Systems Simulation

Prepared By: Dr. Ahmad S. Tarawneh

Updated By: Dr. Saed Alqaraleh

Outline

- Introduction
- What are queuing systems
- Single-server queuing system
  - Example
- Muti-server queuing system
  - Example

Simulation Examples

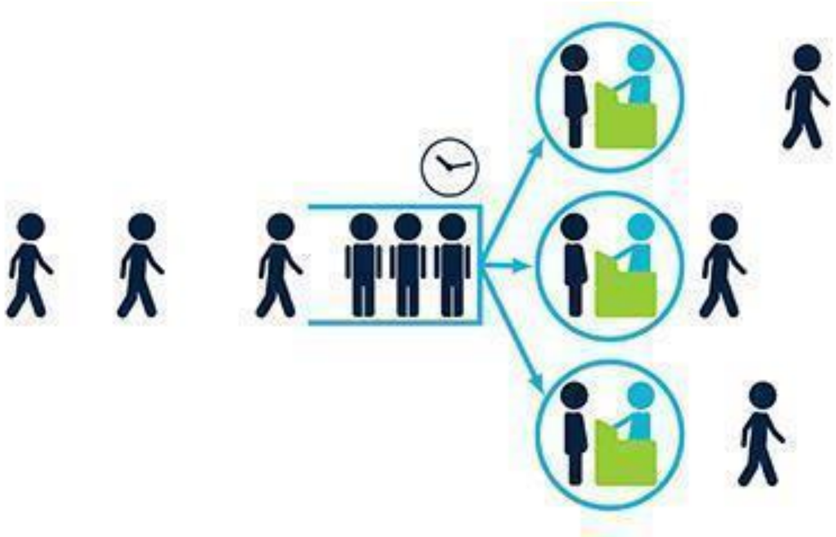
- Three steps of the simulations
  - Determine the characteristics of each of the inputs to the simulation. Quite often, these may be modelled as **probability distributions**, either continuous or discrete.
  - **Construct a simulation table**. Each simulation table is different and developed specifically for the problem at hand.
  - For each repetition *i*, generate a value for each of the p inputs and evaluate the function, calculating a value of the response y_i. The input values may be computed by sampling values from the distributions determined in step 1. A response typically depends on the inputs and one or more previous responses.

Simulation Table

- The simulation table provides a systematic method for tracking system state over time.

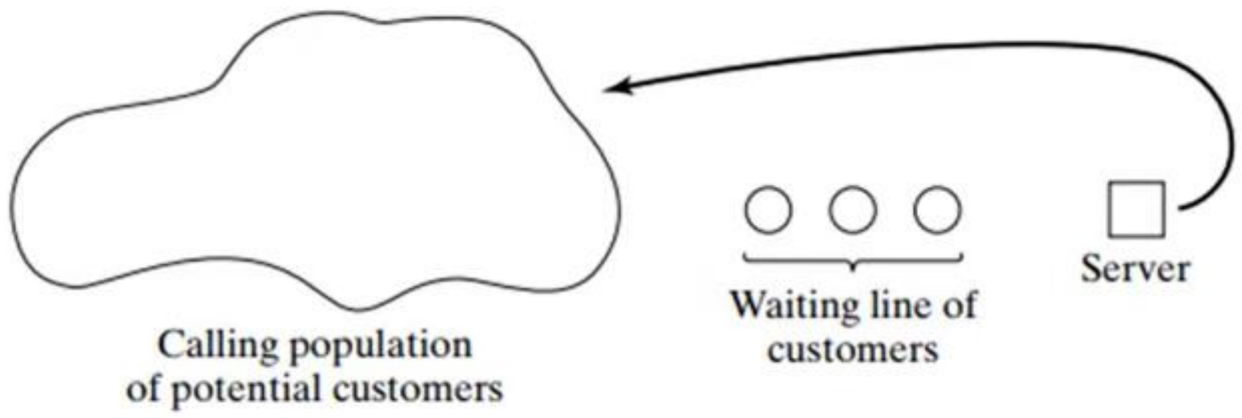
	Inputs						Response
Repetitions	X_{i1}	X_{i2}	…	X_{ij}	…	X_{ip}	y_i
1							
2							
.							
.							
.							
n							

Queuing Systems Simulation



Intro. Queuing systems (QS)

- One of the common systems to simulate are queuing systems.
- **Typically, queuing systems contain customers who arrive from time to time, join a queue, are served and then leave the system.**



Cont.

