Course: Computer and Communication Networks

Topic: Basic Queueing Theory - I (Analysis of M/M/-/- Type Queues and M/G/1 Queues)

Presentation by

Ajay Kakkar

Assistant Professor

Department of Electronics and Communication Engineering,

Thapar Institute of Engineering & Technology
(Deemed to be University)
Bhadson Road, Patiala, Punjab, Pin-147004
Contact No. : +91-175-2393201
Email : info@thapar.edu

THAPAR INSTITUTE
OF ENGINEERING & TECHNOLOGY
(Deemed to be University)

Basic Queueing Theory

Kendall's Notations for Queues

A is the Inter-arrival time distribution
B is the Service time distribution
C is the Number of servers
D is the Maximum number of jobs that can be there in the system (waiting and in service)
Default ∞ for infinite number of waiting positions
E is the Queueing models (FCFS, LCFS, SIRO, PS) *Default is FCFS*

M: exponential
D: deterministic
Ek: Erlangian (order k)
G: general

M/M/1 or M/M/1/∞
Single server queue with Poisson arrivals, exponentially distributed service times and infinite number of waiting positions

The Poisson distribution with parameter λ is given by $\frac{(\lambda t)^n e^{-\lambda t}}{n!}$

Where n is the number of arrivals. We find that if we set n= 0, the Poisson distribution gives us

$$e^{-\lambda t}$$

which is equal to P(T > t) from the exponential distribution.

The exponential distribution with parameter λ is given by

$$\lambda e^{-\lambda t} \text{ for } t \geq 0$$

If T is a random variable that represents inter arrival times with the exponential distribution, then

$$P(T \leq t) = 1 - e^{-\lambda t} \text{ and } P(T > t) = e^{-\lambda t}.$$

Reference: 1. D. Gross and C. Harris, Fundamentals of Queueing Theory, 3rd Edition, Wiley 2. https://nptel.ac.in/courses/117/103/117103017/

Basic Queueing Theory

Little's Theorem states that
N=λW

Result holds in general for virtually all types of queueing situations
where λ = Mean arrival rate of jobs that actually enter the system

Note: Jobs blocked and refused entry into the system will not be counted in λ

The Poisson Arrivals See Time Average (PASTA) Property

pk(t) =P{system is in state k at time t }
qk(t) =P{an arrival at time t finds the system in state k}
N(t) be the actual number in the system at time t
A(t, t+Δt) be the event of an arrival in the time interval (t, t+Δt)

We have

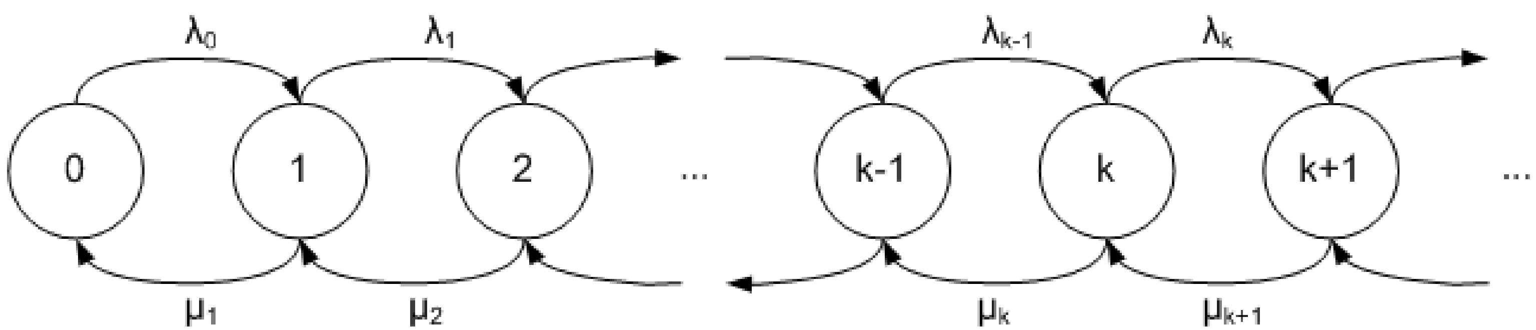
$$q_k(t) = \lim_{\Delta t \to 0} P\{N(t) = k \mid A(t, t+\Delta t\}$$
$$= \lim_{\Delta t \to 0} \frac{P\{A(t, t+\Delta t \mid N(t) = k \mid\} P\{N(t) = k\}}{P\{A(t, t+\Delta t\}} = p_k(t)$$

