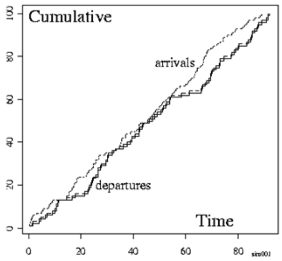
Modeliranje prometa v paketnih omrežjih

as.mag. Iztok HUMAR
prof.dr. Janez BEŠTER

Agenda

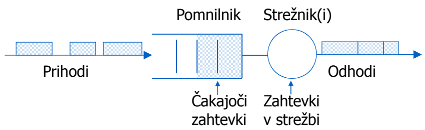
- Model splošnega strežnega sistema
- Parametri splošnega strežnega sistema
- Označevanje splošnih strežnih sistemov
- Strežni sistemi M/M/1
  - Čas med prihodi
  - Število prihodov v intervalu
  - Strežba
- Sistemska stanja
  - Verjetnosti
  - Prihodi med stanji
- Parametri zmogljivosti sistema
- Littleov Teorem

Strežni sistem



- Kumulativna razporeditev prihodov in strežbe
  - razliko med krivuljama predstavlja število elementov v vrsti
- Od sistema pričakujemo, da streže zahtevam v končnem času:
  - krivulji ne smeta divergirati druga od druge

Model strežnega sistema s čakalno vrsto



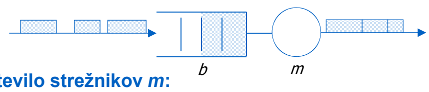
- Strežni sistem vsebuje enega ali več strežnikov
- Zahtevki prihajajo po določeno storitev
- Zahtevki, ki pridejo v sistem in ne prejmejo storitve takoj, so razporejeni v čakalno vrsto (pomnilnik)

Razlogi za zakasnitev v omrežju

- Čas procesiranja
  - Privzamemo lahko, da procesorska moč ne predstavlja omejitve
- Čas čakanja v vrsti
  - Čas čakanja v vrsti pred oddajo
- Čas prenosa po povezavi
  - Čas, ki je porabljen za prenos preko povezave (prenos električnega, optičnega signala)
  - Neodvisen od prometa, ki obremenjuje omrežje

- Največji po iznosu: čas čakanja v vrsti

Parametri strežnega sistema



- Število strežnikov *m*:
  - Eden
  - Več
  - Neskončno
- Velikost pomnilnika *b*
- Razvrščanje zahtevkov v vrsto (scheduling):
  - FCFS = FIFO,
  - LCFS = LIFO,
  - Processor Sharing (PS) ...
- Proces prihodov
- Proces strežbe

Kendall-Lee označevanje strežnih sistemov

- Omogoča enostaven opis strežnega sistema s čakalno vrsto

A/B/c:(XXXX/m/n)

- *A* določa proces prihodov
- *B* določa proces strežbe
  - M: porazdelitev Markovskega (eksponentna porazdelitev časa med prihodi)
  - D: deterministična (degenerate/deterministic) porazdelitev
  - E_k: Erlangova porazdelitev s parametrom k
  - G: splošna porazdelitev
- *c* določa število strežnikov/kanalov

Kendall-Lee označevanje strežnih sistemov

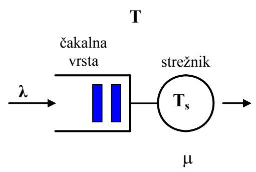
A/B/c:(XXXX/m/n)

- XXXX določa način strežbe
  - FCFS=FIFO, LCFS=LIFO, SIRO = Service In Random Order, prioritete
- *m* določa največje možno število uporabnikov v sistemu – skupaj v pomnilniku in prejemajočih storitev. Vsi ostali prihodi so zavrnjeni
- *n* določa največje število populacije, iz katere izvirajo uporabniki. To posredno omejuje hitrost prihodov.
- Pogosto se uporablja samo prva trojica oznake. V tem primeru se za drugi del privzame FIFO/∞/∞

