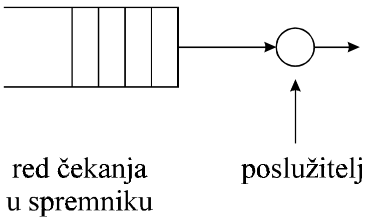
7. SUSTAVI SA POSLUŽIVANJEM

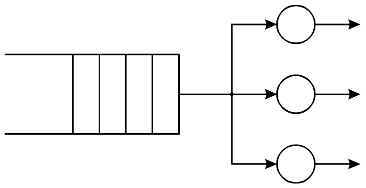
7.1 UVOD

Svaki paket kojeg se šalje kroz mrežu s prospajanjem paketa prolazi kroz više čvorišta, gdje najčešće mora čekati dok ne budu posluženi paketi prije njega. Dok čeka, paket treba zapamtiti. Paketi koji čekaju na predaju (posluživanje) pamte se u spremniku. Spremnik se formira ispred predajnika (poslužitelja), s kojim čini sustav s posluživanjem, s jednim redom čekanja i jednim poslužiteljem (Slika 7.1).



Slika 7.1. Sustav s jednim poslužiteljem i jednim redom čekanja

Postoje i sustavi sa više poslužitelja (Slika 7.2.), a mogući su i sustavi s više redova čekanja.

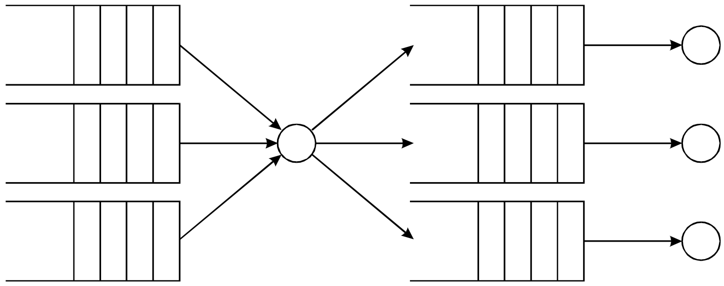


Slika 7.2. Sustav s više poslužitelja i jednim redom čekanja

Sustav može biti sa prioritetom i bez prioriteta. Kod sustava bez prioriteta prvo se poslužuje paket koji je prvi stigao u spremnik (FIFO - first in first out). Paketi koji naiđu na popunjen spremnik se tiho odbacuju. Sustav s prioritetom se primjenjuje ako neki tokovi ne dozvoljavaju kašnjenje.

Veličina spremnika je značajan parametar sustava s posluživanjem. Za slučaj prevelikog spremnika, paketi bi se nepotrebno gomilali u redu čekanja. Ta pojava je naročito izražena kod protokola koji za indikaciju zagušenja koriste gubitak PDU (npr. TCP protokol Interneta). Za slučaj spremnika malog kapaciteta, gubici paketa zbog popunjenosti spremnika bili bi česti. Gubitke bi izazivala i trenutna zagušenja. Stoga je potrebno naći optimalnu veličinu spremnika, koja će biti dovoljna za prihvaćanje praskova paketa trenutnog zagušenja, a neće pri tome izazvati nepotrebno gomilanje paketa, odnosno na vrijeme će gubitkom paketa dojaviti zagušenje izvorištu. Algoritmi kao što je RED upravljaju veličinom reda, koja ne mora biti ograničena fizičkim kapacitetom spremnika.

Usmjernik se može modelirati složenim sustavom s više redova i poslužitelja (Slika 7.3).



Slika 7.3. Model usmjernika kao sustava s posluživanjem

Paketi pristižu s ulaznih kanala i smještaju se u (kratke) ulazne spremnike, gdje čekaju na prosljeđivanje prema izlaznim kanalima. Čekanje na ulazu je nužno, jer postoji vjerojatnost istovremenog pristizanja paketa na više ulaznih kanala. Poslužitelj prospajanja uzima redom pakete iz ulaznih redova i prospaja ih prema izlaznim kanalima, koristeći se pritom tablicama usmjeravanja. Paketi se privremeno smještaju u izlazne redove čekanja (u izlazne spremnike). Čekanje na izlazu je nužno, jer postoji mogućnost da više uzastopnih paketa, pristiglih po raznim ulaznim kanalima, bude usmjereno na isti izlazni kanal.

