Unit-3 Queuing System

DATE

3.1. Characteristics and structure of Basic Queuing System & Models of queuing system

* Concept of Basic Queuing System

→ A queuing system is a facility consisting of one or several servers designed to perform certain tasks or process certain jobs and a queue of jobs waiting to be processed.

→ Jobs arrive at the queuing system, wait for an available server, get processed by the server, and leave.

→ Ex: An Isp whose customers connect to the internet, browse & disconnect.

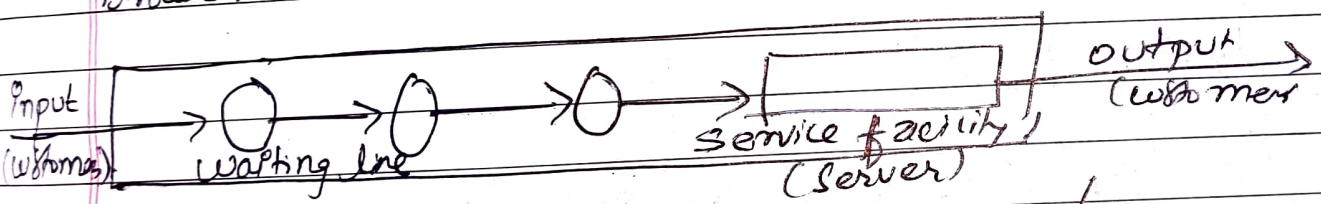


fig: Structure of a basic queuing system (queuing model)

* Characteristics or elements of the queuing system

The elements or characteristics of queuing system a

- Calling population
- System capacity
- Arrival process
- Queue behaviour / Queue discipire.
- Service process

Note: Customer - people, machine, patients,
Server - receptionist, mechanics, doctors

classmate

PAGE

DATE

1. Calling population.

It is the population of potential customers. The calling population may be finite or infite population model. The assumption of an infinite population is such that the rate of arrival of customers is not affected by the number of customers that have already joined the system. It means, the rate of arrival is constant throughout time

2. System capacity

In many queuing systems, there is a limit to the no of customers that may be in the waiting line or syste An arriving customer who finds the system full does enter but returns immediately to the calling population However, there are other systems that have infinite capacity So, System capicity is either limited or unlimited

3. Arrival process.

The arrival process for infinite popn models is usually Characterized in terms of inter-arrival times of successive customers. Arrivals may occur at scheduled times or random times. When at random times, the inter-arrival times are usually characterized by a probability distribution.

4. Queue Behaviour & Queue Discipline.

Queue behaviour refers to the actions of customers while in a queue waiting for service to begin. They are possibly walk, Renege or jockey queue behaviour

classmate

PAGE

DATE

Queue discipline refers to the logical ordering of customers in a queue & determines which customer will be chosen for service when a server becomes free. Common queue disciplines are FIFO, LIFO, Service In Random Order (SIRO), Priority and so on. (SPT (Shortest process time)

5. Service Process.

Once the entities/customers have entered the system, they must be served. The physical meaning of "service" may vary from system to system. From the modeling point of view, we care about whether customers are processed in FCFS order or according to some kind of priority, etc.

Markov/exponential service process: It is a special service process in which entities are processed one at a time in FCFS order and serve time is independent & exponential. It is a memoryless service process.

Note:

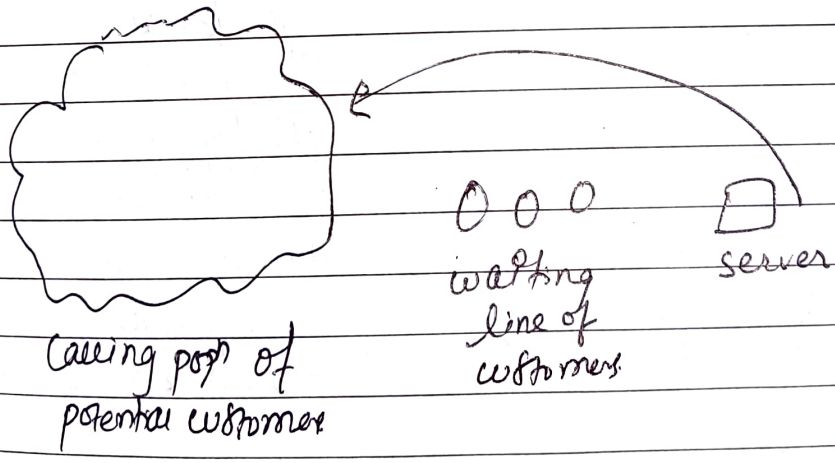


fig: Simple queuing model.

