Unit IV

Queuing Models

Unit IV: Syllabus

- Queuing Models:
  - Little's Theorem
  - Analytical Results for
    - M/M/1
    - M/M/1/N
    - M/M/c
    - M/G/1 and
  - Other Queuing Models

Learn M & S : From B K Sharma

What is Queuing Models?

Describe the behavior
And
Evaluate performance

of queuing systems.

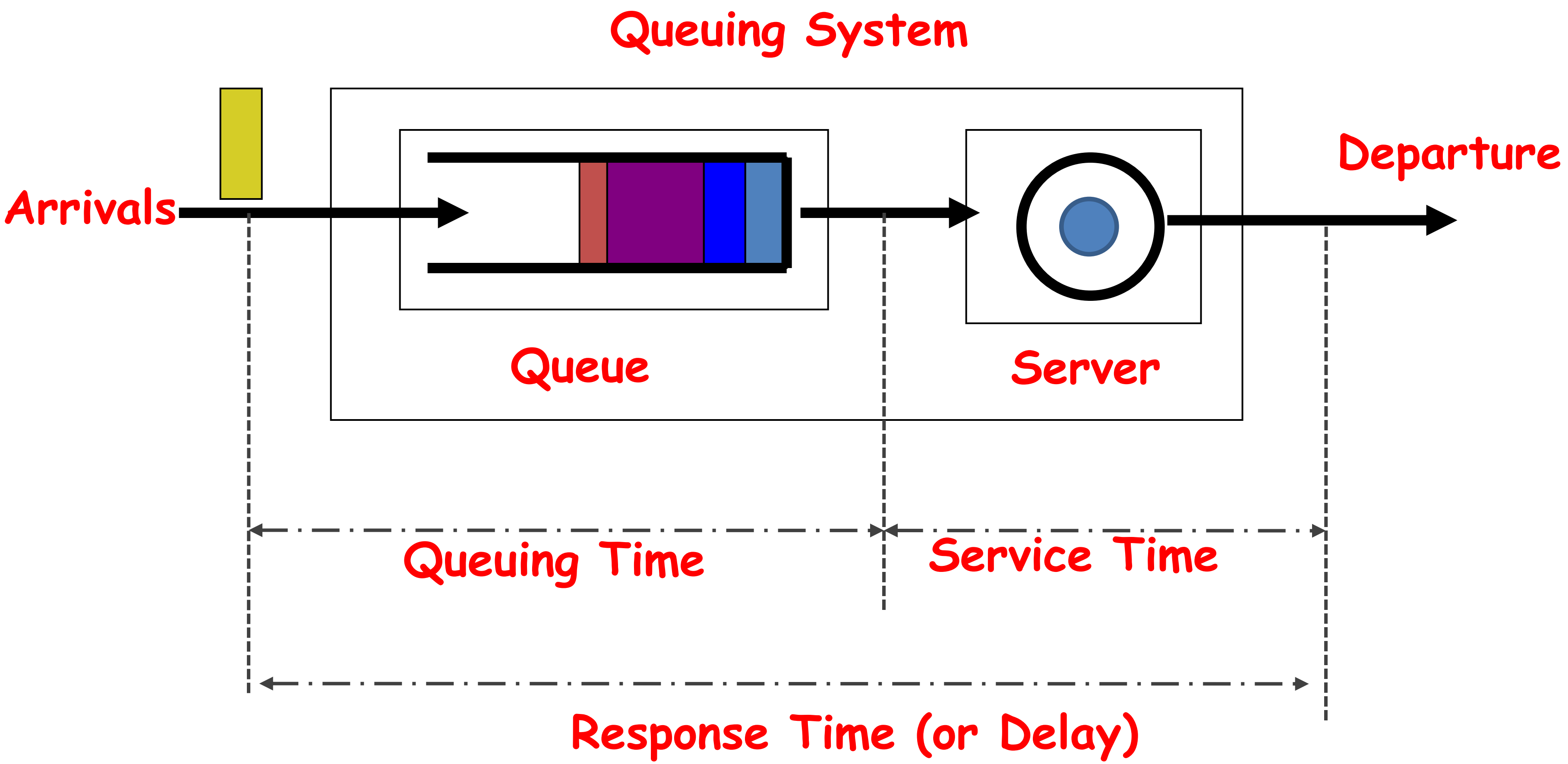
What is Queuing System?