Ukupno vrijeme čekanja T_d pojedinog paketa u sustavu sastoji se od:

- čekanja na prospajanje (zanemarivo)
- vremena prospajanja (zanemarivo)
- čekanja na predaju (T_q, ovisi o broju paketa u redu ispred promatranog)
- vremena predaje (T_s, ovisi o brzini kanala b/s i veličini paketa b)
- vremena kašnjenja na kanalu (T_p, ovisi o duljini kanala i brzini prostiranja signala)

Vrijedi:

$$T_d = T_q + T_s + T_p$$

Kod modeliranja komunikacijskih sustava sustavima s posluživanjem vrijeme prostiranja, koje je konstantno, se često zanemaruje.

7.2 STATISTIČKA SVOJSTVA DOLAZAKA I POSLUŽIVANJA

Interesira nas broj paketa u redu čekanja n(t) u nekom trenutku t. Pod pretpostavkom da je u trenutku t=0 red bio prazan, broj paketa u redu bit će razlika broja paketa koji su ušli u red l(t) i paketa koji su izašli iz reda (posluženi) m(t) do trenutka t. Pri tome iz reda ne može izaći više paketa nego što je u njega ušlo.

$$n(t) = l(t) - m(t) \quad ; \quad m(t) \leq l(t)$$

Za neki poseban slučaj možemo izmjeriti stvarne l(t) i m(t), dok za opći slučaj te funkcije moramo procijeniti. Jedna je mogućnost modeliranje procesa dolazaka paketa (zahtjeva) i predaje paketa (posluživanja) kao stohastičkih procesa. To znači da ćemo te procese prikazati njihovim statističkim svojstvima, kao što su srednja vrijednost, varijanca i funkcija razdiobe gustoće vjerojatnosti.

Proces dolazaka zahtjeva ovisi o ponašanju predajnika paketa, a to su stotine ili tisuće korisnika priključenih na neku mrežu. Uobičajena je pretpostavka da su dolasci paketa međusobno nezavisni, iako u stvarnosti postoj znatna međuovisnost paketa koji pripadaju istom toku, pa i međuovisnost tokova koje aktivira isti proces (npr. jedna sjednica dohvata WEB stranice). Proces dolazaka kvantitativno se određuje intenzitetom dolazaka λ paketa (okteta, bita) u jedinici vremena.

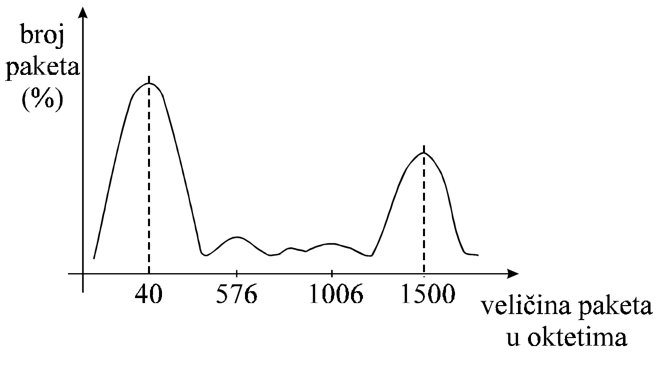
Posluživanje pojedinog paketa ovisi o brzini kanala c (bita sekundi) i duljini paketa M (bita):

$$T_s = M/c$$

Proces posluživanja ovisi o razdiobi duljina paketa na mreži. Kvantitativno ga opisujemo intenzitetom posluživanja paketa srednje duljine:

$$\mu = 1/T_s = c/\overline{M}$$

Mjerenja su pokazala da se Internetom prenosi velik broj kratkih paketa duljine oko 40 okteta, koji su rezultat praznih TCP potvrda i Telnet prometa. Ovi paketi ne opterećuju znatno kapacitet kanala, ali predstavljaju osjetno opterećenje za usmjernike. Nešto manji je broj dugačkih paketa duljine 1500 okteta, kojima se prenosi gotovo 90% ukupnog prometa. Ti paketi su rezultat masovne primjene Ethernet lokalnih mreža, za koje je maksimalna veličina SDU 1500 (Ethernet II) i 1492 (SNAP) okteta. Nešto je paketa duljine 1006 okteta koje generiraju računala priključena posredstvom SLIP protokola, a ima paketa duljine 576 okteta, što je garantirana duljina koja će Internetom biti prenesena bez fragmentacije na mrežnoj razini (Slika 7.4).