- **The main goal of such systems is to help decision-makers improve their actual system based on the simulation outputs**
- Queuing simulation systems can provide several performance measurements, such as:
    1. **Server utilization**
    2. **Number of waiting customers in the queue (line)**
    3. **Delay of the customers inside the system**
- **In general, a good system is a system that provides a good tradeoff between server utilization and customer satisfaction**
    - **Large server utilization** => happy customers **BUT** sad employees (**fast failure of machines**, etc.)
    - **Low server utilization** => happy employees (long-live machines, etc.) **BUT** angry customers

Description of QS

- The main elements in any queuing system are
  - Server(s)
  - Customers
- Customers can be any entity that needs a service, including machines, cars, people, or orders in a restaurant; the entity that arrives at a system and waits to be served
- Server(s) on the other hand, can be any entity that provides the required service, such as tellers, machines, the guy who makes sandwiches, etc.

Cont.

- Such a system has several components:

1. The Calling Population
2. The System Capacity
3. The Arrival Process
4. Queue Behavior and Queue Discipline
5. Service Time and Service mechanism

The Calling Population

Calling Population refers to the group of individuals or entities that can potentially request a service.

- This refers to the **source of the requests or customers** that enter the system seeking service.
- Characteristics of the calling population include size (finite or infinite), behavior (e.g., arrival patterns), and requirements (type of service needed).
- **Understanding the calling population is crucial for designing an efficient queue system that meets customer needs without excessive cost or complexity**

Cont.

- **A finite population** refers to a system where the number of potential customers or entities that can request service is limited.
- **This limitation means that the probability of a new arrival changes as the number of entities currently being served changes**
- For example, in a maintenance system for a fleet of 50 trucks, if one truck is being serviced, the population size effectively becomes 49.
- Finite populations are often seen in closed systems, where the pool of requestors is known and limited.
- **Modeling often requires more complex calculations** due to the changing arrival rates

Cont.

- **An infinite population** refers to a model assumption where the number of potential customers or entities is considered unlimited.
- In such models, the arrival of one entity does not significantly change the probability of the next arrival.
- This assumption simplifies the analysis and modeling of the queue system because the arrival process can often be modeled using a single distribution, assuming a constant average rate of arrivals over time.
- **Simplifies mathematical modeling** of the queue system, making it easier to predict system behavior and performance under various conditions

Cont.

- **The main difference between finite and infinite population models is how the arrival rate is defined.**
- In an ***infinite*** population model, the arrival rate (i.e., the average number of arrivals per unit of time) is not affected by the number of customers who have left the calling population and joined the queueing system.

Finite vs. Infinite Calling Populations

Characteristic	Finite Calling Population	Infinite Calling Population
Size	Limited number of potential requests	Essentially limitless number of potential requests
Examples	A small town with a single doctor, a factory with limited machines, a call centre handling a specific campaign	Website selling a popular product, toll-free customer service line, public library
Impact of arrival	Arrival of one request significantly reduces the number of remaining potential requests	Arrival of one request has minimal impact on the number of remaining potential requests
Additional examples	Students in a classroom, cars waiting at a toll booth, patients in a doctor's office	Raindrops in a storm, phone calls to a hotline, social media followers

The System Capacity

- System capacity pertains to the **maximum number of customers or requests the system can handle at one time.**
- This includes
  - The service channels (e.g., Tellers, servers, lanes)
  - The space available for waiting.