Primeri označevanja

- M/M/1: Naključni prihodi in strežba, časi med prihodi in strežbo porazdeljeni po eksponentni porazdelitvi, en strežnik, neskončen pomnilnik in neskončna populacija, FIFO način razvrščanja
- M/M/c: Enaka porazdelitev kot pri prejšnjem, le da je v tem primeru c strežnikov
- M/G/3: Eksponentna porazdelitev časa prihodov, splošna porazdelitev časa strežbe, trije strežniki, neskončen pomnilnik in neskončna populacija, FIFO način razvrščanja
- */D/∞ : Sistem s konstantno zakasnitvijo
- M/M/c:FIFO/c: Eksponentna porazdelitev časa prihodov, splošna porazdelitev časa strežbe, c strežnikov, pomnilnika ni, neskončna populacija, FIFO način razvrščanja

Strežni sistem M/M/1

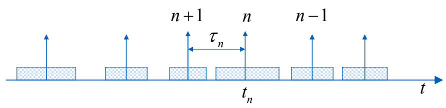


- En sam strežnik
- Neskončen pomnilnik
- Proces prihodov: Poissonov s številom zahtevkov λ
- Proces strežbe: eksponenten s parametrom $\mu = \frac{1}{T_S}$
- Stabilno stanje: λ < μ
- Časi strežbe in intervali prihodov so med seboj neodvisni

Naključni prihodi

- λ predstavlja povprečno število zahtevkov (arrival rate), ki vstopijo v sistem v časovni enoti.
- Verjetnost prihoda novega paketa v kratkem intervalu *h* je λh = konstanta.
- Verjetnost novega prihoda ni odvisna od časa zadnjega prihoda.
- Tovrstni proces prihodov imenujemo *Poissonov proces*.
- Pogoji za uporabo:
  - Število zahtev je veliko
  - Vpliv posamezne zahteve na celoten sistem je majhen
  - Vse zahteve so med seboj neodvisne t.j.: njihova odločitev za uporabo sistema je neodvisna od ostalih zahtev

Čas med prihodi

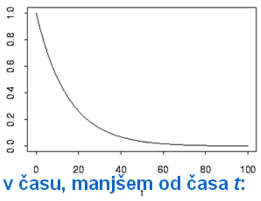


- τ_n čas med zahtevkoma n in n+1.
- τ_n je naključni čas.
- $\{\tau_n,\ n \geq 1\}$ predstavlja stohastični proces.
- Časi med zahtevki so enakomerno porazdeljeni in imajo skupno povprečno vrednost.

Eksponentna porazd. časa med prihodi

- Čas med prihodi τ_n predstavlja naključno spremenljivko z eksponentno porazdelitvijo s parametrom λ.
- Naključna spremenljivka τ_n ima eksponentno porazdelitev s parametrom λ, če je verjetnostna gostota funkcije podana kot:

$$f_{\tau_n}(t) = \begin{cases} \lambda e^{-\lambda t} & \text{if } t \geq 0 \\ 0 & \text{if } t < 0 \end{cases}$$



- Verjetnost, da pride do prihoda v času, manjšem od časa *t*:

$$F_{\tau_n}(t) = P(\tau_n \leq t) = \begin{cases} 1 - e^{-\lambda t} & \text{if } t \geq 0 \\ 0 & \text{if } t < 0 \end{cases}$$

Povprečni čas med prihodi

- Povprečna vrednost in varianca:

$$E[\tau_n] = \frac{1}{\lambda}, \quad \mathrm{Var}[\tau_n] = \frac{1}{\lambda^2}$$

- Dokaz:

$$E[\tau_n] = \int_0^\infty t f_{\tau_n}(t)\,dt = \int_0^\infty t\lambda e^{-\lambda t} dt =$$