Slika 7.4. Broj paketa u ovisnosti o njegovoj veličini

7.3 KENDALOVA NOTACIJA

Očito je da ni dolazni proces ni proces posluživanja nije lako modelirati. Stoga se koriste različiti stohastički modeli, a sustavi s posluživanjem se označavaju Kendalovom notacijom u šest parametara:

A/B/C/k/n/z

z - disciplina (ako nije naglašeno, podrazumijeva se da je FIFO);
n - broj izvorišta (ako nije naglašeno, podrazumijeva se da je ∞);
k -kapacitet spremnika (ako nije naglašeno, podrazumijeva se da je ∞);
C - broj poslužitelja (ako nije specificirano, pretpostavlja se da je 1);
B - statistički izražena raspodijela vremena između 2 poslužitelja.
A - statistički izražena raspodijela vremena između 2 dolaska zahtijeva;

Ako je k= ∞ , n= ∞ , a z=FIFO, Kendelova notacija se piše kao A/B/C; inače je A/B/C/k/n/z. Umjesto A/B simbola u notaciji možemo pisati oznake razdioba :

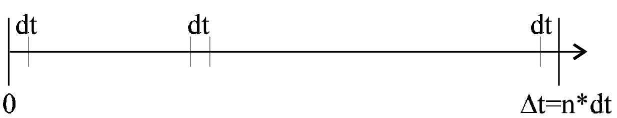
- M - eksponencijalna razdioba (Markovljeva razdioba), gdje je funkcija gustoće vjerojatnosti proporcionalna sa $f(t) = e^{-\mu \cdot t}$. Analitički se rezultati dobiju jednostavno.
- E_r - Erlangova razdioba stupnja r : $b(x) = \frac{r \cdot \mu}{(r-1)!}(r \cdot \mu \cdot x)^{r-1} e^{-r \cdot \mu \cdot x}$.

  Za r=1 imamo eksponencijalnu razdiobu;
- H_r - hiper eksponencijalna razdioba r-tog stupnja : $b(x) = \sum_{i=1}^{R} \alpha_i \cdot \mu_i \cdot e^{-\mu_i \cdot x}$;
- D - deterministička raspodijela : $b(x) = u_0\left(x - \frac{1}{\mu}\right)$;
- G - generalna razdioba. b(x) je proizvoljan, modelira se Gaussovom (normalnom) razdiobom: $b(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x-m)^2}{2\sigma^2}}$. Dobra je za simulacije, dok je dobivanje rezultata analitičkim putem otežano;
- P - Paretova razdioba. Primjenjuje se u situacijama kada nemamo pakete srednje veličine, a imamo mnogo malih i velikih paketa. Još se naziva i "razdioba s debelim repom".

7.4 EKSPONENCIJALNA RAZDIOBA

Eksponencijalna se raspodjela vrlo često koristi za modeliranje dolaznog procesa i procesa posluživanja. Razlog tome je jednostavnost u postizanju analitičkih rezultata, te činjenica da eksponencijalna razdioba može poslužiti kao najgori slučaj.

Eksponencijalna razdioba je nedjeljivo povezana s Poissonovim procesima. To su procesi kod kojih događaje (pakete) generira beskonačna populacija predajnika, tako da su oni međusobno potpuno nezavisni. Promatrajmo period Δt, i vjerojatnost $p_k(\Delta t)$ da se je u tom periodu desilo upravo k događaja. Podijelimo Δt na n po volji kratkih perioda dt ($\Delta t = n \cdot dt$), tako da je vjerojatnost pojave dvostrukog ili višestrukog događaja u periodu dt zanemariva (Slika 7.5)



