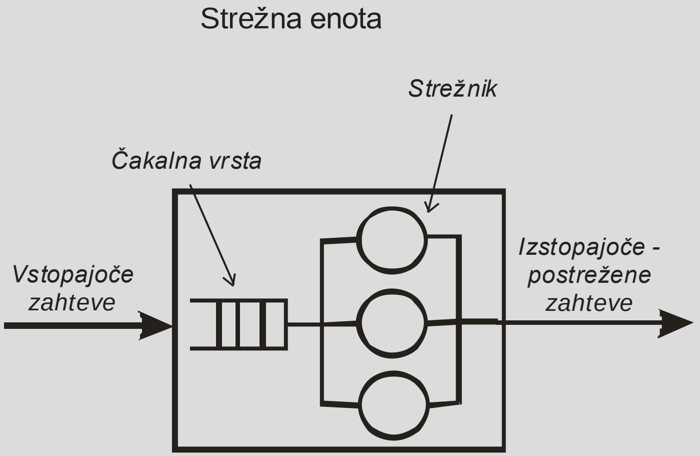
Univerza v Ljubljani
Fakulteta
za računalništvo
in informatiko

2. Osnove teorije strežbe

Vsebina 1.poglavja predavanj OMIS
(3.VSŠ/LS+PO)

2.1. Strežna enota, strežna mreža

- Strežna enota (SE):
  - 1 čakalna vrsta;
  - n paralelno vezanih strežnikov (n>=1)
  - Ponuja 1 strežbo
- Strežna mreža (SM): poljubna vezava m strežnih enot (m>1)



• Strežno enoto lahko opišemo z Kendallovo notacijo: A/B/m/K/M/Q

- A: porazdelitev medprihodnih časov zahtev
- B: porazdelitev strežnih časov zahtev
- m: število strežnikov v SE (paralelno vezanih)
- K: kapaciteta sistema, privzeto ∞ (K=dolžina vrste+m=maskimalno število zahtev v sistemu)
- M: velikost populacije zahtev (v sistemu in njegovi zunanjosti), privzeto ∞
- Q: opis strežne discipline, privzeto: FIFO

- Poziciji A in B lahko zasedeta vrednosti:
  - M – eksponentna porazdelitev
  - D – deterministična porazdelitev
  - E – Erlangova porazdelitev
  - G – Splošna verjetnostna porazdelitev
- Zgled; M/M/1//20 predstavlja SE z:
  - Esponentno porazdelitvijo časov
  - 1 strežnik
  - Velikost populacije zahtev je končna (20)

- Strežne discipline:
  - FIFO (angl. First in First out)
  - LIFO (angl. Last in First Out – skladovna strategija)
  - Prioritetna disciplina
  - Naključna disciplina
  - Disciplina razporejanja časa (angl. Time sharing)

2.2. Vhodni proces

- Medprihodni časi zahtev (konstantni, naključni ali časovno spremenljivi)
- λ - intenzivnost prihajanja zahtev (št.zahtev/časovno enoto)

2.3. Strežna enota

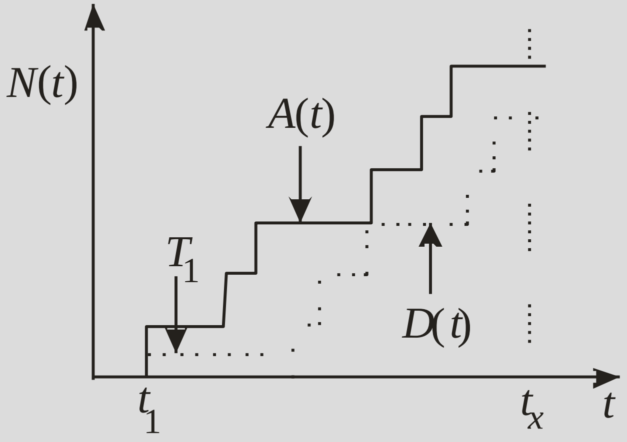
- Čakalna vrsta in paralelno vezanih n strežnikov;
- Posamezni strežnik: μ - intenzivnost strežbe (število postreženih zahtev/časovno enoto)
- n strežnikov: maksimalna intenzivnost strežbe v SE $= \mu * n$
- povprečni čas strežbe posamezne zahteve $x=1/\mu$ v posameznem strežniku