Because P{A(t, t+Δt)|N(t) = k} = P{A(t, t+Δt)}

Reference: 1. D. Gross and C. Harris, Fundamentals of Queueing Theory, 3rd Edition, Wiley 2. https://nptel.ac.in/courses/117/103/117103017/

Basic Queueing Theory: Equilibrium Solutions for M/M/-/- Queues

M/M/1 (or M/M/1/∞) Queue



$$\lambda_k = \lambda \qquad \forall k$$

$$\mu_k = 0 \qquad k = 0$$
$$\quad = \mu \qquad k = 1, 2, 3, \ldots\ldots$$

For $\rho < 1$

$$p_k = p_0 \left(\frac{\lambda}{\mu}\right)^k = p_0 \rho^k$$

$$p_0 = (1 - \rho)$$

$$N = \sum_{i=0}^{\infty} i p_i = \sum_{i=0}^{\infty} i \rho^i (1 - \rho) = \frac{\rho}{1 - \rho} \qquad W = \frac{N}{\lambda} = \frac{1}{\mu(1 - \rho)}$$

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

Birth-death process

The values in the circles represent the state of the birth-death process, *k*.
The system transitions between values of *k* by "births" and "deaths" which occur at rates given by various values of λ_i and μ_i, respectively.

For a queueing system, *k* is the number of jobs in the system
Furthermore, for a queue, the arrival rates and departure rates are generally considered not to vary with the number of jobs in the queue.

Hence, to consider a single average rate of arrivals/departures per unit time to the queue.

Therefore, for a queue, the figure shown above has arrival rates of $\lambda = \lambda_1, \lambda_2, \ldots, \lambda_k$ and departure rates of $\mu = \mu_1, \mu_2, \ldots, \mu_k$

Reference: 1. D. Gross and C. Harris, Fundamentals of Queueing Theory, 3rd Edition, Wiley 2. https://nptel.ac.in/courses/117/103/117103017/

Basic Queueing Theory: M/M/1/∞ Queue with Discouraged Arrivals

$$\lambda_k = \frac{\lambda}{k+1} \qquad \forall k$$

$$\mu_k = 0 \qquad k = 0$$
$$= \mu \qquad k = 1, 2, 3, \ldots\ldots$$

For $\rho = \lambda/\mu < \infty$

$$p_k = p_0 \prod_{i=0}^{k-1} \frac{\lambda}{\mu(i+1)} = p_0 \left(\frac{\lambda}{\mu}\right)^k \frac{1}{k!}$$

$$p_0 = \exp(-\frac{\lambda}{\mu})$$

$$N = \sum_{k=0}^{\infty} k p_k = \frac{\lambda}{\mu}$$

$$\lambda_{eff} = \sum_{k=0}^{\infty} \lambda_k p_k = \mu \left[1 - \exp(-\frac{\lambda}{\mu})\right]$$

$$W = \frac{N}{\lambda_{eff}} = \frac{\lambda}{\mu^2 \left[1 - \exp(-\frac{\lambda}{\mu})\right]}$$

Reference: 1. D. Gross and C. Harris, Fundamentals of Queueing Theory, 3rd Edition, Wiley 2. https://nptel.ac.in/courses/117/103/117103017/

Basic Queueing Theory: M/M/m/∞ Queue (m servers, infinite number of waiting positions)

$$\lambda_k = \lambda \qquad \forall k \qquad \mu_k = k\mu \qquad 0 \le k \le (m-1)$$
$$= m\mu \qquad k \ge m$$