A Queue System is characterized by:

Queue (Buffer)
Server
Events

Learn M & S : From B K Sharma

What is Queuing Models?



Two Types of behaviour of System

Behaviour in Transient Period

And

Behaviour in Steady State Period

The transient period occurs at the initial time of operation.

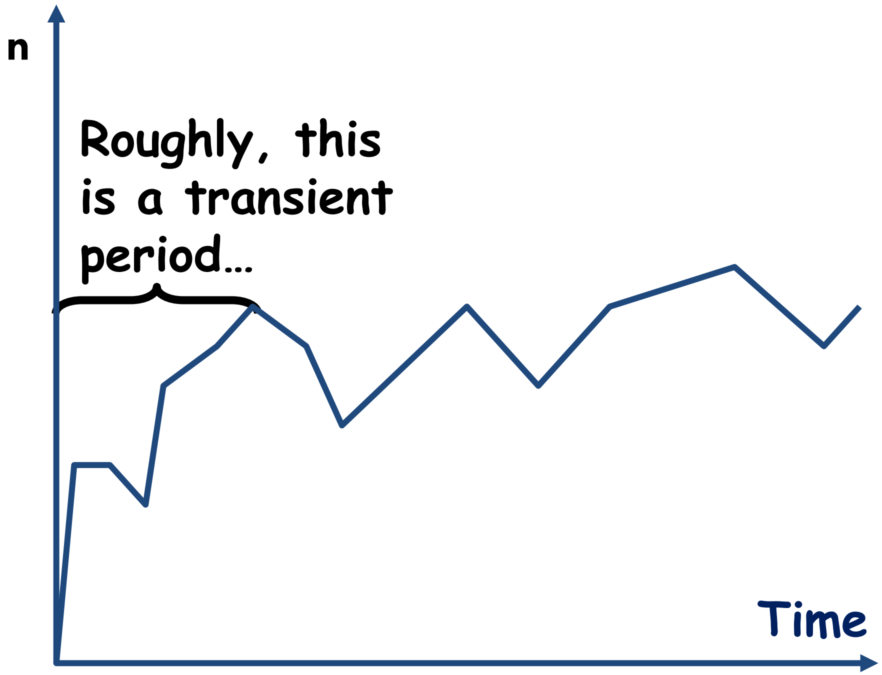
Initial transient behavior is not indicative of long run performance.

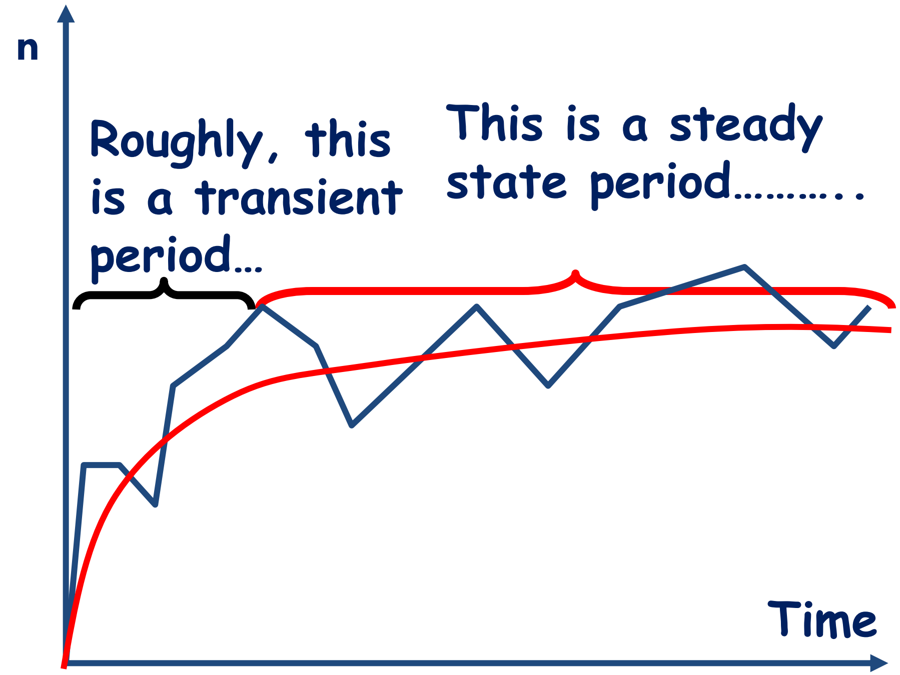
The steady state period follows the transient period.