Slika 7.5. Računanje vjerojatnosti k događaja za Poissonov proces

Vjerojatnost k događaja proporcionalna je broju kombinacija s kojima se desi upravo k događaja u n perioda dt, a u odnosu na sve moguće brojeve događaja. Pustimo li da $n \to \infty$ dobijemo:

$$p_k(\Delta t) = \frac{(\lambda \Delta t)^k}{k!} e^{-\lambda \cdot \Delta t}$$

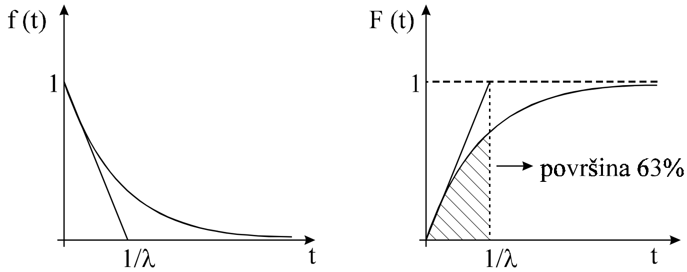
Nas zapravo interesira razdioba gustoće vjerojatnosti vremena između dva dolaska paketa, a to je upravo vjerojatnost da u promatranom periodu Δt neće stići niti jedan zahtjev:

$$p_0(\Delta t) = e^{-\lambda \cdot \Delta t}$$

Pri tome je srednje vrijeme između dva dolaska obrnuto proporcionalno intenzitetu dolazaka:

$$T_d = 1/\lambda$$

Eksponencijalna razdioba prikazana je na Slici 7.6.

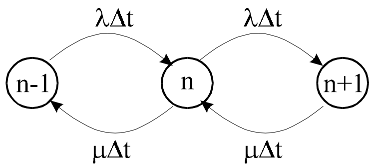


Slika 7.6. Eksponencijalna razdioba

Karakteristika je eksponencijalne razdiobe, da ako uzmemo dovoljno mali period Δt, vjerojatnost jednog dolaska (paketa) prelazi u $\lambda\Delta t$, dok je vjerojatnost višestrukih dolazaka zanemariva.

Istim postupkom dobijemo da vjerojatnost jednog posluživanja (paketa) prelazi u $\mu\Delta t$, dok je vjerojatnost višestrukih posluživanja zanemariva.

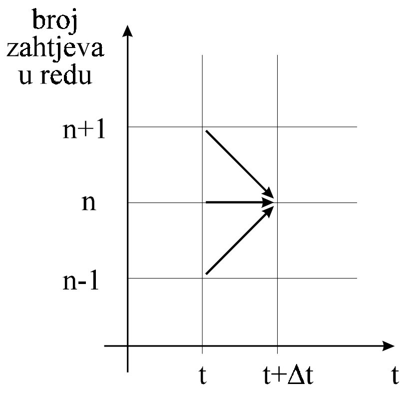
Pod gornjim uvjetima, procesi dolazaka i posluživanja prelaze u procese rađanja i umiranja i nazivaju se Markovljevi procesi. Karakteristika je Markovljevih procesa (lanaca) da sustav iz stanja n može prijeći samo u stanje n-1 ili u stanje n+1, odnosno u stanje n doći samo iz tih stanja (Slika 7.7).



Slika 7.7. Markovljev lanac: proces rađanja i umiranja

7.5 SUSTAV M/M/1

Broj paketa u spremniku sustava s posluživanjem možemo modelirati sustavom rađanja i umiranja, a sustav označavamo kao M/M/1. Pri tome se podrazumijeva da je kapacitet spremnika beskonačan, da je populacija predajnika beskonačna, te da je disciplina posluživanja FIFO. Vremensko - prostorni dijagram za Markovljev lanac prikazan je na Slici 7.8.



Slika 7.8. Vremensko - prostorni dijagram za Markovljev lanac

Pokušajmo izračunati vjerojatnost da se u trenutku $t + \Delta t$ u redu čekanja nađe upravo n paketa. Ta vjerojatnost jednaka je vjerojatnosti postizanja stanja n, a za to imamo četiri mogućnosti:

1. ako je broj zahtjeva u trenutku t jednak n, a nije se desio ni jedan dolazak niti jedno posluživanje.
2. ako je broj zahtjeva u trenutku t jednak n, i desio se jedan dolazak i jedno posluživanje.
3. ako je broj zahtjeva u trenutku t jednak n+1, i dogodilo se jedno posluživanje i nijedan dolazak.
4. ako je broj zahtjeva u trenutku t jednak n-1, i dogodio se jedan dolazak i nijedno posluživanje.

