VI_Single queue modeling

Perf. Predictions:

- User: time to obtain or wait for the service
- System: # users served or resources utilization level

Measures:

- T: interval
- A: # arrivals in T
- C: # completions in T
- B: busy period in T

Derived:

- lambda = A/T arrival rate
- X = C/T Throughput

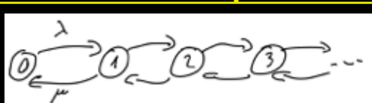
Utilization factor (in single version):

- Rho = B/T Server utilization
- S = B/C avg sac completion x request
- Rho = B/T = S C/T = X S Utilization Law

Little's Law

- W: time in sys by all requests in T
- N = W/T avg # requests in sys
- R = W/C avg residence time x req

Little's law: N = X R = W/T = R C/T



$$\lambda p_0 = \mu p_1 \quad \text{balance}$$

$$p_k = \frac{\lambda}{\mu} p_{k-1} = \frac{\lambda}{\mu}\left(\frac{\lambda}{\mu} p_{k-2}\right) \dots = p_0 \left(\frac{\lambda}{\mu}\right)^k$$

$$\sum p_k = 1 \Rightarrow p_0 = \left[\sum \left(\frac{\lambda}{\mu}\right)^k\right]^{-1} = 1 - \frac{\lambda}{\mu}$$

$$\Rightarrow p_k = p_0 \rho^k \rightarrow \rho = \frac{\lambda}{\mu}$$

Queueing systems

Set of interconnected queues If queues have independent behaviors each queue can be analyzed separately (as Markov Chain)

Kendall's notation

AD/STD/N1/N2/N3
(arrival distribution, service time distribution, #servers, max # users, # potential users)

M/M/1 infinite queues

1 server, A = lambda, sac time = 1/mu
In balanced conditions incoming flow is equal to the outgoing
U = Rho = 1-p0=lambda/mu Utilization factor
<u>N</u> = U/(1-U) Exp # users in sys
R = <u>N</u>/X = S/(1-U) Avg response time (con S=1/mu)
E[L] = rho2/(1-rho) Exp. # users in queue
E[W] = E[L]/lambda = rho/mu(1-rho) Avg time in queue
E[T] = E[N]/lambda = 1/mu(1-rho) Avg time in sys

M/M/1 finite queues

P0{[1-(lambda/mu)^W+1]/[1-(lambda/mu)]} = 1
U = 1- p0. <u>N</u> R = <u>N</u>/X (con S = 1/mu)

Multi server queues