香港城市大學
City University of Hong Kong

396EM Airline Operations and Scheduling/ 6075MAA Airline Scheduling and Operations

Lecture 5a
Fundamental Simulation Concepts Queuing Models

Developed & Revised :
Dr Yang Dai & Dr Ammar Bazi (Coventry)
Kimman Lui & Rossella Lau (Scope)
Presented: Kinki Leung

專業 創新 胸懷全球
Professional · Creative
For The World

SCOPE
School of Continuing and Professional Education

Coventry University

Outline of the Lecture

- Airline operation simulation
  - Fundamental simulation concepts
    - Simulation definition
    - Systems and operations systems
    - Pros and cons
  - Conceptual modelling
  - System analysis model

- Queueing models
  - Introduction to queueing
  - Queue modelling
  - Queuing model

Learning Outcome

Understand the concepts of simulation

Understand the system concept, its components and its modelling

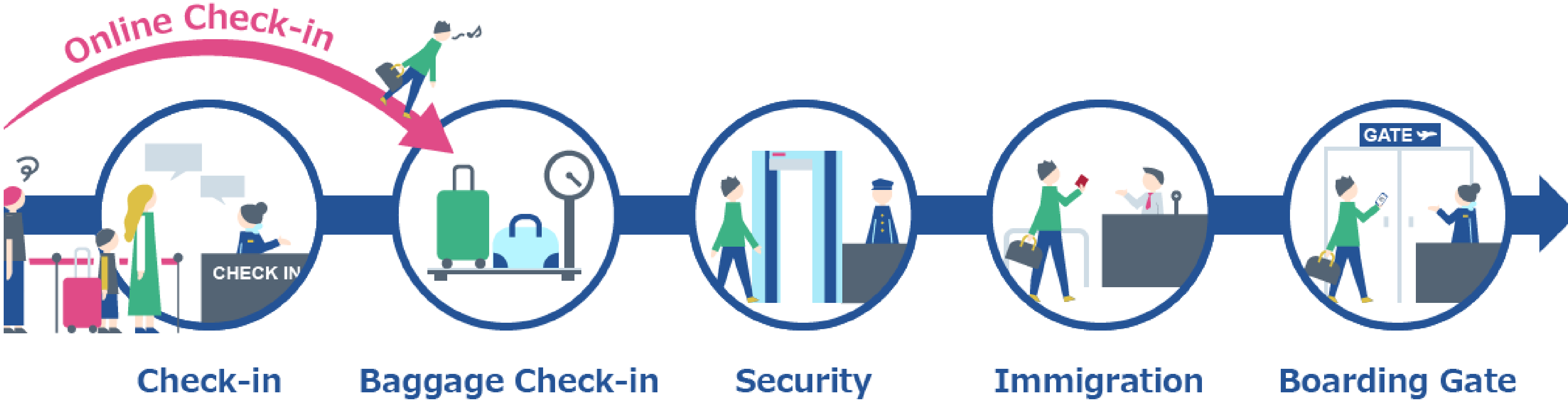
Understand the basics of queue and queuing models