Vjerojatnost da se sustav nalazi u stanju n u trenutku $t+\Delta t$ jednaka je sumi vjerojatnosti gornja četiri moguća događaja:

$$p_n(t+\Delta t) = p_n(t)\left[(\mu\Delta t)(\lambda\Delta t) + (1-\mu\Delta t)(1-\lambda\Delta t)\right] + p_{n-1}(t)\left[\lambda\Delta t(1-\mu\Delta t)\right] + p_{n+1}(t)\left[(1-\lambda\Delta t)\mu\Delta t\right]$$

Izraz vrijedi za n >0 (broj zahtijeva ne može biti negativan).

Ako je $\Delta t <<$, za stacionarno stanje vrijedi aproksimacija

$$p_n(t+\Delta t) = p_n(t)$$

pa imamo nakon množenja:

$$p_n(t) = p_n(t)\left[\lambda\mu\Delta t^2 + 1 - \lambda\Delta t - \mu\Delta t + \lambda\mu\Delta t^2\right] + p_{n-1}(t)\left[\lambda\Delta t - \lambda\mu\Delta t^2\right] + p_{n+1}(t)\left[\mu\Delta t + \lambda\mu\Delta t^2\right]$$

Ako je $\Delta t <<$, onda je Δt^2 zanemarivo:

$$p_n(t) = p_n(t)\left[1 - \lambda\Delta t - \mu\Delta t\right] + p_{n-1}(t)\cdot\lambda\Delta t + p_{n+1}(t)\cdot\mu\Delta t \;|:\Delta t$$

Dijeljenjem s Δt dobije se jednadžba ravnoteže vjerojatnosti:

$$p_n(t)(\mu+\lambda) = p_{n-1}(t)\cdot\lambda\cdot p_{n+1}(t)\cdot\mu$$

Da bi se izračunale pojedinačne vjerojatnosti, primijenit ćemo dva granična uvijeta :

1. $\sum p_n = 1$, odakle slijedi da funkcija gustoće vjerojatnosti mora padati da bi vrijedilo $p_n \to 0$ za $n \to \infty$, odnosno $p(\infty) = 0$.
2. problem početnog stanja : Ako je $n = 0,1,\ldots,\infty$ i ako promatramo slučaj n = 0, slijedi $p_n = p_0$, pa vjerojatnost $p_{n-1} = 0$ ne postoji (u redu ne može biti negativan broj zahtjeva).

Iz drugog uvjeta imamo:

$$p_0 = p_0(1 - \lambda\Delta t - \cancel{\mu\Delta t}) + p_1\cdot\mu\Delta t\,.$$

Ako smo imali n=0 zahtijeva, ne možemo izvršiti posluživanje, te dobijemo :

$$p_0\lambda\Delta t = p_1\mu\Delta t \quad\Rightarrow\quad p_1 = \frac{\lambda}{\mu}p_0\,.$$

λ/μ je odnos između inteziteta (srednje vrijdnosti) dolazaka i inteziteta odlazaka (posluživanja). Predstavlja faktor opterećenja sustava:

$$\rho = \frac{\lambda}{\mu}$$

Može se pisati:

$$p_n\cdot(\mu+\lambda) = p_{n-1}\cdot\lambda + p_{n+1}\cdot\mu \;|:\mu$$

$$p_n\cdot(1-\rho) = p_{n-1}\cdot\rho + p_{n+1}$$

Slijedi da je:

$$p_1 = \rho\cdot p_0$$

$$p_2 = (\rho+1)\cdot p_1 - \rho\cdot p_0 = (\rho+1)\cdot\rho\cdot p_0 - \rho\cdot p_0 = \rho^2\cdot p_0$$

i konačno:

$$p_n = \rho^n\cdot p_0$$

Korištenjem prvog graničnog uvjeta za $\rho < 1$ imamo:

$$\sum_{n=0}^{\infty} p_n = \sum_{n=0}^{\infty} \rho^n \cdot p_0 = \frac{p_0}{1-\rho}$$