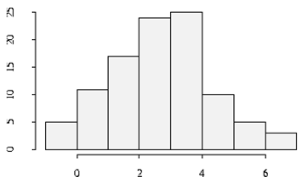
$$= -te^{-\lambda t}\Big|_0^\infty + \int_0^\infty e^{-\lambda t} dt = \frac{1}{\lambda}$$

$$E[\tau_n^2] = \int_0^\infty t^2 \lambda e^{-\lambda t} dt = -t^2 e^{-\lambda t}\Big|_0^\infty + 2\int_0^\infty te^{-\lambda t} dt = \frac{2}{\lambda} E[\tau_n] = \frac{2}{\lambda^2}$$

$$\mathrm{Var}(\tau_n) = E[\tau_n^2] - (E[\tau_n])^2 = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}$$

Število prihodov v intervalu

- V intervalu dolžine *D* je število prihodov naključna spremenljivka s Poissonovo porazdelitvijo s parametrom m=λ*D*.



- Slika prikazuje histogram porazdelitve za realen Poissonov proces.
- Pričakovano število prihodov v intervalu dolžine *D* je m=λ*D*.

Poissonova porazdelitev števila prihodov

- Proces ustreza Poissonovi porazdelitvi s parameterom *m*, če je verjetnostna funkcija podana kot:

$$P(X = x) = e^{-m}\frac{m^x}{x!}$$

- Široka uporabnost Poissonovega procesa pri modeliranju naključnih dogodkov, ki se zgodijo v danem časovnem intervalu.
  - Uporabniki, ki pridejo v trgovino tekom dneva
  - Klici na napačno številko tekom tedna
  - Študentje, ki pridejo v kabinet v času govorilnih ur
  - Paketi, ki prispejo v omrežno stikalo

Poissonova porazdelitev števila prihodov

- Srednja vrednost in varianca:

$$E[X] = m, \quad \mathrm{Var}(X) = m$$

- Dokaz:

$$E[X] = \sum_{x=0}^{\infty} xP(X = x) = e^{-m}\sum_{x=0}^{\infty} x\frac{m^x}{x!} = e^{-m}\sum_{x=0}^{\infty}\frac{m^x}{(x-1)!}$$

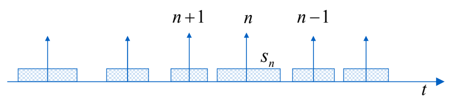
$$= e^{-m} m \sum_{j=0}^{\infty}\frac{m^j}{j!} = e^{-m} m e^m = m$$

$$E[X^2] = \sum_{x=0}^{\infty} x^2 P(X = x) = e^{-m}\sum_{x=0}^{\infty} x^2\frac{m^x}{x!} = e^{-m}\sum_{x=0}^{\infty} x\frac{m^x}{(x-1)!}$$

$$= e^{-m} m \sum_{j=0}^{\infty}(j+1)\frac{m^j}{j!} = m\sum_{j=0}^{\infty} je^{-\lambda}\frac{m^j}{j!} + me^{-m}\sum_{j=0}^{\infty}\frac{m^j}{j!} = m^2 + m$$

$$\mathrm{Var}(X) = E[X^2] - (E[X])^2 = m^2 + m - m^2 = m$$

Čas strežbe



- s_n : Čas strežbe zahtevkov n strežnika
- $\{s_n,\ n \geq 1\}$ je stohastični proces
- Časi strežbe so enakomerno porazdeljeni in imajo skupno povprečno vrednost
- Ali je za pakete v stikalu čas strežbe res naključen?

Naslov predstavitve, predavanja

Eksp. porazdelitev trajanja strežbe

- μ predstavlja število zahtev, ki jih strežnik lahko postreže v časovni enoti (service rate).
- Naključna spremenljivka s_n ima eksponentno porazdelitev s parametrom μ, če je verjetnostna gostota funkcije podatna kot:

$$f_{s_n}(t) = \begin{cases} \mu e^{-\mu t} & \text{if } t \geq 0 \\ 0 & \text{if } t < 0 \end{cases} \qquad E[s_n] = \frac{1}{\mu}, \qquad Var[s_n] = \frac{1}{\mu^2}$$