Meaningful long run performance measures can be calculated for the system when in steady state.

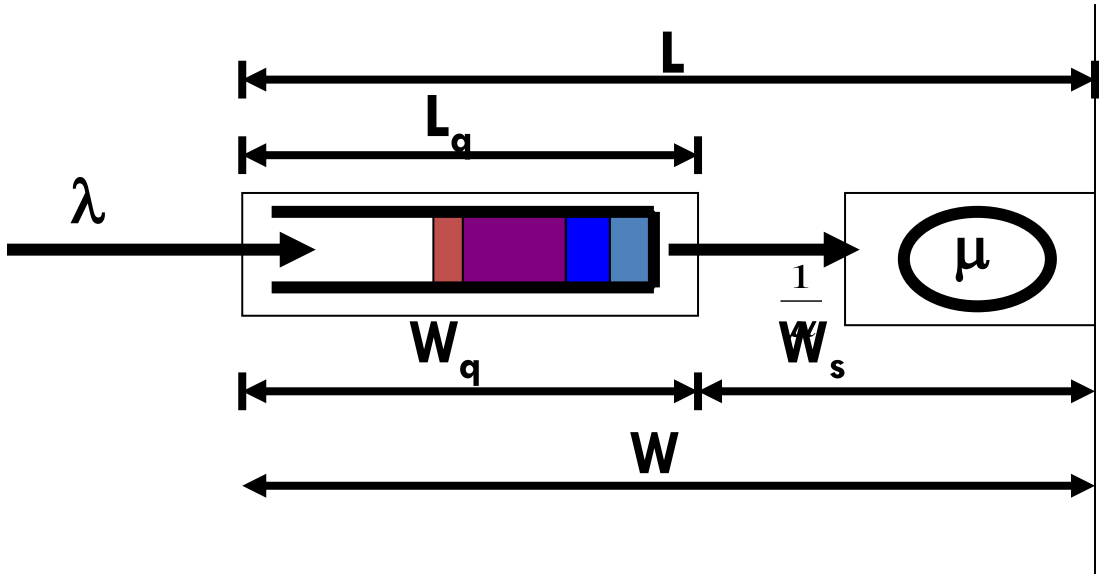
Learn M & S : From B K Sharma

Two Types of behaviour of System





State Performance Measures of Queuing System



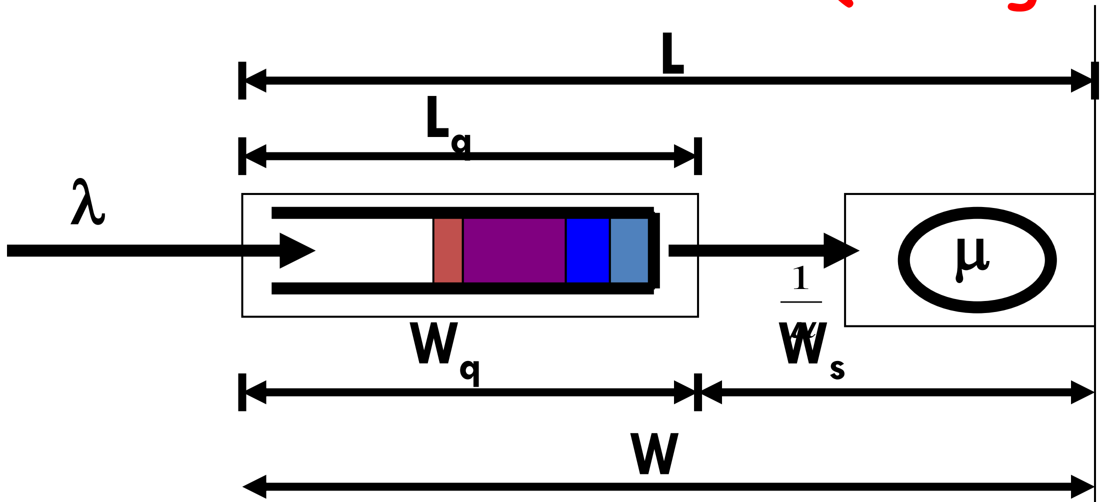
P_0 = Probability that there are no customers in the system.

P_n = Probability that there are "n" customers in the system.

L = Average number of customers in the system.

L_q = Average number of customers in the queue.

State Performance Measures of Queuing System



W = Average time a customer spends in the system.

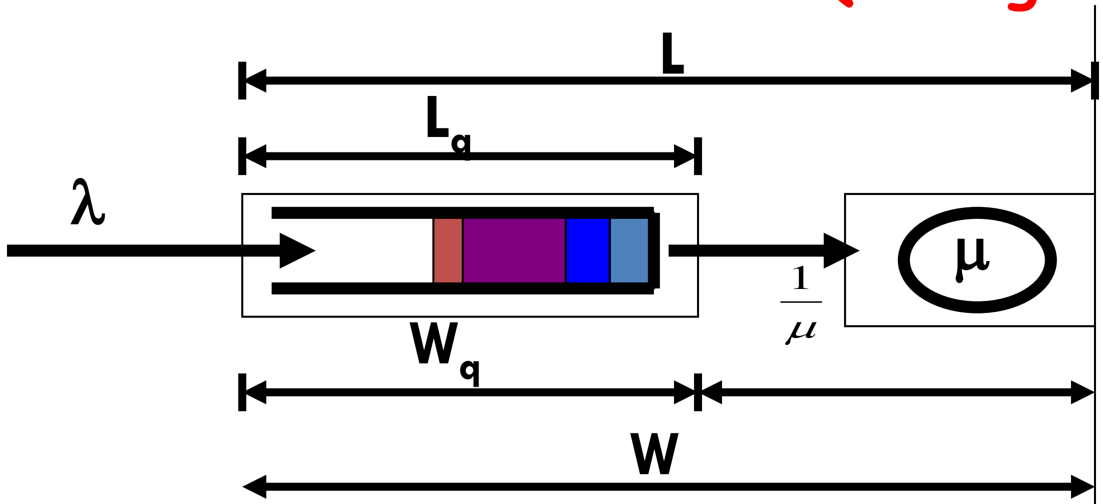
W_q = Average time a customer spends in the queue.

P_w = Probability that an arriving customer must wait for service.

ρ = Utilization rate for each server (the percentage of time that each server is busy).

Learn M & S : From B K Sharma

State Performance Measures of Queuing System



λ = average number of arrivals *entering* the system per unit time.

1/λ = mean inter arrival time, time between arrivals.

μ = mean service rate per server = average number of units that a server can process per period.

1/μ = mean service time

Little Law

$$E[N] = \lambda \cdot E[T]$$

Expected number of customers in the system

Expected time in the system

Arrival rate IN the system

The long-term average number of customers in a stable system N, is equal to the long-term average arrival rate, λ, multiplied by the long-term average time a customer spends in the system, T.