Airport/Airlines Management

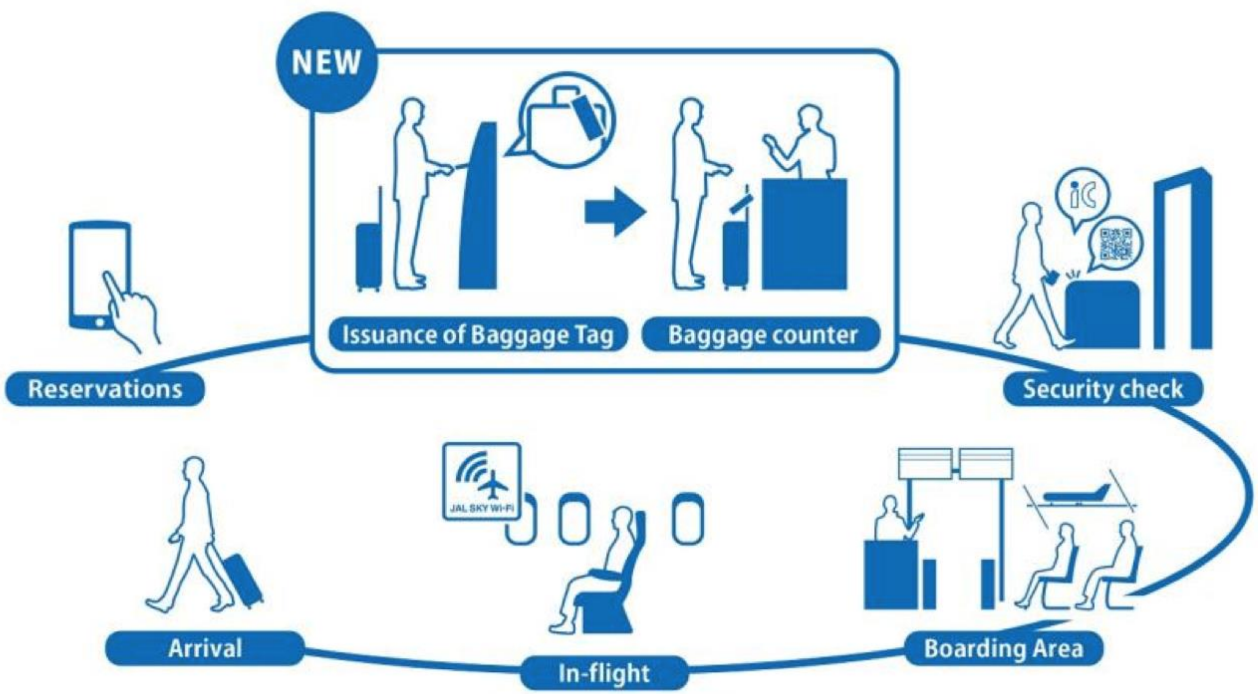
- The management of an airport are planning the facilities that are required in a new terminal building.

- Important decisions need to be made:
  - Number of staff to employ and the shifts they should work need to be determined
  - Number of check-in desks devoted to each airline
  - Size of the baggage handling system
  - Amount of security check positions
  - Number of departure gates.

HKIA Map



Boarding Procedures (sourced from [ANA check-in guide]())



JAL streamline check-in process with self-tagging service (sourced from JAL Flyer)

Determine the number of resources

- One approach would be to build the terminal and hope that it works! This seems very risky with so much at stake.

- A much more effective approach is likely to be a simulation of the proposed airport terminal.
  - For example: Simulate the flow of passengers and their bags through each of the key stages from arrival to departure

What is Simulation?

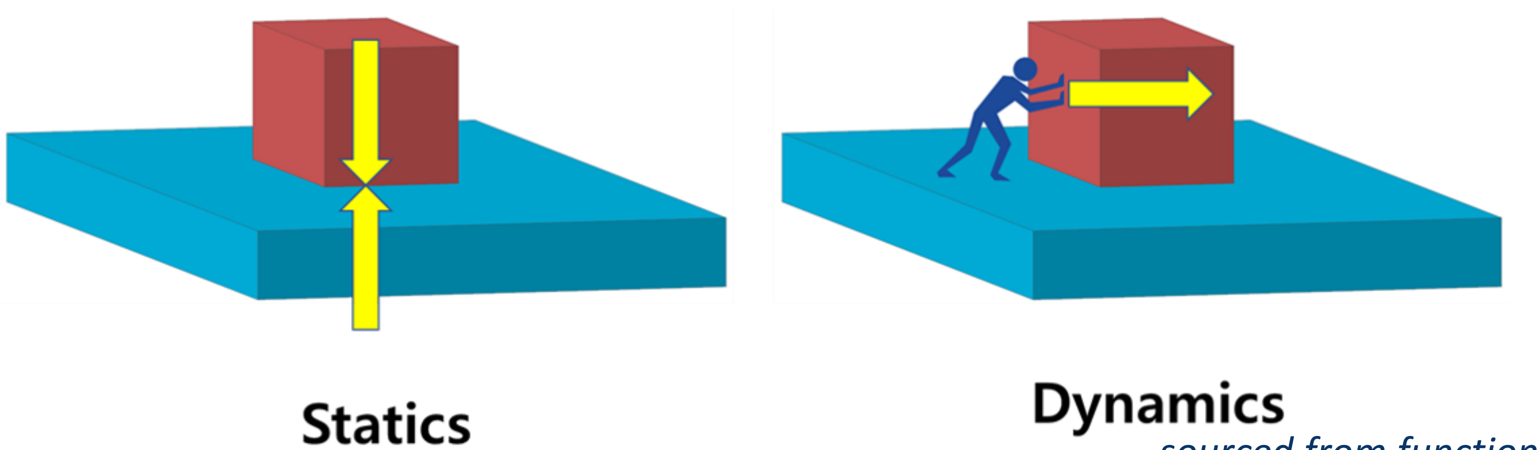
Quick answer:

Static Simulation

- imitation of a system at a point in time.

Dynamic Simulation

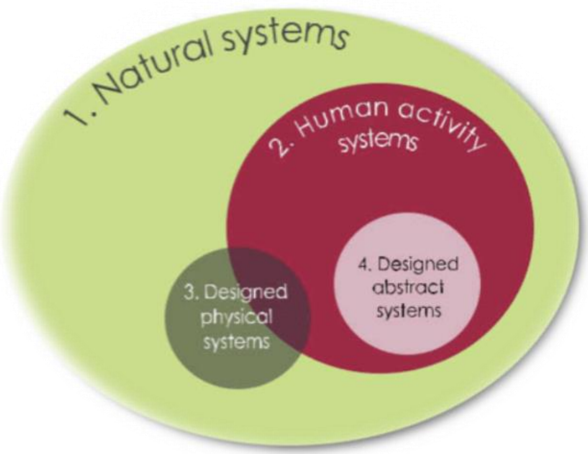
- imitation (on a computer) of a system as it progresses through time.



sourced from functionbay.com

Four main classes of system (Checkland 1981)

- Natural systems: systems whose origins lie in the origins of the universe, e.g. the atom, the Earth's weather system, biological system, disease.
- Designed physical systems: physical systems that are a result of human design, e.g. a house, a car and a production facility.
- Designed abstract systems: abstract systems that are a result of human design, e.g. mathematics, literature stimulatory models.
- Human activity systems: systems of human activity that are consciously, or unconsciously, ordered, e.g. a family, a city, work organizations, department, committee.



sourced from [ResearchGate]

Operations Systems

- When describing and understanding airport and airlines two classes of systems are of prime concern, that is, designed physical and human activity systems

- For instance, a simulation might be developed to intimate check-in processes of an airline (a human activity system) with an automated baggage handling facility (a designed physical system)

- System with operations involving elements of both classes of systems can be referred to as operations systems or operating systems.

Operation simulation

- To simulate the check-in operations in order to understand and to make decisions, e.g., number of counters required

- To achieve the purpose, we need a way to describe the system (modelling)

- The imitation should be easily understood (simple) and can be conducted for "As is" and "What if" different situations (experimentation)

Definition of Simulation

- Experimentation with a simplified imitation (on a computer) of an operations system as it progresses through time, for the purpose of better understanding and/or improving that system.

Nature of operations systems

1 variability 2 interconnectedness 3 complexity

Variability

- Many operations systems are subject to variability. This might be predictable variations, for instance, changing the number of operators in a call centre during the day to meet changing call volumes or planned stoppages in a production facility.

- It might also be variations that are unpredictable, such as the arrival rate of patients at a hospital emergency department or the breakdown of equipment in a flexible manufacturing cell.

- Both forms of variability are present in most operations systems.

Nature of operations systems

- Interconnectedness

  Operations systems are also interconnected.

  Components of the system do not work in isolation, but affect one another.

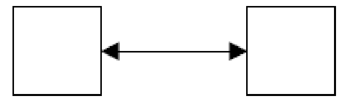
  - A change in one part of a system leads to a change in another part of the system.

- Complexity

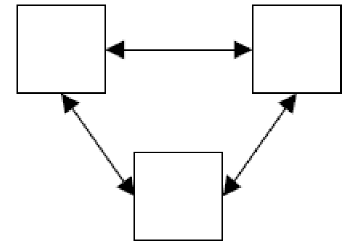
  Many operations systems are also complex. It is difficult to provide an exact definition of the word complexity.