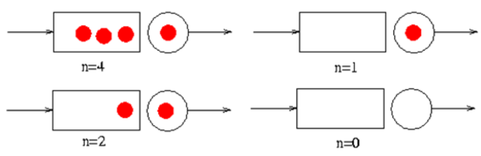
- Verjetnost, da bo čas strežbe manjši od časa *t* je:

$$F_{s_n}(t) = P(s_n \leq t) = \begin{cases} 1 - e^{-\mu t} & \text{if } t \geq 0 \\ 0 & \text{if } t < 0 \end{cases}$$

- μ predstavlja hitrost strežbe. Podaja število postreženih zahtev na časovno enoto neprestano obremenjenega strežnika.
- Verjetnost, da zahteva zapusti zaseden strežnik v intervalu *h* je μh=*konstanta.* Zahteve zapuščajo strežnik po Poissonovem procesu.

Sistemska stanja in verjetnosti

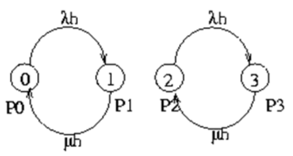
- Sistem z vrsto se nahaja v določenem stanju:
  - Stanje sistema M/M/1 določa število uporabnikov v sistemu – *n*
  - Sistem nima pomnilnika za pretekle dogodke



- Za vsako stanje obstaja verjetnost, da se sistem nahaja v njem: P_n.

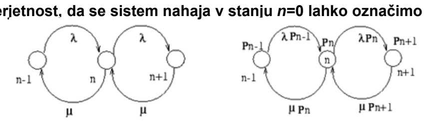
Diagram prehoda (Transition diagram)

- Pri Markovskih sistemih (sisteme s Poissonovimi prihodi in eksponentno porazdelitvijo časa strežbe) je verjetnost prehoda odvisna zgolj od trenutnega stanja.
  - Ni odvisna od tega, koliko časa se sistem nahaja v preteklem stanju.
- Za sisteme M/M/1 je v intervalu dolžine *h* konstanta:
  - Verjetnost prihoda zahtevka: λh
  - Verjetnost odhoda zahtevka - strežbe: μh



Prehod med stanji

- Delujoč Markovski sistem se premika iz stanja v stanje z določeno verjetnostjo.
- Po dolgem času se sistem ustali v stabilno stanje.
  - Verjetnost, da se sistem nahaja v stanju *n*, je konstantna (časovno neodvisna)
  - Verjetnost, da se sistem nahaja v stanju *n*=0 lahko označimo s P_0



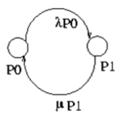
- Da zagotovimo stabilno stanje, morajo biti prehodi med vsakim parom stanj P_n in P_{n-1} med seboj uravnoteženi. Zapišemo lahko:

$$\lambda P_{n-1} = \mu P_n \rightarrow P_n = \frac{\lambda}{\mu} P_{n-1} = \rho P_{n-1}$$

Rekurzivna formula za prehod med stanji

- Zapišimo za n=1: $P_1 = \rho P_0$
- Zapišimo za splošen n>1: $P_2 = \rho P_1 = \rho^2 P_0$

$$P_n = \rho^n P_0$$

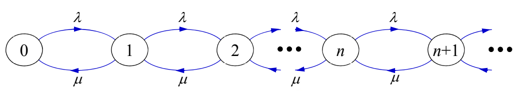


- Vsota verjetnosti je 1:

$$1 = \sum_{n=0}^{\infty} P_n = \sum_{n=0}^{\infty} \rho_n P_0 = P_0 \sum_{n=0}^{\infty} \rho_n = \frac{P_0}{1-\rho}$$

- Od tod dobimo splošno enačbo za P_n:

$$P_n = \rho^n (1-\rho)$$



Parametri zmogljivosti sistema

- Zanimajo nas parametri:
  - Povprečen čas čakanja
  - Povprečna zakasnitev pri prehodu zahteve preko sistema
  - Povprečno število čakajočih zahtev
- ρ: Izkoriščenost sistema (utilization, traffic intensity)