Example of queuing systems:

Systems	Customers	Server(s)
Reception desk	People	Receptionist
Garage	Trucks	Mechanic
Airport	Planes	Runway

DATE

3.2 Queuing Notation

* Kendall's notation for queuing system

Different notations are frequently used in queuing system and kendall's notation is one of them. Kendall proposed a notational system for parallel servers systems which has been widely adopted. It can be represented in the form A/B/C/N/K/D where.

A → interarrival time distribution
B → Service time distribution
C → No. of parallel servers
N → System capacity
K → Size of calling population
D → Queuing discipline.

Common symbols for A & B include M: Exponential/Markov
D: Constant or Deterministic
Ek: Erlang of order k
PH: Phase-type
H: Hyper-exponential
G: Arbitary or General
GI: General Independent

When final three parameters are not specified eg. M/H/1, it is assumed to be N=∞, K=∞, and D=FIFO

Ex: M/H/1 or M/H1/∞/∞ both same
M → Exponential/Markov distributed interarrival time
M → Exponential/Markov distributed service time
1 → Single server system
∞ → Unlimited system capacity
∞ → Infinite calling pop with FIFO queuing discipline.

PAGE

G/M/5/100

→ General time dist.
→ Markov service time dist
→ 5 parallel servers system
→ 100 system capacity
→ ∞ calling popn
→ FIFO queuing discipline

PH/G1/2/50/1000/PR

→ Phase-type time dist and
→ General Indep. service time dist
→ 2 parallel server system
→ 50 system capacity
→ 1000 calling popn
→ priority queuing discipline

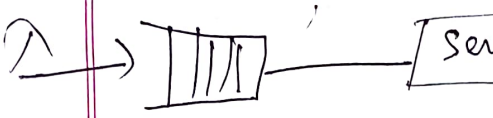
3.3 Single Server and Multiserver queuing system

* Single Server queuing system

→ It is a queuing system with only one server for any number of clients.
→ It is a FIFO queuing system with Kendall notation M/M/1 with poisson input, exponential service time and unlimited waiting positions.
→ The model is based on following assumption.

- The arrival follows poisson distribution with mean arrival rate λ
- The service time has exponential distribution, average service rate μ
- Arrivals are infinite population
- Customers are served on FIFO basis.
- There is only a single server.

In a single server queuing system, there is an infinite number of waiting positions in the queue. Hence, there can be any number of customers in the queue. So, it becomes a challenge to maintain the service rate in such a way to matchup with continuous arrivals of customers.



DATE

$$Q = \begin{pmatrix} -\lambda & \lambda & & & \\ \mu & -(\mu+\lambda) & \lambda & & \\ & \mu & -(\mu+\lambda) & \lambda & \\ & & \mu & -(\mu+\lambda) & \lambda \\ & & & \vdots & \end{pmatrix}$$

The model is considered stable only if $\lambda < \mu$. We write $\rho = \lambda/\mu$ for the utilization of buffer & require $\rho < 1$ for the queue to be stable.

Performance of M/M/1 System:- (measurement of queue system perf

Given, λ = Arrival rate of customers
μ = service rate of the server.

- P_0 = probability that there are no customers in the system. (Server idle ratio)
$$P_0 = 1 - \lambda/\mu$$
- P_n = Probability that there are "n" customers in the syst
$$P_n = [1 - \lambda/\mu][\lambda/\mu]^n$$
- L = Average no. of customers in the system
$$L = \lambda/\mu - \lambda$$
- L_q = Average no. of customers in the queue
$$L_q = \left(\lambda/\mu - \lambda\right)\left(\lambda/\mu\right)$$
- W = Average time a customer spends in the system
$$W = 1/\mu - \lambda$$

classmate

PAGE

- W_q = Average time a customer spends in queue.