The Arrival Process

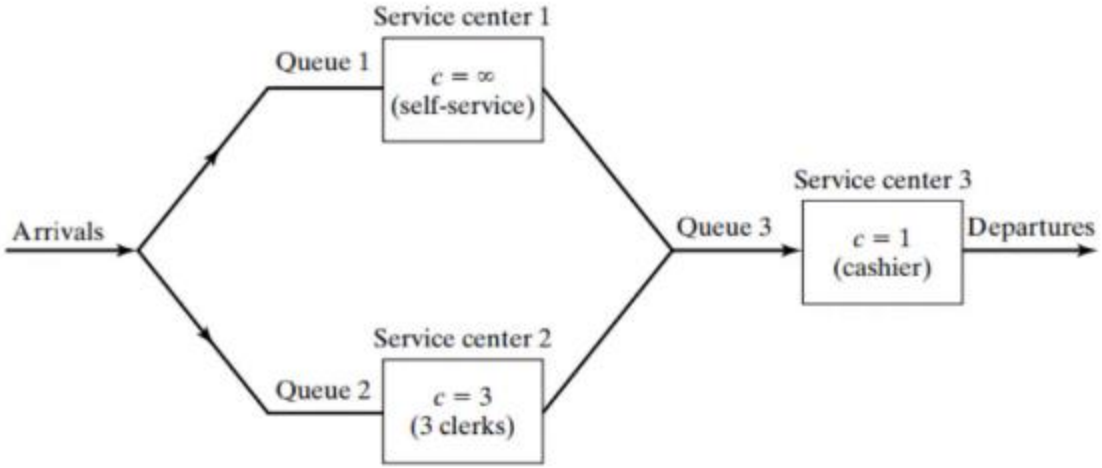
- This involves the pattern by which customers or requests arrive at the system.
- The arrival process can be deterministic (arrivals at regular intervals) or stochastic (arrivals at random intervals).
- If the arrival at random, it is modeled using a probability distribution.

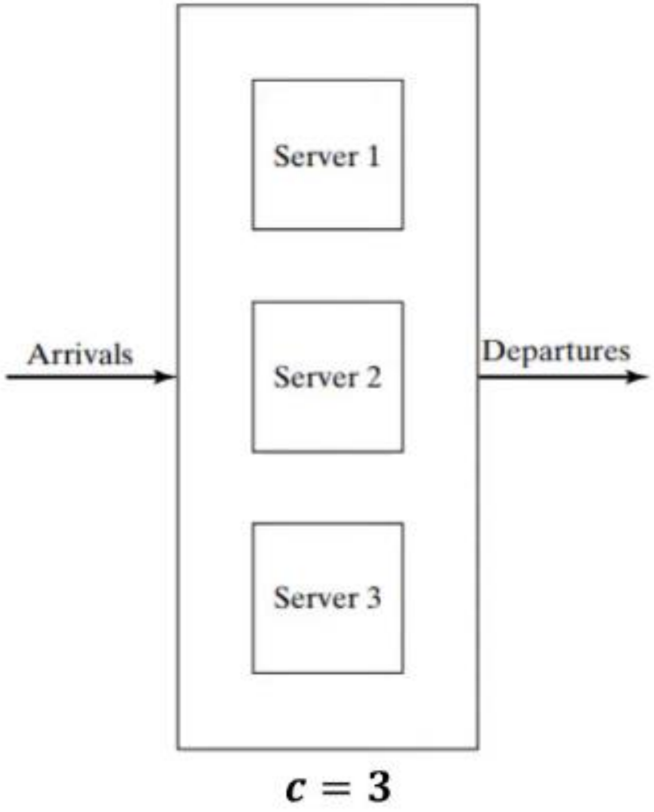
Queue Behavior and Queue Discipline

- **Queue behavior** refers to how customers behave while waiting in the queue, including:
  - They are willing to wait
  - They might leave the queue (balk, bored, jockeying)
  - Join and leave without being served (renege, changed his mind).
- **Queue discipline** is the order in which customers are served (called for service), such as
  - First-in-first-out (FIFO)
  - Last-in-first-out (LIFO)
  - Based on priorities (e.g., emergency cases first),
  - Short processes first

Service Time and Service mechanism

- Service time is the amount of time required to serve a customer, which can vary widely depending on the type of service and other factors.
- The service mechanism entails how services are delivered, is it by one server or parallel servers? Is it single phase or multiphase?