Example

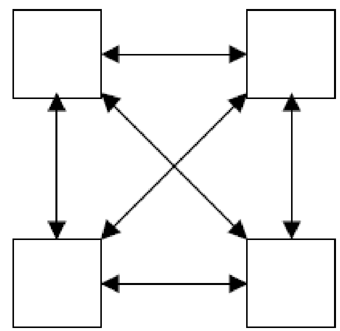
- Job Shop Systems: Interconnections and Combinatorial Complexity



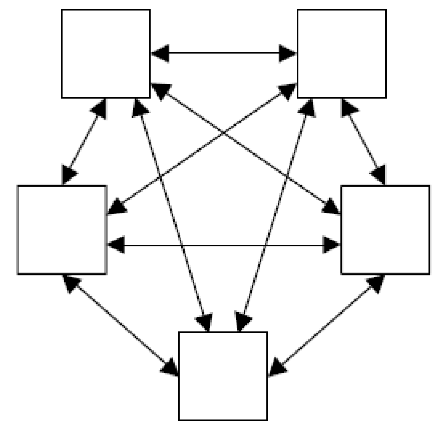
2 interconnections



6 interconnections



12 interconnections



20 interconnections

Software for simulation

- Simulators (for work, for operation…)

- Discrete-event simulation software
  - Being VIS (Visual Interactive Simulation)
  - Software
    - Spreadsheets
    - Programming Languages
    - Specialist Simulation Software: Arena, SIMUL8, FlexSim, ……

  - SIMUL8 Examples: Call Center, Healthcare, Service

Advantages of simulation – I

- Simulation versus experimentation with the real system

- Rather than develop and use a simulation model, experiments could be carried out in the real system.

- Cost: Experimentation with the real system is likely to be costly. It is expensive to interrupt day-to-day operations in order to try out new ideas.

- Time: It is time consuming to experiment with a real system. It may takes many weeks or months before a true reflection of the performance of the system can be obtained.

Advantages of simulation – II

- Control of the experimental conditions. When comparing alternatives, it is useful to control the conditions under which the experiments are performed so direct comparisons can be made. This is difficult when experimenting with the real system.

- It is also likely that experimentation with the real system will lead to the Hawthorne effect, where staff performance improves simply because some attention is being paid to them.

- The real system does not exist. A most obvious difficulty with real world experimentation is that the real system may not yet exist.

Disadvantages of simulation – I

- Expensive
  - Simulation software is not necessarily cheap
  - Cost of model development and use may be considerable, particularly if consultants have to be employed.

- Time consuming
  - Simulation is a time-consuming approach and adds to the cost of its use and means that the benefits are not immediate.

