Service mechanism

- The **service mechanism** is the way that customers receive **service** once they are selected from the front of a **queue**.
- It is the pattern according to which the customers are served.
- A specification of the service mechanism includes a description of time to complete a service and the number of customers who are satisfied at each service event.
- The service mechanism also prescribes the number and configuration of servers. If there is more than one service facility, the calling unit may receive service from a sequence of these. At a given facility, the unit enters one of the parallel service channels and is completely serviced by that server. Most elementary models assume one service facility with either one or a finite number of servers.

  Single server facility

  Multi server facility

Customer's Behaviour

1. **Balking.** A customer may not like to join the queue due to long waiting line.
2. **Reneging.** A customer may leave the queue after waiting for sometime due to impatience. The customer enters the line but decides to leave before being served.
3. **Collusion.** Several customers may cooperate and only one of them may stand in the queue.

4. **Jockeying.** When there are a number of queues, a customer may move from one queue to another in hope of receiving the service quickly. In this customer enters one line and then switches to a different one in an effort to reduce the waiting time

Server's Behaviour

- **Failure.** The service may be interrupted due to failure of a server (machinery).
- **Changing service rates.** A server may speed up or slow down, depending on the number of customers in the queue. For example, when the queue is long, a server may speed up in response to the pressure. On the contrary, it may slow down if the queue is very small.
- **Batch processing.** A server may service several customers simultaneously, a phenomenon known as batch processing.

SYMBOLS AND NOTATIONS

The following symbols and notations will be used in connection with the queuing systems:

n = number of customers in the system, both waiting and in service.

λ = average number of customers arriving per unit of time.

μ = average number of customer being served per unit of time.

λ / μ = traffic intensity.

C = number of parallel service channels (servers).

E(n) = average number of customers in the system. Both waiting and in service.

E(m) = average number of customers waiting in the queue.

E(v) = average waiting time of a customer in the system, both waiting and in service.

E(w) = average waiting time of a customer in the queue.

$P_n(t)$ = Probability that there are n customers in the system at any time t, both waiting and in service.

P_n = time independent Probability that there are n customers in the system at any time, both waiting and in service.

CLASSIFICATION OF QUEUING MODELS AND THEIR SOLUTIONS

Different models in queuing theory are classified by using special (or standard) notations described initially by **D.G.Kendall** in 1953 in the form (a/b/c). Later A.M.Lee in 1966 added the symbols d and c to the Kendall notation. Now in the literature of queuing theory the standard format used to describe the main characteristics of parallel queues is as follows:

$$\{(a/b/c) : (d/c)\}$$

Where

a = arrivals distribution

b = service time (or departures) distribution

c = number of service channels (servers)

d = max. number of customers allowed in the system (in queue plus in service)

e = queue (or service) discipline.

Certain descriptive notations are used for the arrival and service time distribution (i.e. to replace notation a and b) as following:

M = exponential inter-arrival times or service-time distribution (or equivalently poisson or markovian arrivel or departure distribution)

D = constant or deterministic inter-arrival-time or service-time.

G = service time (departures) distribution of general type, i.e. no assumption is made about the type of distribution.

GI = Inter-arrival time (arrivals) having a general probability distribution such as as normal, uniform or any empirical distribution.

E_k = Erlang-k distribution of inter-arrival or service time distribution with parameter k (i.e. if k= 1, Erlang is equivalent to exponential and if k = , Erlang is equivalent to deterministic).

For example, a queuing system in which the number of arrivals is described by a Poisson probability distribution, the service time is described by an exponential distribution, and there is a single server, would be designed by M/M/I.