Characteristics

- The queuing system's characteristics can be denoted using the notation **A/B/c/N/K**

A represents the interarrival-time distribution

B represents the service-time distribution

c represents the number of parallel servers

N represents the system capacity

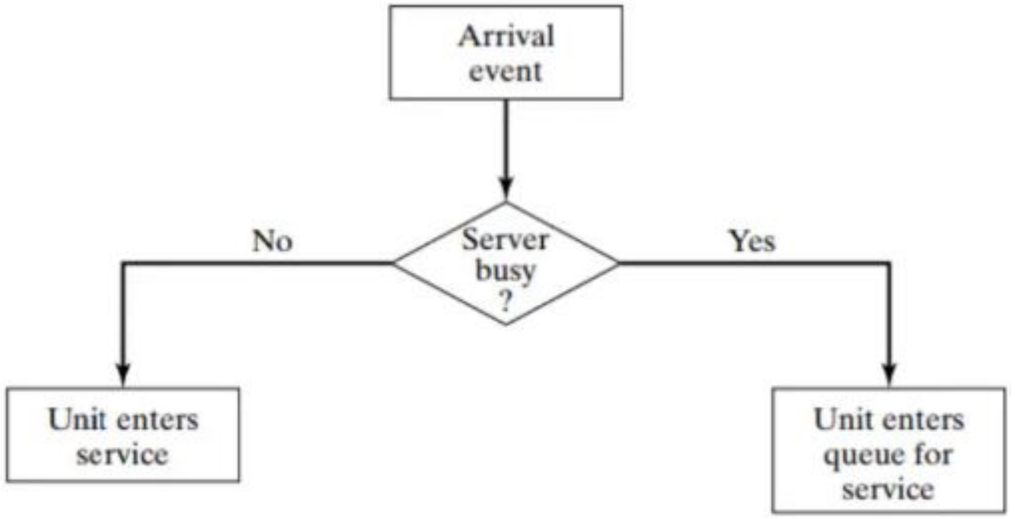
K represents the size of the calling population

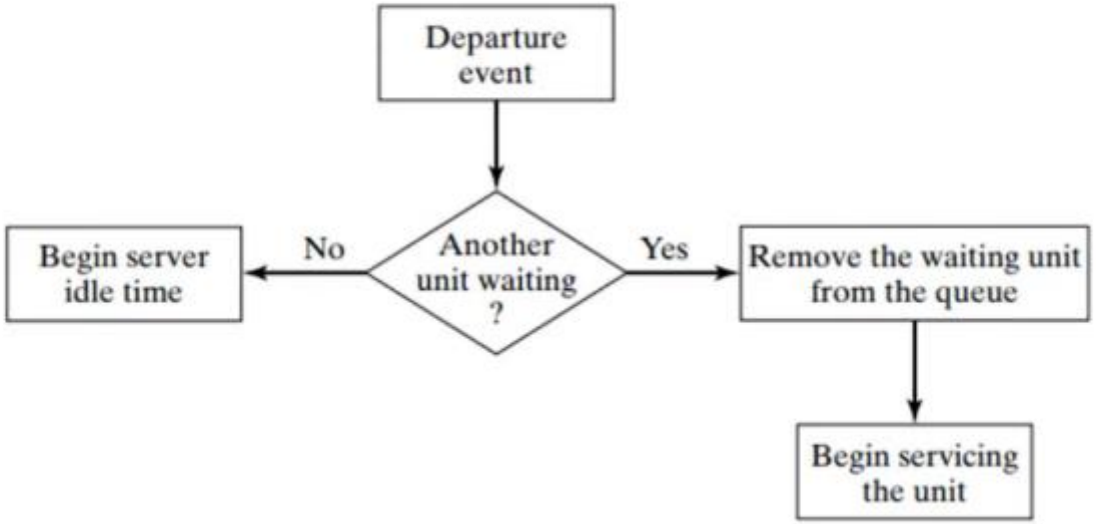
Queuing system state

- The state of the system is the quantities that completely describe the system at any point of the simulation, considering the study objectives:
  - **The number of units in the system: the number of units waiting in the line and the units being served.**
  - **Status of the server(s): is the server busy serving a customer or idle waiting for a customer**

Cont.

