

# Analytic Modeling

## Queuing Models with Single Service Facility

1

## Characteristics of a Service Facility

- Number of servers
- Queuing discipline
  - default is FCFS
- Number of queuing spaces (or buffers) default is infinite

2

# Terminology

- Entity
  - a component of the system that requires explicit representation in the model
  - e.g., server, customers
- Attribute
  - a property of an entity
  - e.g., capacity of a server, service requirement of a customer
- State
  - a variable that describes the system at a given point in time, based on its entities
  - e.g., number of customers in the system

3

# Customer Arrivals -Infinite Population Model

- Number of users of the service facility is large (potentially infinite)
- Pattern of customer arrivals is based on combined behavior of the customers, and is assumed to be independent of the state of the system

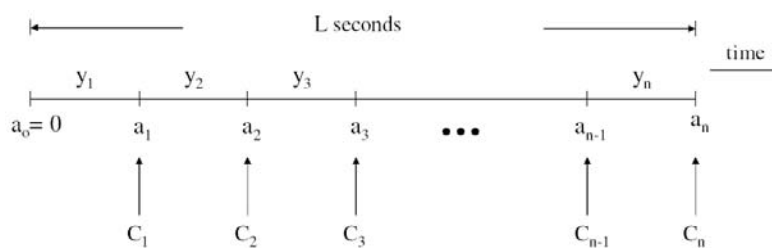
4

## Arrival Definition

- Arrival time
  - time at which a customer arrives at a service facility
- Interarrival time
  - time between two successive arrivals to a service facility
- Arrival rate
  - number of arrivals per unit of time

5

## Timing Diagram



$C_j$  – customer  $j$

$a_j$  – arrival time of  $C_j$

$y_j$  – interarrival time between  $C_{j-1}$  and  $C_j$

6

## Relationship between Arrival Rate and Mean Interarrival Time

○  $L = \sum_{j=1}^n y_j$

○ Mean interarrival time =  $\frac{L}{n}$

○ Arrival rate =  $\frac{n}{L} = \frac{1}{\text{mean interarrival time}}$

7

## Customer Arrivals -Finite Population Model

- Number of users is not large
- The behavior of each user is modeled explicitly as far as arrival pattern is concerned
- Arrival rate is dependent on the state of the system
- Definition
  - think time: elapsed time from completion of previous request to submission of next request

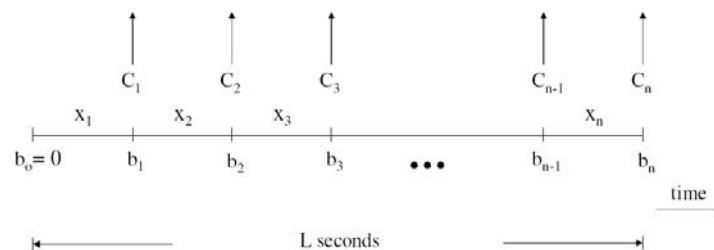
8

# Service Definition

- Service requirement - in unit of work
  - CPU - unit of work is “instruction”
  - communication channel - unit of work is “bit”
- Server capacity - in units of work per second
  - CPU - server capacity is in “number of instructions executed per second”
  - communication channel - server capacity is in “number of bits transmitted per second”
- Service time =  $\frac{\text{service requirement}}{\text{server capacity}}$
- Service rate
  - Number of customers served per second (assuming no idle time)

9

# Timing Diagram



- $C_j$  - customer  $j$
- $x_j$  - service time of  $C_j$
- $b_{j-1}$  - time at which  $C_j$  starts service

10

## Relationship between Service Rate and Mean Service Time

○  $L = \sum_{j=1}^n x_j$

○ Mean service time =  $\frac{L}{n}$

○ Service rate =  $\frac{n}{L} = \frac{1}{\text{mean service time}}$

11

## Performance Measures

- Utilization factor
  - percentage of time that server is busy
- Response time
  - elapsed time from arrival to departure
- Waiting time
  - time spent in queue
- Number of customers in system
- Number of customers in queue
- Throughput
  - rate at which customers leave a service facility after completing services

12