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

  - Ko gre λ -> μ, gre ρ -> 1
- Verjetnost, da je v sistemu (vrsti in strežbi) *n* zahtev:

$$P_n = \rho^n (1-\rho)$$

  - geometrijsko zaporedje

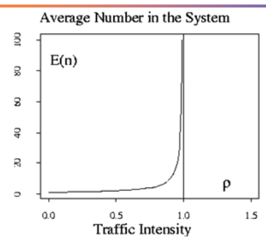
Povprečno število zahtev v sistemu

- L=E(n) predstavlja povprečje geometrijske porazdelitve:

$$L = E(n) = \sum_{n=0}^{\infty} nP_n = \sum_{n=0}^{\infty} n\rho^n (1-\rho) =$$
$$= (1-\rho)\sum_{n=0}^{\infty} n\rho^n = (1-\rho)\rho\sum_{n=0}^{\infty} n\rho^{n-1} =$$
$$= \rho(1-\rho)\frac{1}{(1-\rho)^2} = \frac{\rho}{1-\rho} = \frac{\lambda}{\mu-\lambda}$$

- Narišimo graf odvisnosti $L(\rho)$

Povprečno število zahtev v sistemu



- Povprečno število zahtev v sistemu je večje od 0, tudi če je izkoriščenost sistema (ρ) relativno nizka.
- Povprečno število zahtev v sistemu raste preko vseh mej, če gre izkoriščenost sistema (ρ) proti 1, kljub temu, da še vedno velja λ < μ.
- Pogoj za stabilnost sistema M/M/1: ρ<1

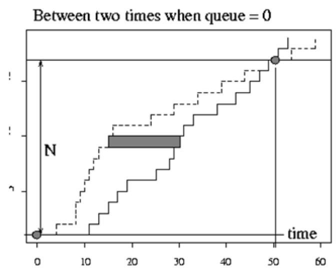
Littleov Teorem

- Littleov Teorem povezuje povprečen čas, ki ga potrebuje neko opravilo v kateremkoli sistemu z številom opravil, čakajočih na sistemu.
- Več o teoremu v:
  - Vignaux, G. A. (1997). *Analyzing Queues using cumulative graphs*. Technical report, School of Mathematical and Computing Sciences, Victoria University of Wellington.
  - Winston, W. L. (1994). *Operations Research, Applications and Algorithms*. Duxbury Press, 3rd edition.
- Izpeljava povprečnih vrednosti iz grafov kumulative prihajajočih in odhajajočih opravil.

Littleov Teorem: povprečni čas v sistemu

- Izračunajmo povprečni čas čakanja zahteve v sistemu z uporabo ploščine med presečišči krivulj.
- Povprečni čas zahteve v sistemu je W:

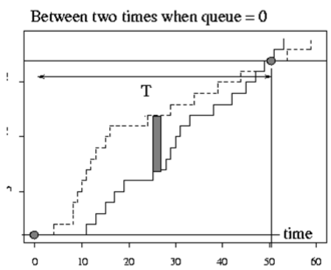
$$W = \frac{Ploščina}{N}$$



Littleov Teorem: povprečno št. v sistemu

- Izračunajmo povprečno število zahtev v sistemu z uporabo ploščine med presečišči krivulj.
- Povprečno število zahtev v sistemu je L:

$$L = \frac{Ploščina}{T}$$



Littleov Teorem

- Ploščina med presečišči krivulj je v obeh primerih enaka.

$$NW = TL \rightarrow L = \frac{N}{T}W$$

- ker je $\lambda = N/T$, dobimo: $\boxed{L = \lambda W}$ kjer je:
  - L= povprečno število zahtev na sistemu
  - λ = povprečno število prihodov zahtev
  - W = povprečni čas, ki ga zahteve porabijo na sistemu
- Videti je, da teorem velja za periodo od ene ničle do druge ničle.
  - Interval lahko raztegnemo od ničle do neskončnosti, in s tem zajamemo ničli; Littleov Teorem torej velja tudi za daljše periode za stabilne sisteme.

