Queueing Theory

[illegible]

A flow of customers from finite / infinite population towards the service facility forms a queue on account of lack of capability to serve them all at one time. In the absence of a perfect balance between the service facilities and the customers, waiting is required either of the service facility or for the customers. The subject of queueing is not directly concerned with optimization, it attempts to explore, understand and compare various queueing situations, thus indirectly achieving approximate optimization.

Elements of queueing system

1. Input process – This element of a queueing system is concerned with the pattern in which the customers arrive for service. This can be described by the following three factors:

(a) Size of queue – If the total no of customers requiring service are few, then the size is said to be finite. If the potential customers are sufficiently large, then the source is considered infinite. If more than one arrival is allowed to enter the system simultaneously, the input is said to occur in bulk or patches

(b) Arrival distribution – If the pattern of arrival of customers at the service system is uncertain between successive arrivals (inter - arrival time) . then the arrival pattern is measured by mean arrival time . These are charaterised by the provability distribution associated with this random process most queueing models assume that the arrival rate follows a Poisson distribution and the inter - arrival time follows an exponential distribution .

(c) Customer's behaviour — A customer may decide to wait no matter how long the queue is or he may decide not to enter the system if the queue is too long. If he decides not to wait, he is said to have 'balked' . He may enter the queue but decide to leave after some time. In this case, he is said to have 'reneged' . In case there is more than one queue, he may shift between queues . This is called 'jocky' for position' .

(d) The input process which does not change with time is called a stationary process If it is time dependent, it is called transient.

2. Queue Discipline — This is the rule according to which customers are served after the queue is formed , the most common ~~dispt~~ discipline is FIFO or FCFS . other disciplines include LIFO and SIRO (service in random order) . There are

A variety of priority schemes in which service is done in preference (priority preemptive). Under this case, service is of two types:

(a) <u>pre-emptive</u> – service is provided to customers of high priority

(b) <u>Low pre-emptive</u> – service is provided to customers of low priority

In case of parallel channels. FSR (Fastest server ~~Route~~ Rule) is adopted.

3 <u>service mechanism</u> – The service mechanism is concerned with service time that is the time interval from the commencement of service to completion of service. Service facilities are of the following type:

(a) single queue - single server

(b) single queue - several server

(c) several queue - single server

~~(d) several queue - several server~~

(d) several servers — series channels / parallel channels

4 <u>capacity of system</u>

The source from which the customers are generated can be finite / infinite

Operating characteristics

$E(n) = L$: expected no. of customers in the system.

$E(m) = L_q$: expected no of customers in queue ; $m = n-1$

$E(v) = W$: expected waiting time in the system.

$E(w) = W_q$: expected waiting time in the queue

Server utilisation factor, $\rho = \frac{\lambda}{\mu}$ & where λ = average customers arriving per unit time, μ = average customers completing service per unit time. ρ is also called traffic intensity.

Probability distribution in queueing system

It is assumed that customer joining the system arrive in a random manner and follow a Poisson distribution the service times are mostly assumed to be exponentially distributed. It implies the probability of service completion is independent of length of time spent in progress. We assume the arrival and service distributions follow a Poisson Queue, under the following axioms:

Axiom 1 : The no. of arrivals in non-overlapping time interval are statistically independent

Axiom 2 The probability of more than one arrival within a small time interval. [$t + \Delta t$] is negligible.

$P_0(\Delta t) + P_1(\Delta t) + O(\Delta t) = 1$

~~P_0~~ $P_n(t) = P$ (n arrivals in a time interval, t)

Axiom 3 : The probability that an arrival occurs ~~with~~ within time [$t, t + \Delta t$] is

$P_1(\Delta t) = \lambda \cdot \Delta(t) + O(\Delta t)$

[Derivations - study from book
(a) distribution of arrival
(b) distribution of departure]

(M/M/1) : (∞ / FIFO) // infinite source
(M/M/1) : (N / FIFO) // finite source.

Pg. 592 [21.5 - 1. Deterministic Queuing syste

XX 2 Distribution of Inter - arrival XX

✓ 3. Di " " departure

21 : 7 classification of Queuing models

21 : 9 model 1 - derivation not needed

characteristics of model 1 - formulae

Relationship ~~model II~~

Pg 608 model III - only eq^ns.