Performance Modeling and Design of Computer Systems- Ch 13 M/M/1 and PASTA

Debobroto Das Robin

Kent State University

Spring 2020

drobin@kent.edu

Overview

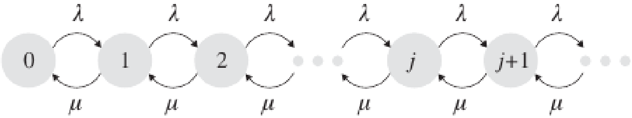
The M/M/1 Queue-1

- **M/M/1 Queue:**simplest queueing model consisting single server
- Service times are i.i.d. Exponential random variables with mean 1μ,
- Jobs arrive according to a Poisson process with rate λ.

- **M/M/p/q:** 4 slot Kendall Notation. **1st M**: arrival process– "memoryless". **2nd M:** distribution of the service – "memoryless" – "exponential" **3rd "p":** number of servers in the system . **4th "q":**upper bound on the capacity of the system in terms of the total space available to hold jobs. **absence of a fourth field indicates that the queue is unbounded and that the scheduling policy is FCFS**

The M/M/1 Queue-2

- **Number of customers** in an M/M/1 system forms a continuous-time Markov chain (CTMC) where the **state** of the system corresponds to the **number of customers** in the system.



- A.K.A. **birth-death process**, with λ – "births" and the μ – "deaths"
- Rate of transitions leaving state i to go to state $i+1 = \pi_i \lambda$
- Solution of Balance Equation: $\pi_i = (\frac{\lambda}{\mu})^i \pi_0$ (page 237 for details)

The M/M/1 Queue-3

Important Metrics : For derivation look at section 13.1

- **server utilization** $= \rho = \frac{\lambda}{\mu}$
- $\rho < 1$ must be met if the system is to be stable (number of jobs not grows without bound). For this condition to be true, we must $\lambda < \mu$
- mean number of customers in the system $= E[N] = \frac{\rho}{1-\rho}$
- variance of the number of customers $= Var(N) = \frac{\rho}{(1-\rho)^2}$

- **Q.** Given an M/M/1 system (with $\lambda < \mu$), if arrival rate λ & service rate μ are both increased k times, What the impact on system performance?
- **Ans:**

$$\lambda_{new} = k\lambda \text{ , } \mu_{new} = k\mu$$

- Utilization rate $= \rho_{new} = \frac{\lambda_{new}}{\mu_{new}} = \rho_{old}$
- Expected number of jobs $E[N_{new}] = \frac{\rho new}{1-\rho new} \frac{\rho old}{1-\rho old} = E[N_{old}]$
- Response time $=$ $E[T_{new}] = \frac{1}{\mu_{new}-\lambda new} = \frac{1}{k(\mu-\lambda)} = \frac{1}{k}E[T_{old}]$

PASTA

Poisson Arrivals See Time Averages)

- Precondition: Arrival process have to be Poisson
- $\pi_n = p_n =$ limiting probability that there are n jobs in the system
- $a_n =$ limiting probability that an arrival sees n jobs in the system
- $d_n =$ limiting probability that a departure leaves behind n jobs in the system when it departs
- **PASTA**: If the arrival process to the system is a Poisson process, then $a_n = p_n$
- Application: If we are simulating a Poisson arrival process to some system and would like to know the mean number of jobs in the system.