2.4. Numerične značilnosti strežnih enot

- $N(t)$ – število zahtev v SE v času t
- N – povprečno število zahtev v SE v določenem časovnem intervalu
- T – povprečni čas bivanja posamezne zahteve v SE
- W – povprečni čas zadrževanja zahteve v vrsti
- $P_k(t)$ – verjetnost k zahtev v SE v času t
- N_q, N_s, povprečno število zahtev v vrsti, povprečno število zahtev v strežbi
- $N= N_q+ N_s$
- $T=W+x$

2.4.1 Little-ov teorem

- $N=T*\lambda$
- Izpeljava na osnovi razlike stopničastih funkcij s slike na desni (N(t)=A(t)-D(t)
- N=N(t)/t
- A – število zahtev, ki so od časa 0 do t vstopile v sistem
- D – število zahtev, ki so v času od 0 do t iz sistema izstopile

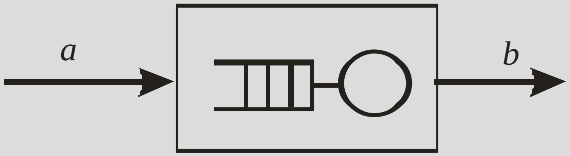


2.4.2. Uporabnostni faktor in intenzivnost prometa

- Uporabnostni faktor:
  - $N_q = \lambda * W$
  - $N_s = \lambda * x = \lambda / \mu = \rho$ (če imamo le en strežnik v SE), brez enote
  - Cilj: $0 =< \rho =< 1$
- Intenzivnost prometa:
  - $\lambda * x = \lambda / \mu = \rho$ (enota Erlang, uporaba za meritev n-kratnika preobremenitve strežbe)

2.4.3. Zakon o ohranitvi pretoka

- Predpostavka a≠b
- Prva možnost a>b vodi v eksplozijo sistema, ker ne moremo realizirati poljubno dolge vrste, zahteve pa se v sistemu vse bolj kopičijo
- Druga možnost a<b vodi v sklep, da se nekje v notranjosti SE zahteve porojevajo, česar pa SE po definiciji ne omogoča
- Sklep: predpostavka a≠b je napačna, torej a=b



2.5. Mesto izstopa zahtev – izhodni proces

- Medizhodni časi iz sistema = časi strežbe zahtev (konstantni, naključni ali časovno spremenljivi)
- Intenzivnost odhajanja zahtev (št.zahtev/časovno enoto) =< m* μ (ob m strežnikih v SE)
- Izhodni proces okarakteriziramo z M, D, E, G (glej Kendallova notacija)

2.6. Poissonov proces (PP)

- Gre za proces, v katerem so časi porazdeljeni eksponentno (M) -> prihajalni in strežni proces sta ob oznaki M Poissonova procesa
- Interpretacija PP: proces štetja naključno se porajajočih točk v časovnem intervalu (0,t)
- P(X(t)=k) verjetnost porajanja k točk na intervalu (0,t)
- $P(X(t)=k)=(\lambda * t)^k e^{-\lambda t}/k!$
- λ*t – povprečno št.pojavitev točk v časovnem intervalu

- Lastnosti PP:
  - Superpozicija: če združimo k neodvisnih PP dobimo nov PP
  - Eksp.porazdeljenost časov <-> PP
  - Verjetnost neporajanja zahteve na (0,t):
    - $P_0(t)=1-e^{-\lambda t}$
  - Čas porajanja nove točke (zahteve) je neodvisen od časa porajanja predhodne točke - zahteve

2.7. Stohastični proces (SP)

- SP je proces, v katerem je dinamika (spreminjanje stanj sistema) vsaj deloma odvisna od statistično pogojenih spremenljivk
- SP determiniran z
  - Prostorom stanj: X(t), zaloga stanj končna (diskretna stanja ali stohastična veriga) ali nekončna
  - Časovnim parametrom:
    - diskretni čas – stohastično zaporedje (X_n)
    - zvezni čas – X(t)

	Diskr.prost.stanj	Zvez.prost.stanj
Diskretni čas	Dis.čas.stoh.veriga	Dis.čas.stoh.proces
Zvezni čas	Zvez.čas.stoh.veriga	Zvez.čas.stoh.proces

2.8. Markovski proces (MP)

- MP: Novo stanje procesa je odvisno le od trenutnega stanja procesa, ne pa od njegovih predhodnikov – brez sposobnosti pomnjenja
- Eksperiment metanja kovanca:
  - Xk: izid cifra(Xi=1)/mož(Xi=0) pri k-tem metu
  - Yk: kumulativa vseh cifer do vključno k-tega meta
  - Yk=Yk-1+Xk, Y0=0
  - Xk – zaporedje naključnih spremenljivk (stohastični proces)
  - Yk – Markovska veriga (novo stanje Yk odvisno le od predhodnega stanja Yk-1)