Example: Amusement park analogy

People arrive, spend time at various sites, and leave.

Over a long horizon: arrivals about the same as departures.

Little Law

$$E[N] = \lambda \cdot E[T]$$

Example: server satisfies I/O request in average of 100 msec(T). I/O rate is about 100 requests/sec *(λ)*. What is the mean number of requests at the server?

Mean number at server = arrival rate x response time
= (100 requests/sec)x(0.1 sec)
= 10 requests

Little Formula

Little's Formulas represent important relationships between L, W, L_q, and W_q.

$$L = \lambda W$$

$$L_q = \lambda W_q$$

$$L = L_q + \lambda/\mu$$

For the case of an infinite population

These formulas apply to systems that meet the following conditions:

Single queue systems,

Customers arrive at a finite arrival rate λ, and

The system operates under a steady state condition.

Kendall Notation

To Specify a Queue, we use Kendal Notation.

There are six Parameters:

(A/B/m(/K/N/X)

1. A: Arrival Distribution
2. B: Service Distribution
3. m: Number of servers
4. K: Storage Capacity (infinite if not specified)
5. N: Population Size (infinite)
6. X: Service Discipline (FCFS/FIFO)

Learn M & S : From B K Sharma

Kendall Notation

A/B/m/K/N/X

Arrival Process
- M: Markovian (Poisson)
- D: Deterministic
- E_r: Erlang
- G: General

Service Process
- M: Markovian (Poisson)
- D: Deterministic
- E_r: Erlang
- G: General

Number of servers
m=1,2,…

Service Discipline
FIFO, LIFO, Round Robin, …

Storage Capacity
K= 1,2,…
(if ∞ then it is omitted)

Number of customers
N= 1,2,…
(for closed networks, otherwise it is omitted)

D: deterministic (known) arrivals and constant service duration. Constant arrival rate or service time.
M: process with exponential distribution of intervals or service duration respectively. Poisson arrivals or exponential service time.

Markovian (Poisson) Arrival/Departure Process

REQUIRED CONDITIONS

Orderliness

at most one customer will arrive/depart in any small time interval of Δt.

Stationarity

for time intervals of equal length, the probability of n arrivals/completing n potential services, in the interval, is constant.

Independence

the time to the next arrival is independent of when the last arrival occurred.

the time to the completion of a service is independent of when it started.

Learn M & S : From B K Sharma

Based on **Queuing Models** Kendall Notation

(A/B/m(/K/N/X)

M/M/1

1. Poisson arrivals (*exponential inter-arrival*)
2. Poisson service (*exponential Service*)
3. 1 server
4. infinite capacity
5. Infinite population
6. FCFS (FIFO)

M/M/1//N/

1. Poisson arrivals (*exponential inter-arrival*)
2. Poisson service (*exponential Service*)
3. 1 server
4. infinite capacity
5. Population Size fixed to N
6. FCFS (FIFO)

M/M/c

1. Poisson arrivals (*exponential inter-arrival*)
2. Poisson service (*exponential Service*)
3. c servers
4. infinite capacity
5. Infinite population
6. FCFS (FIFO)

M/G/1

1. Poisson arrivals (*exponential inter-arrival*)
2. General service(can be anything)
3. 1 server
4. infinite capacity
5. Infinite population
6. FCFS (FIFO)

M / M /1 Queue - Performance Measures

$$P_0 = 1 - (\lambda/\mu)$$

$$P_n = [1 - (\lambda/\mu)](\lambda/\mu)^n$$

$$L = \lambda / (\mu - \lambda)$$

$$L_q = \lambda^2 / [\mu(\mu - \lambda)]$$

$$W = 1 / (\mu - \lambda)$$

$$W_q = \lambda / [\mu(\mu - \lambda)]$$

$$P_w = \lambda / \mu$$

$$\rho = \lambda / \mu$$

The probability that a customer waits in the system more than "t" is $P(X>t) = e^{-(\mu - \lambda)t}$

λ is constant and $\mu > \lambda$ (average service rate > average arrival rate).