$$W_q = \left(\frac{1}{\mu - \lambda}\right)\left(\frac{\lambda}{\mu}\right)$$

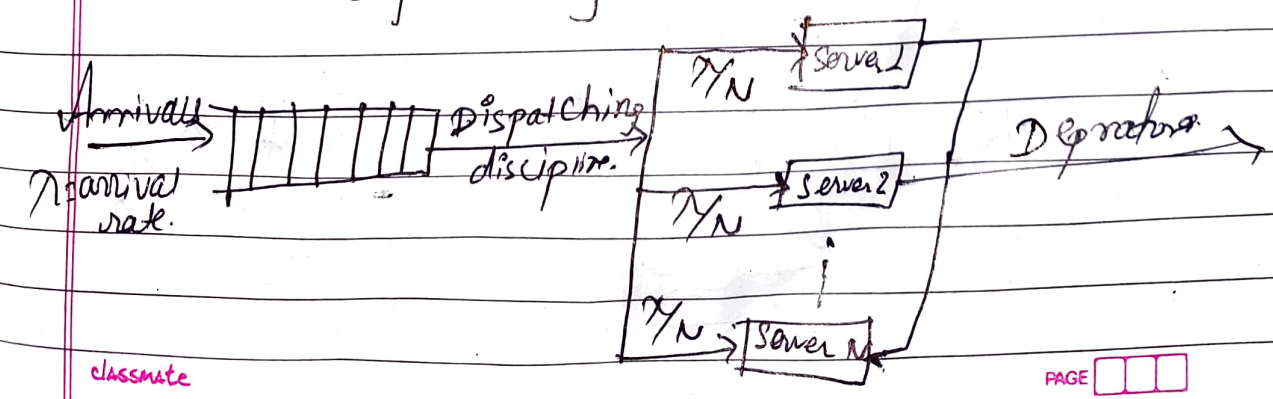
- P_w = Probability that an arriving customer must wait for service

$$P_w = \lambda/\mu$$

- ρ = Server utilization ratio = λ/μ

Multiserver Queuing System

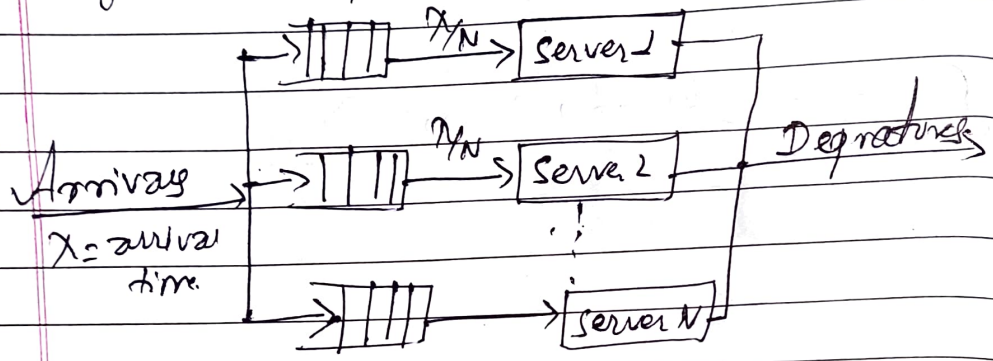
→ It is the queuing system with more than one server.
→ In this system, all share a common queue.
→ If an item arrives and at least one server is available, then the item is immediately dispatched to the server.
→ It is assumed that all servers are identical, it makes no difference which server is chosen for an item.
→ If all servers are busy, a queue begins to form
→ Multi-server queuing system is represented by M/M/c where arrivals form single queue & are governed by a poisson process. There are c-servers & job service times are exponentially distributed.



DATE

The total Server Utilization in case of Multi-Server queue for N server queue is $\rho = \lambda/c\mu$ where μ is the service rate & λ is the arrival rate.

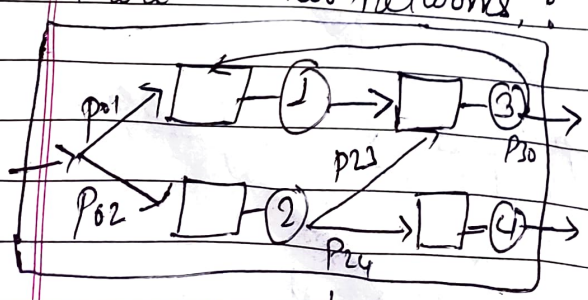
➡ There is another concept which is called multiple Single server queue system



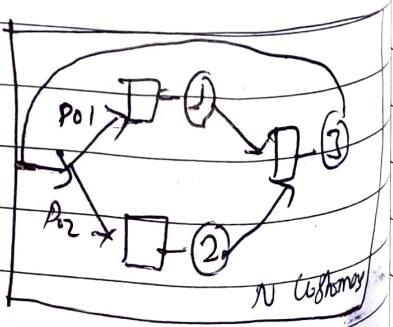
9.3.1 Networks of queuing (Networks of queues)

→ Networks of queues are the systems in which a number of queues are connected by customer routing.
→ When a customer is serviced at one node, it can join another node & queue for service, or leave the network.
→ A queuing network is a system composed of several interconnected stations, each with a queue.