Na kraju dobijemo:

$$\frac{p_0}{1-\rho} = 1 \quad \Rightarrow \quad p_0 = 1-\rho$$

Vjerojatnost da u spremniku imamo n poruka je jednaka:

$$p_n = (1-\rho) \cdot \rho^n$$

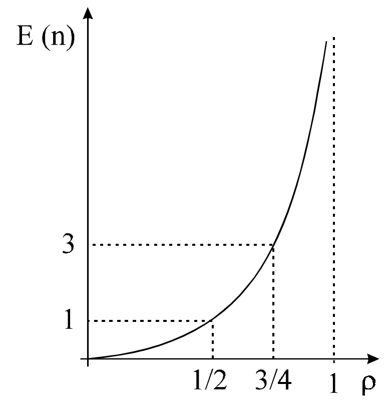
Sada se može izračunati prosječna duljina reda čekanja kao očekivanje od n, E(n). Očekivanje je po definiciji:

$$E(n) = \sum_{n=0}^{\infty} n \cdot p_n = \sum_{n=0}^{\infty} n \cdot (1-\rho) \cdot \rho^n$$

Za $\rho < 1$ se dobije:

$$E(n) = \frac{\rho}{1-\rho}$$

Prosječna duljina reda čekanja E(n) je funkcija opterećenja sustava ρ. Red je približno prazan kada je $\rho = 0$ (mreža podopterećena), a teži u ∞ kada je $\rho = 1$ (mreža preopterećena, Slika 7.9).



Slika 7.9. Prikaz očekivanja E(n) u funkciji faktora opterećenja ρ

Gledajući kratkoročno, ρ može biti veći od q, jer može stići više zahtjeva nego što će ih biti posluženo, i ti će zahtjevi biti privremeno pohranjeni u spremnik. Dugoročno mora biti $\rho \leq 1$.

Prosječno očekivano vrijeme čekanja je jednako nekom očekivanom vremenu posluživanja i prosječnom vremenu čekanja tog istog zahtjeva na posluživanje prethodnih paketa:

$$E(T) = \frac{1}{\mu} + \frac{E(n)}{\mu} = \frac{1}{\mu}(1 + E(n)) = \frac{1}{\mu}\left(1 + \frac{\rho}{1-\rho}\right) = \frac{1}{\mu} \cdot \frac{1}{1-\rho}$$

$$\rho\mu = \frac{\lambda}{\mu} \cdot \mu = \lambda$$

$$E(T) = \frac{1}{\mu - \lambda}$$

Ako istu jednadžbu pomnožimo sa μ, dobit ćemo Littleovu formulu:

$$E(T)=\frac{1}{\mu}+\frac{E(n)}{\mu} \qquad /\cdot\mu$$

$$E(n)=\mu E(T)-1=\frac{\mu}{\mu-\lambda}-\frac{\mu-\lambda}{\mu-\lambda}=\frac{\lambda}{\mu-\lambda}=\lambda E(T)$$

Littleova formula

$$E(n)=\lambda E(T)$$

u sustavima sa posluživanjem daje vezu između dužine reda čekanja i dužine čekanja, preko intenziteta prometa. Znači, iz dvije izmjerene veličine možemo izračunati treću. Važnost je ove formule, da vrijedi za sve razdiobe.

7.6 SUSTAV M/M/1/N

U realnim sustavima spremnike je ipak konačnog kapaciteta, te je vjerojatnost da u spremniku imamo određen broj poruka iznad predviđenog kapaciteta =0. Takav se sustav naziva sustav s gubicima, jer se zahtjevi koji su došli, a nisu imali mjesta u spremniku, smatraju izgubljenima.

Za konačni kapacitet spremnika $N < \infty$ vrijedi:

$$\sum_{n=0}^{\infty} p_n = p_0 \sum_{n=0}^{\infty} \rho^n = p_0 \frac{1-\rho^{N+1}}{1-\rho} = 1$$

$$p_0 = \frac{1-\rho}{1-\rho^{N+1}}$$

Korištenjem jednadžbe ravnoteže za beskonačni spremnik dobije se:

$$p_n = \frac{1-\rho}{1-\rho^{N+1}} \cdot \rho^n$$

Vjerojatnost punog spremnika je:

$$p_N = \frac{(1-\rho)\cdot\rho^N}{1-\rho^{N+1}}$$

Vjerojatnost da dođe do gubitaka p_B je jednaka vjerojatnosti da je red pun (n=N) i da u tom trenutku dođe jedan zahtjev, a ne izvrši se nijedno posluživanje:

$$p_B = p_N\,(\lambda\,\Delta t)(1-\mu c\,\Delta t)$$

Ovu vjerojatnost izračunamo iz jednadžbe ravnoteže za cijeli sustav koja kaže da je vjerojatnost neblokiranog sustava u trenutku dolaska paketa jednaka vjerojatnosti posluživanja u trenutku kada red nije prazan:

$$\lambda(1-p_B)=\mu\,(1-p_0)$$

Vjerojatnost gubitka je:

$$p_B = \frac{\lambda^N(\mu-\lambda)}{\mu^{N+1}-\lambda^{N+1}}$$

7.7 SUSTAV M/G/1

Ako sustav M/M/1 ne modelira dovoljno dobro izlazni proces, koristi se M/G/1. Kod M/G/1 je proces posluživanja opisan srednjom vrijednošću i varijancom, a korištenjem Gaussove razdiobe dobije se:

$$E(n)=\frac{1}{1-\rho}\left\{\rho-\frac{1}{2}\rho^2\left[1-\mu^2\sigma^2\right]\right\}$$

Primjenom Littleove formule dobijemo vrijeme čekanja, tzv. Pollaczek-Khintchine formulu:

$$E(T) = \frac{1}{\lambda} E(n) = \frac{1}{2\mu(1-\rho)} \left\{ 2 - \rho \left[1 - \mu^2 \sigma^2 \right] \right\}$$

Razdioba M/M/1 je poseban slučaj razdiobe M/G/1, a dobije se uz σ^2:

$$\sigma^2 = \frac{1}{\mu^2}$$

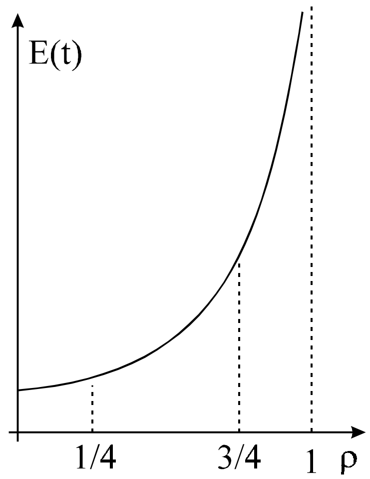
7.8 SUSTAV M/D/1

Ako imamo sustav kod kojeg je duljina paketa konstantna (ATM), konstantno je i vrijeme posluživanja, pa je $\sigma^2 = 0$:

$$E(n) = \frac{\rho}{1-\rho} \left[1 - \frac{1}{2} \rho \right]$$

7.9 PROCJENA RADA MREŽE SUSTAVOM M/M/1

Modelira li se čitava mreža sustavom s posluživanjem koji raspolaže jednim redom čekanja i jednim poslužiteljem, moguće je dobiti relativno grub, ali koristan uvid u rad mreže. Pri tome koristimo M/M/1 sustav obzirom da je on zbog visoke varijance približan najgorem slučaju. Pogledajmo krivulju kašnjenja ovisno o opterećenju ρ (Slika 7.10).



Slika 7.10. Kašnjenje ovisno o opterećenju mreže

Prihvatimo li kašnjenje kao kriterij kakvoće usluge, možemo uočiti faze rada mreže:

1. slabo opterećena mreža (0 - 25%) – može zadovoljiti svakog korisnika, mreža kod puštanja u pogon treba biti slabo opterećena kako bi mogla zadovoljiti potrebe korisnika u prihvatljivom periodu korištenja (1-3 godine);
2. srednje opterećena mreža (25% - 75%) – kompromis između kvalitete usluge i iskorištenja kapaciteta. Svako preopterećenje sustava vodi u područje velikog opterećenja mreže. Kod opterećenja iznad 50% treba započeti planiranje i izgradnju proširenja mreže;
3. vrlo opterećena mreža (75% - 100%) – dolazi do smanjenja kvalitete usluge i znatnih gubitaka. Planirano proširenje treba biti spremno za puštanje u pogon, nakon čega se nastavlja ciklus od koraka 1.