M/M/1//N/ Queue - Performance Measures

For $\rho = (\lambda/\mu) \neq 1$

$$P_0 = \frac{1-\rho}{1-\rho^{K+1}}$$

$$P_n = \rho^n P_0 = \frac{1-\rho}{1-\rho^{K+1}} \cdot \rho^n$$

$$L = \frac{\rho}{1-\rho} - \frac{(K+1)\rho^{K+1}}{1-\rho^{K+1}}$$

$$L_q = L - (1 - P_0)$$

$$W = L / \overline{\lambda}$$

$$W_q = L_q / \overline{\lambda}$$

Where $\overline{\lambda} = \sum_{n=0}^{\infty} \lambda_n P_n$

M/M/c Queue - Performance Measures

Allows for c identical servers working independently from each other.

$$P_0 = \left(\sum_{n=0}^{c-1} \frac{(\lambda/\mu)^n}{n!} + \frac{(\lambda/\mu)^c}{c!} \cdot \frac{1}{1-(\lambda/(c\mu))} \right)^{-1}$$

$$P_n = \begin{cases} \dfrac{(\lambda/\mu)^n}{n!} P_0 & \text{for} \quad n = 1,2,\ldots,c \\[2ex] \dfrac{(\lambda/\mu)^n}{c!\,c^{n-c}} P_0 & \text{for} \quad n = c+1, c+2, \ldots \end{cases}$$

$$L_q = \sum_{n=c}^{\infty} (n-c) P_n = \ldots = \frac{(\lambda/\mu)^c \rho}{c!(1-\rho)^2} P_0$$

Little's Formula $\Rightarrow$ $W_q = L_q/\lambda$

M/M/c Queue - Performance Measures

$W = W_q + (1/\mu)$

Little's Formula $\Rightarrow$ $L = \lambda W = \lambda(W_q + 1/\mu) = L_q + \lambda/\mu$

Steady State Condition:

$\rho = (\lambda / c\mu) < 1$

M/G/1 Queue - Performance Measures

$P_0 = 1 - \rho$

$$L = \rho + \{\lambda^2 (\mu^{-2} + \sigma^2)\} / \{2(1 - \rho)\}$$
$$= \rho + \{\rho^2 (1 + \sigma^2 \mu^2)\} / \{2(1 - \rho)\}$$

$$L_q = \{\lambda^2 (\mu^{-2} + \sigma^2)\} / \{2(1 - \rho)\}$$
$$= \{\rho^2 (1 + \sigma^2 \mu^2)\} / \{2(1 - \rho)\}$$

M/G/1 Queue - Performance Measures

$W = \mu^{-1} + \{\lambda (\mu^{-2} + \sigma^2)\} / \{2 (1 - \rho)\}$

$W_q = \{\lambda (\mu^{-2} + \sigma^2)\} / \{2 (1 - \rho)\}$

$\rho = \lambda / \mu$

Questions

- Explain Little Theorem in the queuing models with the help of an example.
- Explain Little's Law of queuing theory.
- List and explain various measures of performance for queuing systems.
- Describe the techniques used for queuing and also discuss queuing discipline.

Questions

- What are queuing Models? Explain the factors that affect Queuing Models. Write down the performance measures for M/M/C.
- Arrival of self-service gasoline pump occurs in an exponential fashion at a rate of 12/hr. Service time has distributed that average 14 minutes. What is expected number of vehicles in the system?
- Compare and Contrast M/M/1 and M/M/1/N queuing model with respect to their performance measures factors.

Questions

- What do you mean by generation of arrival pattern? Briefly explain it.
- How various arrival patterns are generated for the queues? Explain with examples.
- Discuss Kendall's notation for specifying the characteristics of a queue with an example.