Department of AI & DS
CSE and CS&IT

COURSE NAME: PROBABILITY, STATISTICS AND QUEUING THEORY

COURSE CODE: 23MT2005

Topic

Introduction to Queuing theory

Session - 20

AIM OF THE SESSION

To familiarize students with the basic concept of Queuing models

INSTRUCTIONAL OBJECTIVES

This Session is designed to:

1. Define Queuing model
2. Describe the Characteristics of Queuing models and queue discipline
3. Describe the performance measures of queuing theory

LEARNING OUTCOMES

At the end of this session, you should be able to:

1. Define Queue and its characteristics
2. Describe the characteristics of queuing theory and queue discipline
3. Summarize the performance measures

Introduction to Queues

A queue is a group of people, tasks, or objects waiting to be served. Waiting is the essence of queueing.

Queueing models, when solved mathematically or analyzed through simulation, provide the analyst with a powerful tool for designing and evaluating the performance of queueing systems.

Introduction to queues

There are many valuable applications of the queueing theory some examples are traffic flow (vehicles, aircrafts, people, communication), scheduling (patients in hospitals, jobs in machines, personals on a computer), and facility design (banks, post offices, amusement parks, fast-food restaurants).

Introduction to Queues

AIRPLANES

JOBS / PRODUCTS

TRAFFIC

Characteristics of a Queueing system

The best way to understand how a queue operates is to examine the characteristics of the basic queue elements. There are six basic characteristics of a queueing system.

a) **Arrival Process:** The arrival process describe the arrival patterns at the queue. Do customers arrive individually, or do they arrive in groups? Do customers arrive at a fairly constant rate, or is there some pattern to their arrivals? Is arrival process predictable or random?

b) **Service Process:** The service process represents the time taken to serve customers referred to as service time. Is the service time constant, or does it vary from customer to customers? Are customers served in bulk, as in an elevator?

c) **Queue discipline:** The queue discipline specifies the order in which the customers in the queue are served. Are customers served in a first-come, first-served (FCFS) basis, or perhaps in last-come, first –served (LCFS) basis, or service in random order (SIRO)?

Characteristics of a Queueing system

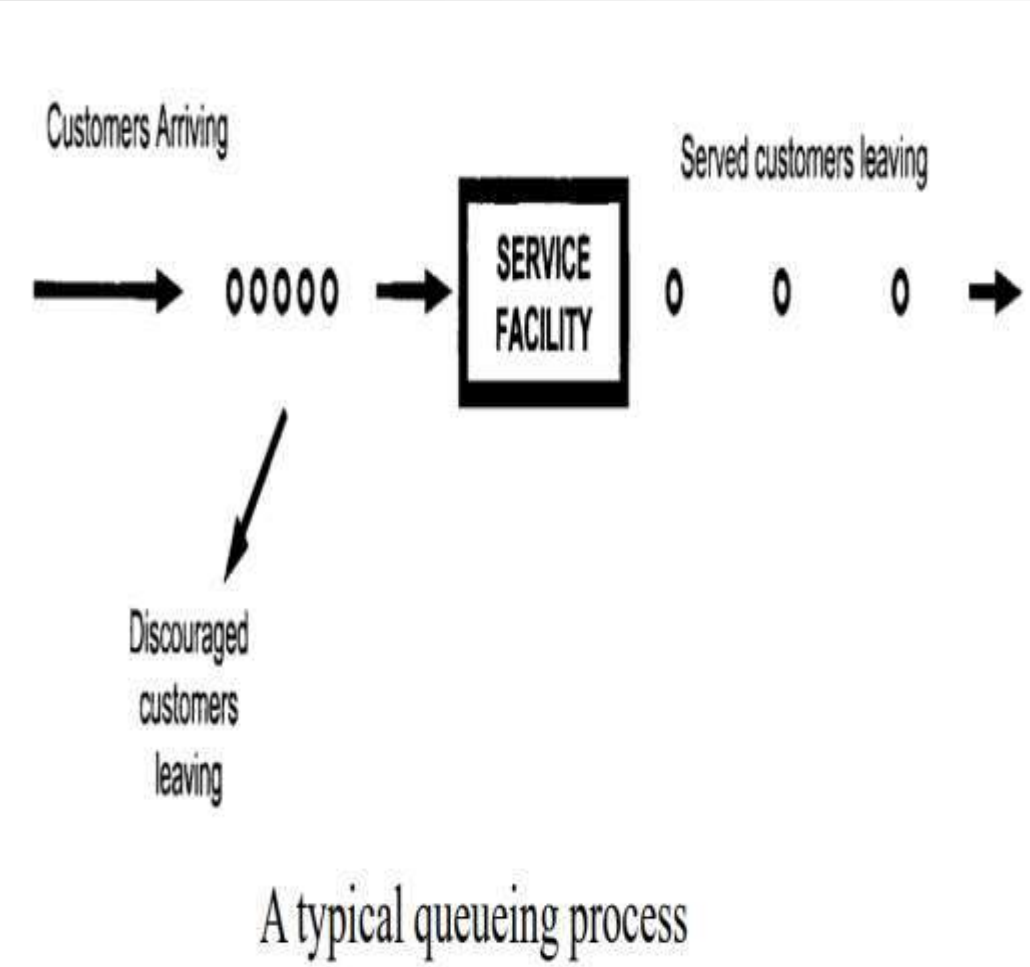
d) No. of servers: Is there a single server or multiple servers. Is there a single queue that feeds all servers, separate queues at each server, or some variation of the two?