2.8.1. Diskretne Markovske verige

- Množica diskretnih stanj (veriga) **z diskretnim časom**
- pij – verjetnost prehajanja iz stanja i v j
- Če je pij v času nespremenljiv imenujemo proces za stacionaren (časovno homogena markovska veriga)
- $X_k=j$: na koraku k smo v stanju j
- $\pi_j^{(k)}=P(X_k=j)$
- $P(X_{k+1}=j)=\Sigma P(X_k=i)\ P(X_{k+1}=j|X_k=i)=\Sigma\ \pi_j^{(k)}$ pij
- Matrični zapis verjetnosti prehajanj (prehajalna matrika M), M=(pij), matrika reda n*n glede na n možnih različnih stanj v verigi

- i-ta vrstica, j-ti stolpec v M: pij – verjetnost prehoda iz i v j
- Sum(pij)=1, j=1..n
- Za poznavanje verige moramo poleg matrike poznati tudi začetno porazdelitev verjetnosti stanj $\pi^{(0)} = (\pi_1^{(0)}, \pi_2^{(0)}, ..., \pi_n^{(0)})$ glede na n možnih različnih stanj v verigi
- V splošnem ob stacionarnosti verjetnosti velja: $\pi^{(k)} = \pi^{(0)} * M^k$

Stacionarnost v Markovski verigi (MV)– limitne verjetnosti stanj (neodvisne od začetne verjetnostne porazdelitve stanj) – stacionarna stanja:

- MV ima limitne verjetnosti, če je ergodična (aperiodičnost, nereducibilnost, časovna homogenost)
- reducibilnost: vsebuje več lot eno izolirano podmnožico stanj
- Periodičnost: vračanje v stanja z isto periodo
- Časovna homogenost: nespremenljivost verjetnosti prehajanj skozi čas

$$\sum_j \pi_j = 1, \pi_j = \sum_i \pi_i p_{i,j}$$

2.9. Zvezna časovna Markovska veriga

- Vpeljava zveznega časa (prehajanje se izvede v poljubni časovni točki z realne osi)
- Stohastičnost
- Brez pomnjenja

$$P(X(t_{k+1}) = j / X(t_1) = i_1, X(t_2) = i_2, ..., X(t_k) = i_k) =$$
$$P(X(t_{k+1}) = j / X(t_k) = i_k), t_1 < t_2 < ... < t_k < t_{k+1}$$

- $q_{i,j}$:intenzvinost prehajanja iz stanja i v stanje j
- verjetnost prehajanja iz i v j je $p_{i,j}$
- intenzivnost prehajanja iz stanja i se izraža kot
- Q matrika intenzivnosti prehajanja stanj $Q=[q_{i,j}]$

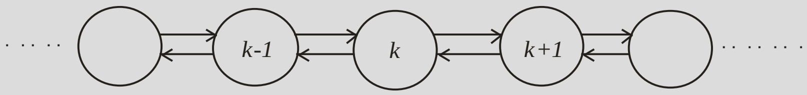
$$p_{i,j} = (t, t+\Delta t) = q_{i,j} * \Delta t$$

$$\sum_{i \neq j} q_{i,j}$$

$$q_{j,j} = -\sum_{k \neq j} q_{j,k} \Rightarrow \sum_{j} q_{i,j} = 0$$

2.10. Rojstno smrtni proces

- Posebna oblika Markovske verige, kjer so prehodi možni le v “sosednja” stanja
- “rojstno smrtni”: velikost populacije v sistemu v času t
- k zahtev -> stanje E_k , E_k -> E_{k+1} rojstvo zahteve, E_k -> E_{k-1} smrt zahteve
- Predpostavka: smrt in rojstvo se ne moreta zgoditi istočasno



$$\lambda_k = q_{k,k+1}, \mu_k = q_{k,k-1}$$

$$Q = \begin{bmatrix} -\lambda_0 & \lambda_0 & 0 & 0 & 0 \\ \mu_1 & -(\mu_1+\lambda_1) & \lambda_1 & 0 & 0 \\ 0 & \mu_2 & -(\mu_2+\lambda_2) & \lambda_2 & 0 \\ 0 & 0 & \mu_3 & -(\mu_3+\lambda_3) & \lambda_3 \\ 0 & 0 & 0 & \mu_4 & -(\mu_4+\lambda_4) \end{bmatrix}$$

