38."
- D. F. Vrsalovic, et. al: "Performance prediction and calibration for a class of multiprocessors", IEEE Transactions on Computers, Volume: 37 Issue: 11, Nov. 1988, pp. 1353 -1365



Rekapitulacija i pitanja za provjeru znanja (1)

- Koje su osnovne veličine u modelu repa čekanja?
 - Vrijeme promatranja (t), broj dolazaka (α), broj odlazaka (β) i vrijeme zaposlenosti poslužitelja (σ).
- Koje su izvedene veličine ?
 - Ulazni intenzitet ($\lambda = \alpha/t$), izlazni intenzitet ($\delta = \beta/t$), prosječno vrijeme posluživanja ($S = \sigma/\beta$) i zaposlenost poslužitelja ($\rho = \sigma/t$)
- Kako se definira stacionarno stanje?
 - $\delta = \lambda$
- Kako glasi Littleov zakon ?
 - Broj zahtjeva u sustavu proporcionalan je intenzitetu dolazaka zahtjeva i vremenu provedenom u sustavu (N = λ x T)
- Kako je definirano vrijeme čekanja u repu?
 - W = N x S
- Kako je definirano ukupno vrijeme provedeno u sustavu?
 - T = S + W



86 od 87

Rekapitulacija i pitanja za provjeru znanja (2)

- Kako se izračunava vrijeme odziva za serijske repove?
 - $T = (Q1 + Q2 + + Qm)/\lambda$
- Kako se izračunava vrijeme odziva za paralelne repove?
 - $T = S/(1-\rho')$
- Kako je definirana iskoristivost poslužitelja u sustavu s m poslužitelja?
 - $\rho' = \rho/m$
- Za koliko će se povećati vrijeme zadržavanja u sustavu s povratnom vezom ako se vjerojatnost vraćanja zahtjeva na ponovnu obradu poveća s 10 % na 20 %?
 - Ta = T1 / (1-0.2)
 - Tb = T1 / (1-0.3)
 - Tb / Ta = (1 0.3) / (1 0.2) = 0.7/0.6 = 1.166 (16.6 %)