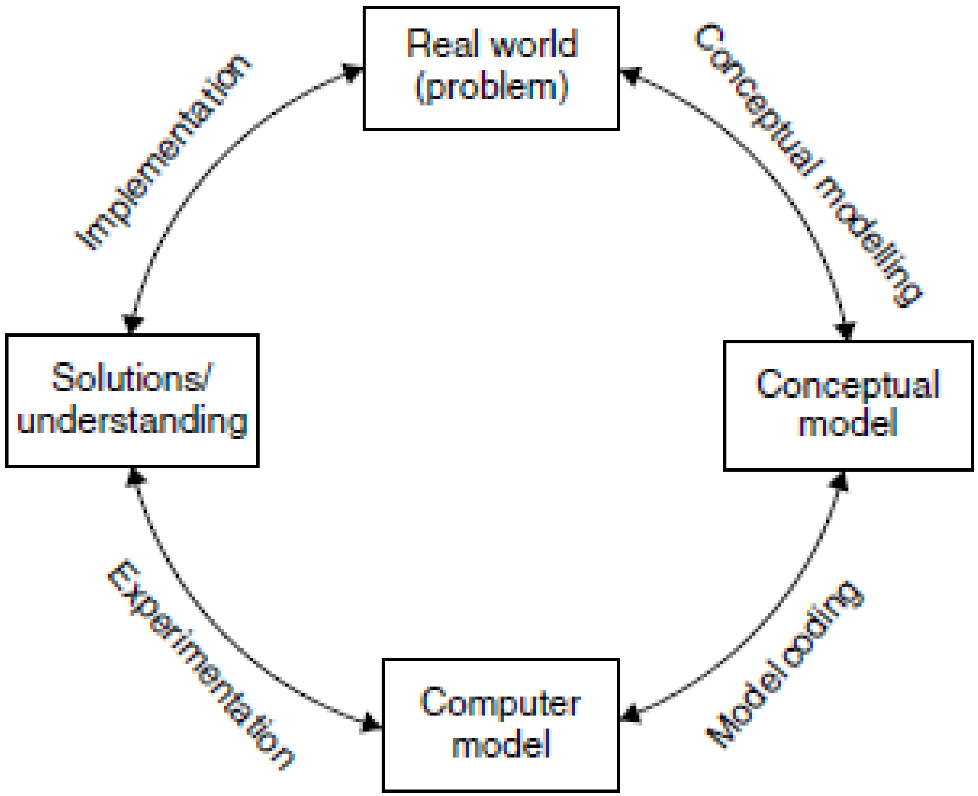
Disadvantages of simulation – II

- Data hungry
  - Require a significant amount of data.
  - not always immediately available and, where it is, much analysis may be required to put it in a form suitable for the simulation.

- Requires expertise
  - Require skills in, among other things, conceptual modelling, validation and statistics, as well as skills in working with people and project management.

Simulation Studies: Key stages and processes

- Conceptual model – description of the model that is to be developed
- Computer model – simulation model implemented on a computer
- Solutions/understanding – derived from the results of the experimentation
- Improvement in the real world – obtained from implementing the solutions and or understanding gained



Concept of modelling

- System
  - Is defined to be a collection of entities, e.g., people or machines, which act and interact together toward the accomplishment of some logical end.
  - Examples are: Airport operations (passengers, security, planes, crews, baggage), Transportation/logistics/distribution operations.

- Model
  - A simpler representation of a system that allows for investigation of the properties of the system and, in some cases, prediction of future outcomes

- Modelling
  - Refers to the process of generating a model as a conceptual representation of some phenomenon.

System Modelling Elements

- Entity
  - An object of interest in the system, e.g. passengers at an airport.
- Attribute
  - A property of an entity, e.g. flight number, ticket type (economy, business).

System Modelling Elements (Con't)

Activity

- Represents a time period of specified length.
- Collection of operations that transform the state of an entity, e.g. check in operation.

Event

- Change in the system state.
- e.g., arrival; departure of passengers, busy; idle status of immigration officers.

System Modelling Elements (Con't)

State Variables

- A collection of variables that are used to describe the performance of a system overtime.
- e.g., A number of passengers (Queue Length) waiting in front of an airline desk. The average waiting time of these passengers (Queue Time)

Note:

- the three concepts(event, process, and activity) give rise to three alternative ways of building discrete event simulation models