e) The calling population: The population of potential customers, referred to as the calling population, may be assumed to be finite or infinite.

In systems with a large population of potential customers, the calling population is usually assumed to be infinite. E.g., customers of a restaurant, bank etc., In a machine shop with 4 machines (machines are treated as customers as they need repair when they breakdown). The calling population is the machines and is finite.

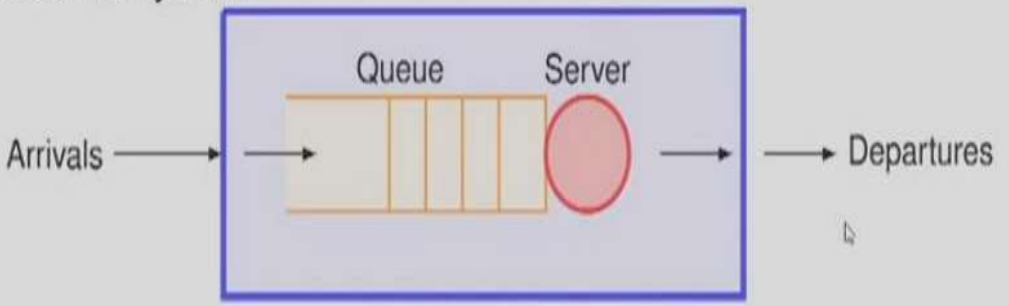
f) System Capacity: In many queuing systems, there is a limit to the number of customers that may be in the waiting lines or system. For example, at a doctor's clinic might have room for only 10 patients to wait (or doctor can give appointment to only 10 patients). An arriving customer finds the system full does not enter for service. Some systems such as banks may be considered as having unlimited capacity.

Characteristics of a Queueing system



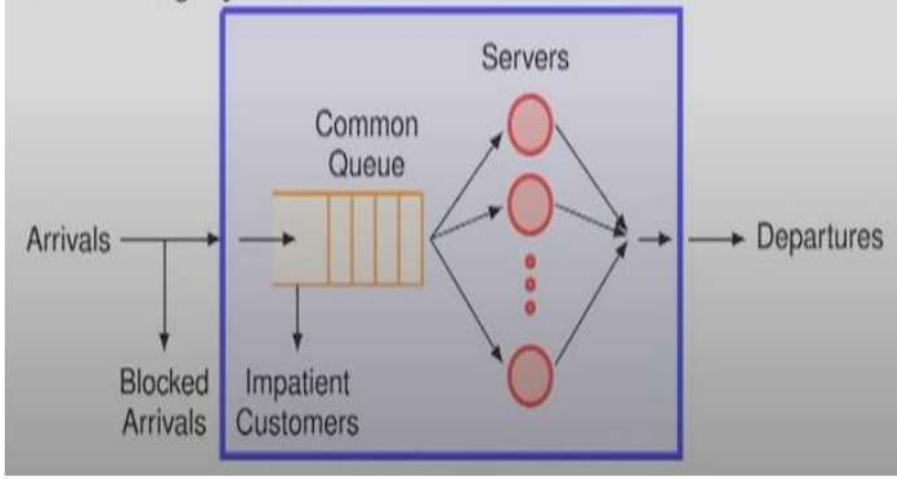
A typical queueing process

Simple Queueing System: Customers arrive, wait for service, receive the service, and then leave the system.