- The state of the system is changing based on the arrivals and departures

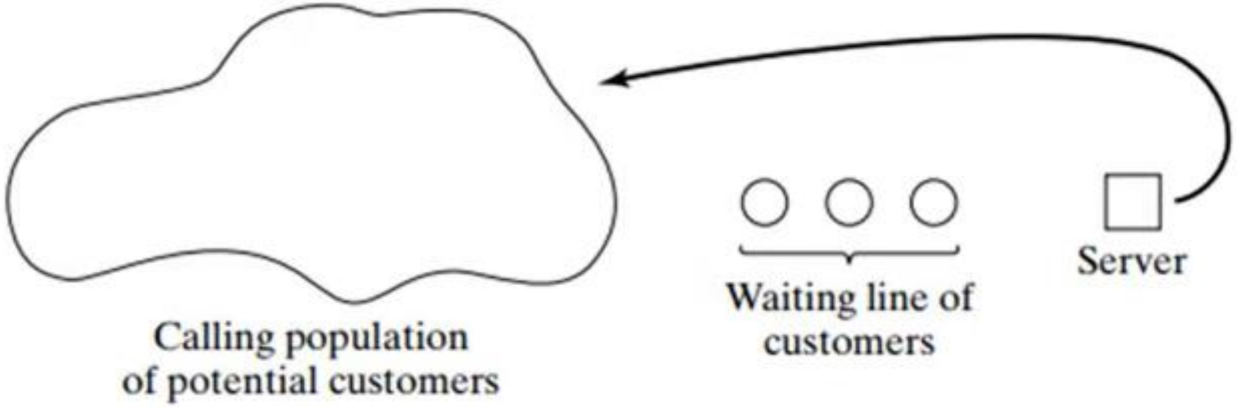




Simulation Tables

- A simulation table is a structured representation used to model and analyze the behavior of a system during the simulation.
  - especially when dealing with stochastic (random) processes.

- The next table shows a simulation system that has single server and serves customers using FIFO.



Using the following info, simulate the university cafeteria, which has only one employee, i.e., one queue system. Find the average waiting time of the customers

Interarrival time(IAT): -, 5, 1, 6, 8, 5, 10

Service Time (ST): 5, 10, 3, 8, 5, 10, 2

Customer #	Interarrival time	Arrival time (AT)	Service Time (AT)	Time Service Begins (TSB)	Time Service Ends(TSE)	Waiting Time(WT)	idle Time(IT)
1	-	0	5	0	5	0	0
2	5	5	10	5	15	0	0
3	1	6	3	15	18	15-6=9	0
4	6	12	8	18	26	18-12=6	0
5	8	20	5	26	31	26-20=6	0
6	5	25	10	31	41	31-25=6	0
7	10	35	2	41	43	41-35=6	0

Calculate the server (employee) utilization. What decisions (suggestions) can you make based on this information?

A simulation system that has a single server and serves customers using FIFO.

The interarrival and service times are taken from distributions!

Customer	Interarrival Time	Arrival Time on Clock	Service Time
1	-	0	2
2	2	2	1
3	4	6	3
4	1	7	2
5	2	9	1
6	6	15	4

The simulation run is build by meshing clock, arrival, and service times!

Customer Number	Arrival Time [Clock]	Time Service Begins [Clock]	Service Time [Duration]	Time Service Ends [Clock]	Waiting Time(WT)	idle Time(IT)
1	0	0	2	2		
2	2	2	1	3		
3	6	6	3	9		
4	7	9	2	11		
5	9	11	1	12		
6	15	15	4	19		

Cont.

- Different visualizations can be used to show the performance of the system

Chronological ordering of events

Clock Time	Customer Number	Event Type	Number of customers
0	1	Arrival	1
2	1	Departure	0
2	2	Arrival	1
3	2	Departure	0
6	3	Arrival	1
7	4	Arrival	2
9	3	Departure	1
9	5	Arrival	2
11	4	Departure	1
12	5	Departure	0
15	6	Arrival	1
19	6	Departure	0

Interesting observations
- Customer 1 is in the system at time 0
- Sometimes, there are no customers
- Sometimes, there are two customers
- Several events may occur at the same time