Littleov Teorem - pregled

- Littleov teorem velja ne glede na: $L = \lambda W$
  - Tip sistema in meje sistema
  - Število strežnikov
  - Karakteristike (naključnost, regularnost) prihodov v sistem
  - Karakteristike strežbe sistema
  - Pristop razvrščanja v vrste ozioma načina strežbe
- Littleov teorem ne velja, če se enote sistema izgubljajo

Povprečen čas zadrževanja zahteve v sist.

- W=E(w)
- Littleov teorem (Little's Theorem): $L = \lambda W$
- Z upoštevanjem teorema lahko zapišemo povprečen čas zadrževanju zahteve v sistemu:

$$W = \frac{L}{\lambda} = \frac{1}{\lambda}\frac{\lambda}{\mu - \lambda} = \frac{\rho}{\lambda(1-\rho)}$$

  - Ta čas je vsota časa čakanja v vrsti in časa strežbe

Povprečno število (čak.) zahtev v vrsti

- L_q = E(n) – E(število zahtev v strežbi)
- Povprečno število v strežbi ni 1, lahko ga izračunamo:

$$E(\text{v str}) = 0P_0 + 1P(>0)$$
$$E(\text{v str}) = 0P_0 + 1(1-P_0) = 1 - P_0 = \rho$$

- Iz česar sledi povprečno število zahtev v vrsti:

$$L_q = E(n) - \rho = \frac{\rho}{1-\rho} - \rho = \frac{\rho^2}{1-\rho}$$

Poprečen čas čakanja v vrsti

- Povprečen čas, ki ga zahteva porabi, preden opravi storitev.
- Od povprečnega časa zadrževanja odštejemo čas strežbe:

$$W_q = W - \frac{1}{\mu} = \frac{\rho^2}{\lambda(1-\rho)}$$

- Do istega rezultata lahko pridemo z uporabo Littleovega teorema:

$$W_q = \frac{L_q}{\lambda}$$

Čakanje na sistemu

- Verjetnost, da je v sistemu manj oz. ravno N paketov (P[Q≤N])

$$P[Q \le N] = \sum_{i=0}^{N} (1-\rho)\rho^i$$

- Verjetnost, da čas čakanja v čakalni vrsti ne preseže m_{Tq} sekund

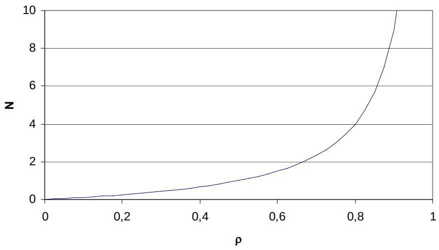
$$m_T(r) = T\ln\left(\frac{100}{100-r}\right)$$

HVALA ZA POZORNOST,
VPRAŠANJA !?

Strežni sistem M/M/1

Pogoj za stabilnost:

- Povprečno število zahtevkov mora biti manjše od povprečnega števila postreženih



Primeri

- Avtomobili na avtocesti:
  - Kot primer si predstavljajmo avtomobile, ki vstopajo na avtocesto. Preverimo, ali ustrezajo pogojem za uporabo:
    1. Celotno število avtomobilov, ki uporablja avtocesto, je relativno veliko.
    2. Posamezni avtomobil uprorablja relativno majhen odstotek avtoceste.
    3. Odločitev posameznega avtomobila o vstopu na avtocesto je neodvisna od odločitev ostalih avtomobilov.
  - Zgornje predpostavke kažejo na to, da proces prihodov avtomobilov na avtocesto lahko aproksimiramo s Poissonovim procesom.
  - Če eden izmed navedenih pogojev ne bi bil izpolnjen, potem ne moremo privzeti Poissonovih prihodov. Če bi – denimo – na avtocesti imeli dirko avtomobilov, potem odločitev posameznika o vstopu na avtocesta ni neodvisna, temveč imajo vsi vozniki skupen cilj.