- Čisti rojstno smrtni proces: $\lambda_k = \lambda > 0$, $\mu_k = 0$

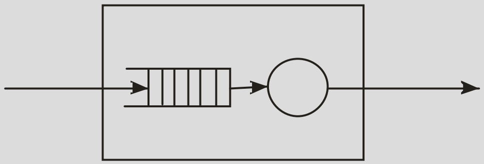
2.11. Markovski strežni sistemi z eno čakalno vrsto

- Prihajanje zahtev modeliramo s Poissonovim procesom
- Strežba zahtev z eskoponento porazdeljenimi časi
- Zanimajo nas P_k (verjetnost k zahtev v sistemu) in pa N (povprečno št.zahtev v sistemu)

$$N = \sum_k kP_k$$

2.11.1. M/M/1 sistem

- $\lambda_k = \lambda$, $\mu_k = \mu$: neodvisnost obeh intenzivnosti od trenutnega števila zahtev v sistemu – neodvisnost intenzivnosti od stanja sistema (stanje sistema ponazarja število zahtev v njem)
- privzete vrednosti: neskončna kapaciteta sistema, neskončna populacija, FIFO strežna disciplina



- Ravnotežne enačbe glede na sosede stanja k:

$$(\lambda + \mu)P_k = \lambda P_{k-1} + \mu P_{k+1}, k \geq 1$$

$$\mu P_1 = \lambda P_{0,} k = 0$$

$$\Rightarrow$$

$$P_0 = 1 - \rho,$$

$$P_k = (1-\rho)\rho^k$$

- Numerične značilke M/M/1:

$$P(N \geq n) = \rho^n$$

$$N = \frac{\lambda}{\mu - \lambda}$$

$$T = N / \lambda = \frac{1}{\mu - \lambda}$$

$$N_s = \rho = 1 - P_0$$

$$W = T - 1/\mu = \frac{\rho}{\mu - \lambda}$$

$$N_q = \lambda * W = \frac{\rho^2}{1 - \rho}$$

- Posebna vrsta M/M/1: omahljiv strežni sistem:
    - $\lambda_k = \lambda/(k+1)$ ... Več je zahtev v sistemu (k), manjša bo intenzivnost prihajanja zahtev

2.11.2 M/M/1/s sistem

- s – (končna) kapaciteta sistema: v sistemu se lahko nahaja največ (s-1) zahtev v vrsti in 1 zahteva v edinem strežniku (s-1)+1=s
- posledično je s tem diagram prehajanja stanj omejene na desni strani
- veljajo izrazi:

$$P_0 \sum_{k=0}^{s} \rho^k = 1 \Rightarrow P_0 = \frac{1-\rho}{1-\rho^{s+1}},$$

$$P_k = \frac{(1-\rho)\rho^k}{1-\rho^{s+1}}$$

- če je v sistemu s zahtev, se novoprispele zahteve izgubljajo: $\lambda = \lambda' + \lambda * p_b$
- λ - intenzivnost prihajajočih zahtev
- p_b – verjetnost polnega sistema in s tem izgube novoprispele zahteve
- λ' – intenzivnost zahtev, ki so vstopile v nezaseden sistem
- $\lambda' = \lambda * (1 - p_b)$

- veljajo izrazi:

$$P_b = \frac{(1-\rho)^s}{1-\rho^{s+1}},$$

$$N = \frac{\rho}{1-\rho} - \frac{\rho}{1-\rho}(s+1) * P_b$$

$$N_s = 1 - P_0 = \rho(1-P_b)$$

$$N_q = N - N_s$$

$$T = \frac{N}{\lambda'}$$

$$W = \frac{N_q}{\lambda'}$$

2.11.3 M/M/m sistem

- m paralelno vezanih funkcionalno in zmogljivostno ekvivalentnih strežnikov
- diagram prehajanja stanj
- strežna intenzivnost m* μ
- veljajo izrazi:

$$\rho = \frac{\lambda}{m * \mu},$$

$$P_d = \frac{P_0 * a^m}{m!(1-\rho)}, a = \frac{\lambda}{\mu},$$

$$N_q = \frac{\rho}{1-\rho} P_d = \frac{\lambda}{m * \mu - \lambda} P_d$$

$$W = N_q / \lambda$$

$$T = W + 1/\mu$$

2.11.4. Ostali sistemi

- M/M/m/m sistem
  - brez čakalne vrste
  - izgubni sistem
  - diagram prehajanja stanj
- Engsetov izgubni M/M/m/m/k sistem:
  - končnost populacije zahtev
- Round Robin sistem