

4. Waiting Line Theory or Queueing Model

Waiting Line Theory or Queueing Model

Definitions :

$(M/N/S) : (C/D)$

M - Arrival Pattern

N - Service Pattern

S - Service Channels

C - Service Capacity

D - Service Discipline

e.g., $(M/N/1) : (\infty/FIFO)$

A queue is formed when there is disbalance between number of servers and number of customers.

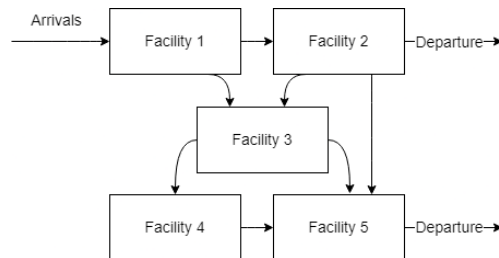
A flow of customers from finite or infinite population towards the service facility forms a queue on account of lack of capability to serve them all at a time.

In absence of perfect balance between the service facilities and the customers, waiting time is required either for the service facilities or the customer's arrival.

Customer - An arriving unit waiting to be serviced.

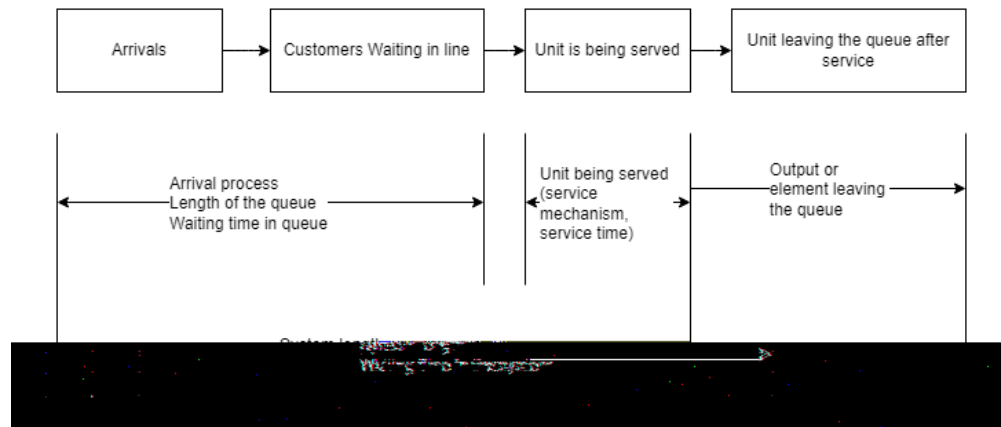
Queue - No of customers waiting to be serviced.

Service Facility - The body providing the service.

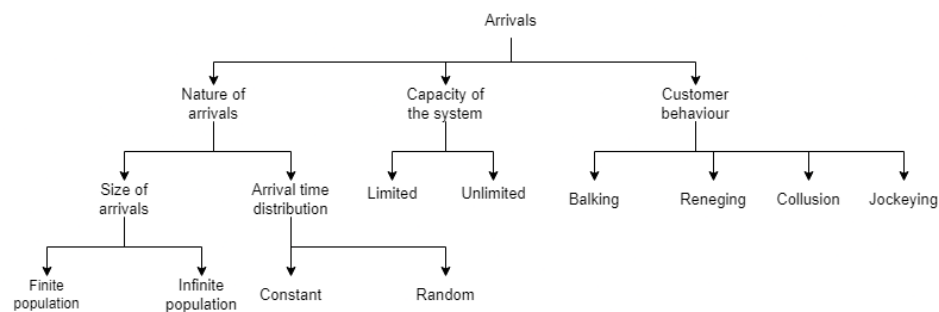


Queueing system / process

- a> Input (Arrival pattern)
- b> Service mechanism / service pattern
- c> Queue discipline
- d> Customer behaviour



Input process



characteristics of arrivals

a) Size of arrivals - Depends on the nature of size of the population (i.e. finite / infinite). More specifically described in terms of probabilities and probability distributions for inter-arrival time (time between two successive arrivals or the distribution of customers arriving in a unit time must be defined).

Note:

For simplicity only Poisson arrivals are considered at the moment.

b) Inter-arrival time - the period between two successive arrivals. Most

queueing models assume that some inter-arrival time distraction exists for all customers exists through the period of study. In most situations, the service time is a random variable with the same distribution for all arrivals but there might be cases where there are clearly two or more classes of customers such as machines waiting for repair with different service time distributions. Here mostly the important service time distributions are considered (*Negative exponential* and *Erlang or Gamma Distribution*)

c) capacity of the system - Space available for arrivals to wait before being taken to service. The space available may be limited or unlimited. When the space is limited, if the length of the queue crosses a certain limit : no more arrivals are permitted to enter the system till the waiting space becomes vacant. This type of system is called *system with finite capacity* and has effects on the arrival pattern of the system.

d) customer behaviour - The length of the queue depends on the behaviour of the customer, i.e. the impatience of the customer during their stay in the queue. Customer behaviour can be classified as

- **Balking** - When the customer does not like to join the queue seeing the length of it. This behaviour might result in losing a customer. A lengthy queue indicates insufficient service facility and the customer(s) may not turn out next time.
- **Reneging** - In this case, a customer joins the queue and after waiting for a certain time loses his patience and leaves the queue. This behaviour may also result in loss of a customer.
- **Collusion** - Several customers may collaborate and only one of them may stand in the queue. One customer represents a group of customers. In this case, the queue itself might be small but service time for an individual will be more. This may lead to other customers becoming impatient.
- **Jockeying** - If there are multiple queues depending on the number of service stations, a customer in one of the queues after seeing the other queue length which is shorter, might leave the present queue and join the other queue with the hopes of getting the service faster. Perhaps the shorter queue might have more collaborated groups which might lead to longer service time. In such case the probability of the customer who has switched queues getting the service may be very less. Because of this customer behaviour, the queue lengths may keep changing from time to time.