There are two networks :



Open network



Closed network

The following results assume a stable system with infinite calling population & no limit on system capacity

- Provided that no customers are created or destroyed in the queue, the departure rate of queue is same as arrival rate in the queue.
- If customers arrive to queue i at rate λ_i, & a fraction $0 \le p_{ij} \le 1$ of them are routed to queue j upon departure, then arrival rate from queue i to queue j is $\lambda_i p_{ij}$
- The overall arrival rate into queue j, λ_j, is the sum of arrival rate from all sources. If customers arrive from outside the network at rate a_i, then, $\lambda_j = a_j + \sum_{\text{all } i} \lambda_i p_{ij}$
- If queue j has $c_j < \infty$ parallel servers, each working at rate μ_j, then the long-run utilization of each server is $\rho_j = \frac{\lambda_j}{c_j \mu_j}$

3.3.4 Applications of queuing system

- Used in designing & operating transportation systems such as airports, freeways, ports, and subways.
- It is used in determining min. no. of servers needed at a service center.
- Used in analysis of production & material handling systems.
- Used in the telephone & communication system
- Used in analysis of telecommunications, computer networks, predicting computer performance, traffic, etc.
- Used in commercial organizations Ex: dentists, bank, ATM, gas stations etc. for serving external customers

DATE

Numericals:

Q. Customers arrive at Mary's Shoes every 12 minutes on the average, according to a Poisson process. Service time is exponentially distributed with an average of 8 min per customer. Management is interested in determining the performance measures for this service system

⇒ Given, $\lambda = \frac{1}{12}$ customers per minute $= \frac{60}{12} = 5$ per hour

$\mu = \frac{1}{8}$ customer per minute $= \frac{60}{8} = 7.5$ per hour

Now, $P_0 = 1 - \frac{\lambda}{\mu} = 1 - \frac{5}{7.5} = 0.3333$

$L = \frac{\lambda}{\mu - \lambda} = 2$

$L_q = \frac{\lambda}{\mu - \lambda} \cdot \frac{\lambda}{\mu} = 1.333$

$W = \frac{1}{\mu - \lambda} = 0.4$ hrs $= 24$ mins

$W_q = \frac{\lambda}{\mu} \cdot \frac{1}{\mu - \lambda} = 0.26667$ hours $= 16$ minutes

Q. Customers arrive in a bank according to a Poisson's process with mean interarrival time of 10 minutes. Customers spend an average of 5 minutes on the single available counter, and leave. Discuss.

i. What is the probability that a customer will not have to wait at the counter?

ii. What is the expected no. of customers in the bank?

iii. How much time can a customer expect to spend in the bank?

DATE

=) Given,

$\lambda = \frac{1}{10}$ customers per minute $= \frac{60}{10}$ customers per hr $= 6$.

$\mu = \frac{1}{5}$ customers per minute $= \frac{60}{5}$ customers per hr $= 12$

i) => The customers will not have to wait if there are no customers in the bank

$\therefore P_0 = 1 - \lambda/\mu = 1 - \frac{6}{12} = 0.5$

ii) => Expected number of customers $= \lambda/\mu - \lambda = 6/6 = 1$

iii) => Expected time to be spent in the bank $= \frac{1}{\mu - \lambda} = 10$ min

Q. Customers arrive at a sales counter manned by a single person according to a poisson process with a mean rate of 20 per hour. The time required to serve a customer has an exponential distribution with a mean of 100 seconds. Find the average waiting time of a customer for the system and queue.

=> Arrival rate (λ) = 20 per hour

Service rate (μ) = $\frac{1}{100}$ customers per second $= \frac{60 \times 60}{100} = 36$ customers per hours.

The average waiting time of a customer in the queue

$= \frac{\lambda}{\mu(\mu - \lambda)} = \frac{20}{36(36-20)} = \frac{5}{36 \times 4}$ hrs $= 125$ seconds

The average waiting time of a customer in the system

$= \frac{1}{\mu - \lambda} = \frac{1}{36-20} = \frac{1}{16}$ hrs $= \frac{60 \times 60}{16}$

$= 225$ sec.

classmate

PAGE

Q. Self service at a university cafeteria, at an average rate of 7 minutes per customer, is slower than attendant service, which has a rate of 6 minutes per student. The manager of the cafeteria wishes to calculate the average time each student spends waiting for service. Assume that customers arrive randomly at each time, at the rate of 5 per hour. Calculate the appropriate statistics for this cafeteria

	Self service	Attended
=> Arrival rate (λ)	5	5
Service rate (μ)	$60/7 = 8.571$ per hr.	$60/6 = 10$ per hr
Expected no. of customers in the system	$\frac{\lambda}{\mu-\lambda} = \frac{5}{8.571-5} = 1.40$	$\frac{5}{10.5} = 1$
Expected no of students waiting for service.	$\frac{\lambda^2}{\mu(\mu-\lambda)} = 0.82$	0.5
Average time in system	$\frac{1}{\mu-\lambda} = 0.28$ hr	$= 0.20$ hr
Average time in queue. -	$\frac{1}{\mu-\lambda} \times \frac{\lambda}{\mu} = 0.1633$ hr	$= 0.10$ hr

classmate