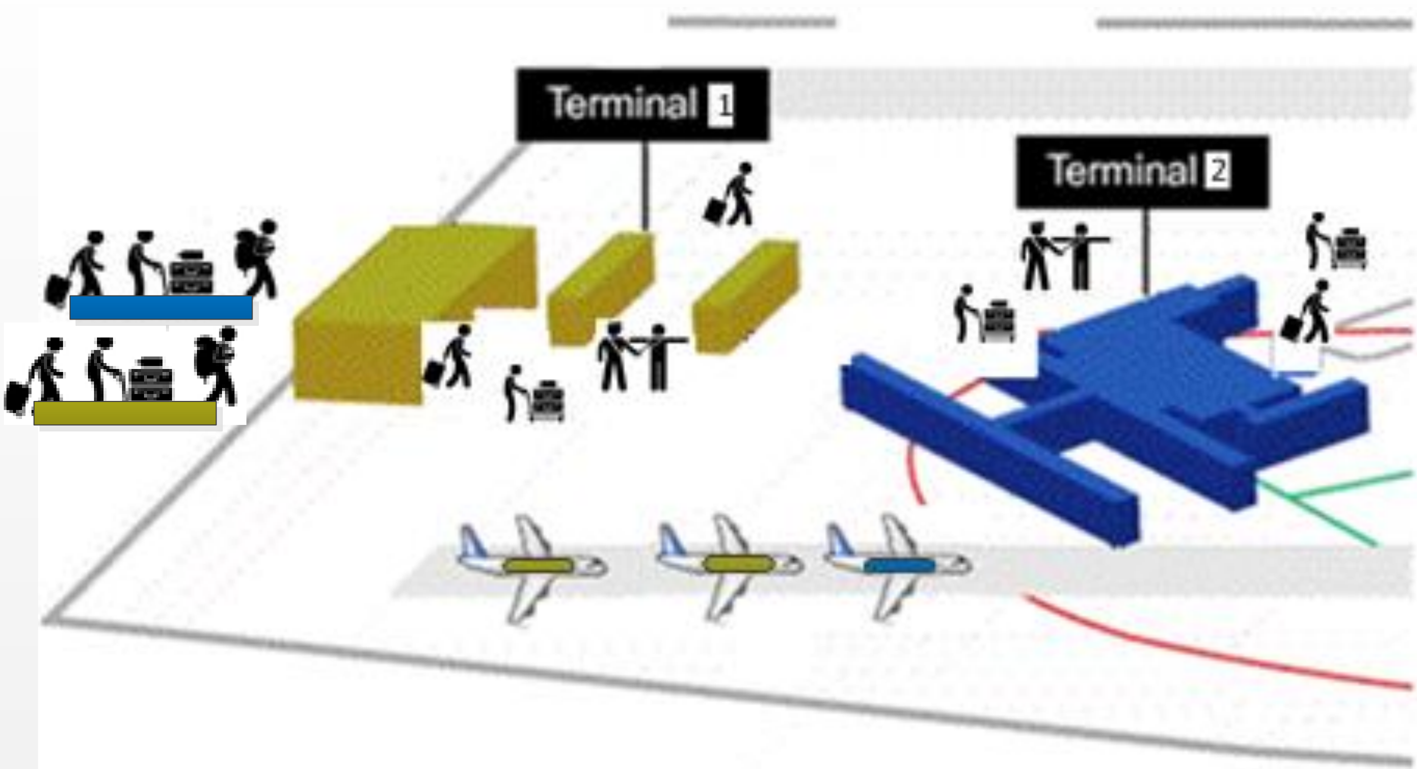
Basics of Selecting Entities

- Should first identify relevant entities in the system
  - Passengers in a terminal
  - Airplanes in an airport
  - Customers and clerks in post office model
- Should distinguish between
  - Permanent entities (Resources)
  - Temporary entities can be created and disappear after getting the service.

Basics of Selecting Attributes

- Entities can have attributes
  - Post office customer could have
    - customer ID,
    - time spent in queue,
    - type of requested service,
    - time spent being served, …
  - Post office waiting line has
    - mean waiting time,
    - maximum waiting time, …

Visual Example of Airport System Modelling Elements



Airport

System Analysis Model

System	Entities	Attributes	Activities	Event	State Variables
Airport	*Passengers (Temporary)*	Passenger booking confirmation Ticket Number Destination Booking Type (Economy, Business)		Arrival, Departure	Queue Length (number of waiting passengers) Average Waiting Time of Passengers
	Resources (Permeant)	Terminal Number Immigration officers Airplane types and capacities Luggage Carousels	Passenger Services Check in operations Guidance and other assistance	Idle, Busy	Utilisation Service efficiency Operating Costs Delay

Queuing

Queueing theory: Little's Law

- The long-term average number of customers in a stable system L is equal to
  - the long-term average effective arrival rate, **λ**, multiplied by
  - the average time a customer spends in the system**,** W;
  - or expressed algebraically:

$$\mathbf{L = \lambda W}.$$

- Although it looks intuitively reasonable, it is quite a remarkable result, as the relationship is

"not influenced by the arrival process distribution, the service distribution, the service order, or practically anything else"

Simchi-Levi, D.; Trick, M. A. (2013)

Explanation of Little's Law

$$\text{Minutes until I get coffee} = \frac{\text{Number of people in line}}{\text{people served per minute}}$$

$$\frac{\text{11 in line}}{\text{1 per minute}} = \text{11 minutes to get coffee}$$

sourced from slideshare.net

Queueing Models: Introduction

- Queues arise when the short term demand for service exceeds the capacity.
- Queue is a line of waiting customers who require service from one or more service providers (servers).
- Queuing models are widely used to estimate desired performance measures of the system.
- It provides rough estimate of a performance measure!