For $\rho = \lambda/\mu < m$

$$p_k = p_0 \frac{\rho^k}{k!} \qquad for \quad k \le m$$
$$= p_0 \frac{\rho^k}{m!\,m^{k-m}} \qquad for \quad k > m$$

$$p_0 = \left(\sum_{k=0}^{m-1} \frac{\rho^k}{k!} + \frac{m\rho^m}{m!(m-\rho)} \right)^{-1}$$

$$P\{queueing\} = \sum_{k=m}^{\infty} p_k = C(m,\rho) = p_0 \frac{m\rho^m}{m!(m-\rho)}$$

Reference: 1. D. Gross and C. Harris, Fundamentals of Queueing Theory, 3rd Edition, Wiley 2. https://nptel.ac.in/courses/117/103/117103017/

Basic Queueing Theory: M/M/m/m Queue (m servers loss system and no waiting)

$$\lambda_k = \lambda(K-k) \qquad k < K$$
$$= 0 \qquad otherwise$$

$$\mu_k = \mu \qquad k \le K$$
$$= 0 \qquad otherwise$$

For

$$p_k = p_0 \rho^k \frac{K!}{(K-k)!} \qquad k=1,\ldots,K$$

$$\rho = \frac{\lambda}{\mu} < \infty \qquad p_0 = \frac{1}{\sum_{k=0}^{K} \rho^k \frac{K!}{(K-k)!}}$$

Reference: 1. D. Gross and C. Harris, Fundamentals of Queueing Theory, 3rd Edition, Wiley 2. https://nptel.ac.in/courses/117/103/117103017/

Basic Queueing Theory: M/G/1 queue

In queueing theory, an M/G/1 queue is a queue model where arrivals are Markovian, service times have a general distribution and it has a single server.

It is an extension of the M/M/1 queue, where service times must be exponentially distributed.

Here, the model name is written in Kendall's notation.

Example: fixed head hard disk.

M(memoryless): Poisson arrival process, intensity λ
G(general): general holding time distribution, mean $\bar{S} = 1/\mu$
1 : single server, load $\rho = \lambda \bar{S}$
(in a stable queue one has $\rho < 1$)

The probability per time unit for a transition from the state {N=n} to the state {N=n−1}, i.e. for a departure of a customer, depends also on the time the customer in service has already spent in the server;

We start with the derivation of the expectation of the waiting time W. W is the time the customer has to wait for the service (time in the "waiting room", i.e. in the actual queue), R is residual service time

$$\mathrm{E}[W] = \underbrace{\underbrace{\mathrm{E}[N_q]}_{\text{number of waiting customers}} \cdot \underbrace{\mathrm{E}[S]}_{\text{mean service time}}}_{\text{mean time needed to serve the customers ahead in the queue}} + \underbrace{\mathrm{E}[R]}_{\text{unfinished work in the server}}$$

Reference: 1. D. Gross and C. Harris, Fundamentals of Queueing Theory, 3rd Edition, Wiley 2. https://nptel.ac.in/courses/117/103/117103017/

Basic Queueing Theory

R is the remaining service time of the customer in the server (unfinished work expressed as the time needed to discharge the work).
If the server is idle (i.e. the system is empty), then R= 0.

•In order to calculate the mean waiting time of an arriving customer one needs the expectation of Nq(number of waiting customers) at the instant of arrival.

• Due to the PASTA property of Poison process, the distributions seen by the arriving

• customer are the same as those at an arbitrary instant. The key observation is that by Little's result the mean queue length E[Nq] can be expressed in terms of the waiting time.

$$\mathrm{E}[N_q] = \lambda \mathrm{E}[W]$$

It remains to determine E[R].

$$\mathrm{E}[W] = \frac{\mathrm{E}[R]}{1-\rho}$$

$\rho = \lambda E[S]$

Reference: 1. D. Gross and C. Harris, Fundamentals of Queueing Theory, 3rd Edition, Wiley 2. https://nptel.ac.in/courses/117/103/117103017/