Service Facility

A Queueing System with Some Features:



Measures of performance of a queueing system

Typical measures of performance include server utilization (% of time a server is busy), length of waiting lines, and waiting times of customers. Queueing model is used to predict these measures of system performance as a function of one or more of input parameters.

The input parameters include

- the arrival rate of customers,
- the service demands of customers,
- the rate at which server works,
- and the number and arrangement of customers.

Queueing Notation

Kendall proposed notational system for parallel service system which has been widely adopted. The abridged form of the notation is A/B/C/N/K.

A – Represents the inter arrival time distribution

B – Represents the service time distribution

C – Represents the number of parallel servers

N – Represents the system capacity

K – Represents the size of the calling population

For example, M/M/1/∞/∞indicates a single-server system, that has unlimited queue capacity and an infinite population of potential arrivals. The inter arrival times and service times are exponentially distributed.

M/M/1/∞/∞ is often shortened as M/M/1.

Transient and steady States:

A system is said be in **transient state** when the behavior of the system is dependent on time.

A system is said to be in **steady state** when the behavior of the system is independent of time. In this topic we study only the steady state analysis.

A list of Symbols:

n = number of customers in the queuing system

$P_n(t)$ = steady state probability of having 'n' customers in the system.

P_n = transient state probability that exactly 'n' customers are in the system at time t.

ACTIVITIES/ CASE STUDIES/ IMPORTANT FACTS RELATED TO THE SESSION

λ= Mean arrival rate ($1/\lambda$ is the inter arrival time)

μ=mean service rate ($1/\mu$ is the mean service time)

s= number of parallel service stations

$\rho=\lambda/(\mu s)$=Traffic intensity (or utilization factor) for the service facility, ie., the expected fraction of time server is busy.

L_s=Expected system length, ie., expected member of customers in the system (number of customers waiting in the queue+ number of customers in service.

L_q= Expected queue length, ie., expected number of customers waiting in the queue.

W_s=Expected waitiing time of an arriving customers in the system.

W_q=Expected waiting time of an arriving customer in the queue (Expected waiting time in the system-expected service tune.

(W/W>0)=Expected waiting time of a customer who has to wait.

(L/L>0)=Expected length of a non-empty queue

P(W>0)= Probability of an arriving customer has to wait.

Examples

- .Waiting in line at a bank or a store
- Waiting for a customer service representative to answer a call after the call has been placed on hold
- Waiting for a train to come

SUMMARY

In this session, Queuing models and its performance measures have discussed.

1. Define queuing theory
2. Performance measures of queuing system
3. Difference between Transient state and Steady state.

TERMINAL QUESTIONS

1. Define queue and also write the characteristics of queuing system
2. Mention the performance measures of queuing system.
3. Explain the terms

   1) Balking 2) Renegining 3) Jockeying

REFERENCES FOR FURTHER LEARNING OF THE SESSION

Reference Books:

1. D. Gross, J.F.Shortle, J.M. Thompson, and C.M. Harris, Fundamentals of Queueing Theory, 4th Edition, Wiley, 2008

2. William Feller, An Introduction to Probability Theory and Its Applications: Volime 1, Third Edition, 1968 by John Wiley & Sons, Inc.

Sites and Web links:

1. https://www.khanacademy.org/math/statistics-probability/significance-tests-one-sample/more-significance-testing-videos/v/small-sample-hypothesis-test

2. J.F. Shortle, J.M. Thompson, D. Gross and C.M. Harris, Fundamentals of Queueing Theory, 5th Edition, Wiley, 2018.

3. https://onlinecourses.nptel.ac.in/noc22_ma17/previ ew 3.

4. https://www.youtube.com/watch?v=Wo75G99F9f M&list=PLwdnzlV3ogoX2OHyZz3QbEYFhbqM 7x275&index=3

THANK YOU

Team – PSQT EVEN SEMESTER 2024-25