- M/M/1 Queueing System
- We have already covered queueing theory basics in a previous article. In this article we will focus on M/M/1 queueing system. As we have seen earlier, M/M/1 refers to negative exponential arrivals and service times with a single server. This is the most widely used queueing system in analysis as pretty much everything is known about it. M/M/1 is a good approximation for a large number of queueing systems.
- The following topics will be discussed in detail:
- Poisson Arrivals Poisson Service Times Single Server M/M/1 Results
- Poisson Arrivals
- M/M/1 queueing systems assume a Poisson arrival process. This assumption is a very good approximation for arrival process in real systems that meet the following rules:
- The number of customers in the system is very large.
- Impact of a single customer on the performance of the system is very small, i.e. a single customer consumes a very small percentage of the system resources.
- All customers are independent, i.e. their decision to use the system are independent of other users.
- Cars on a Highway
- As you can see these assumptions are fairly general, so they apply to a large variety of systems. Lets consider the example of cars entering a highway. Lets see if the above rules are met.
- Total number of cars driving on the highway is very large.
- A single car uses a very small percentage of the highway resources.
- Decision to enter the highway is independently made by each car driver.
- The above observations mean that assuming a Poisson arrival process will be a good approximation of the car arrivals on the highway. If any one of the three conditions is not met, we cannot assume Poisson arrivals. For example, if a car rally is being conducted on a highway, we cannot assume that each car driver is independent of each other. In this case all cars had a common reason to enter the highway (start of the race).
- Telephony Arrivals
- Lets take another example. Consider arrival of telephone calls to a telephone exchange. Putting our rules to test we find:
- Total number of customers that are served by a telephone exchange is very large.
- A single telephone call takes a very small fraction of the systems resources.
- Decision to make a telephone call is independently made by each customer.
- Again, if all the rules are not met, we cannot assume telephone arrivals are Poisson. If the telephone exchange is a PABX catering to a few subscribers, the total number of customers is small, thus we cannot assume that rule 1 and 2 apply. If rule 1 and 2 do apply but telephone calls are being initiated due to some disaster, calls cannot be considered independent of each other. This violates rule 3.

- Poisson Service Times
- In an M/M/1 queueing system we assume that service times for customers are also negative exponentially distributed (i.e. generated by a Poisson process). Unfortunately, this assumption is not as general as the arrival time distribution. But it could still be a reasonable assumption when no other data is available about service times. Lets see a few examples:
- Telephone Call Durations
- Telephone call durations define the service time for utilization of various resources in a telephone exchange. Lets see if telephone call durations can be assumed to be negative exponentially distributed.
- Total number of customers that are served by a telephone exchange is very large.
- A single telephone call takes a very small fraction of the systems resources.
- Decision on how long to talk is independently made by each customer.
- From these rules it appears that negative exponential call hold times are a good fit. Intuitively, the probability of a customers making a very long call is very small. There is a high probability that a telephone call will be short. This matches with the observation that most telephony traffic consists of short duration calls. (The only problem with using the negative exponential distribution is that, it predicts a high probability of extremely short calls).
- This result can be generalized in all cases where user sessions are involved.
- Transmission Delays
- Lets see if we can assume negative exponential service times for messages being transmitted on a link. Since the service time on a link is directly proportional to the length of the message, the real question is that can we assume that message lengths in a protocol are negative exponentially distributed?
- As a first order approximation you can assume so. But message lengths aren't really independent of each other. Most communication protocols exchange messages in a certain sequence, the length distribution is determined by the length of the messages in the sequence. Thus we cannot assume that message lengths are independent. For example, internet traffic message lengths are not distributed in a negative exponential pattern. In fact, length distribution on the internet is bi-modal (i.e. has two distinct peaks). The first peak is around the length of a TCP ack message. The second peak is around the average length of a data packet.