Modelling the progression of time

- Discrete-Event Simulation Approach
  - Basic events
  - Customer arrives
  - Server (operator) starts service
  - Server completes service
  - Customer in queue – no server available

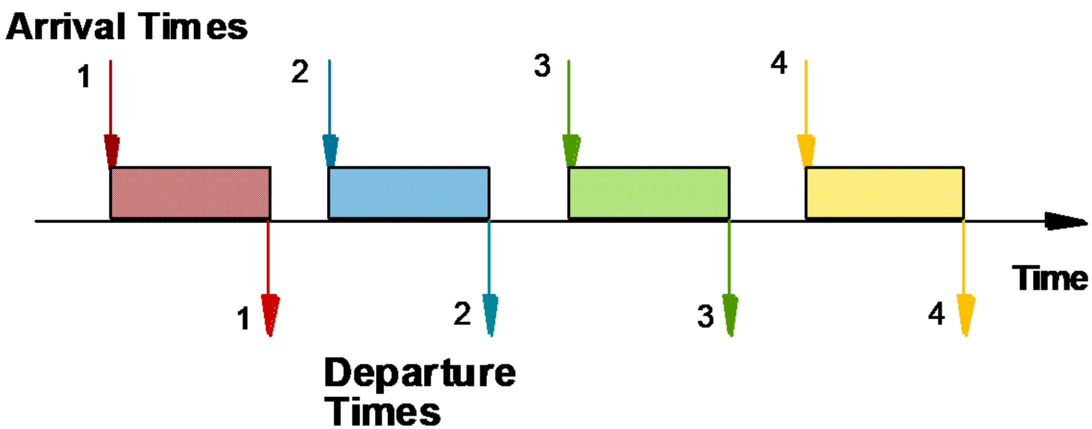
Importance of Queue Modelling

- Capacity problems are very common in industry and one of the main drivers of process redesign and improvement.

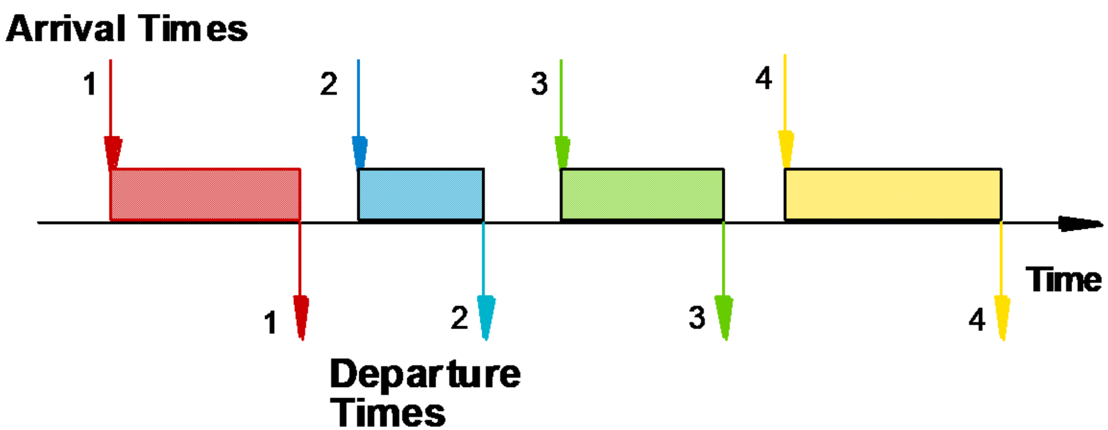
- Queuing and waiting time analysis is particularly important in service systems.

Delay is Caused by Packet Interference

- If arrivals are regular or sufficiently spaced apart, no queuing delay occurs



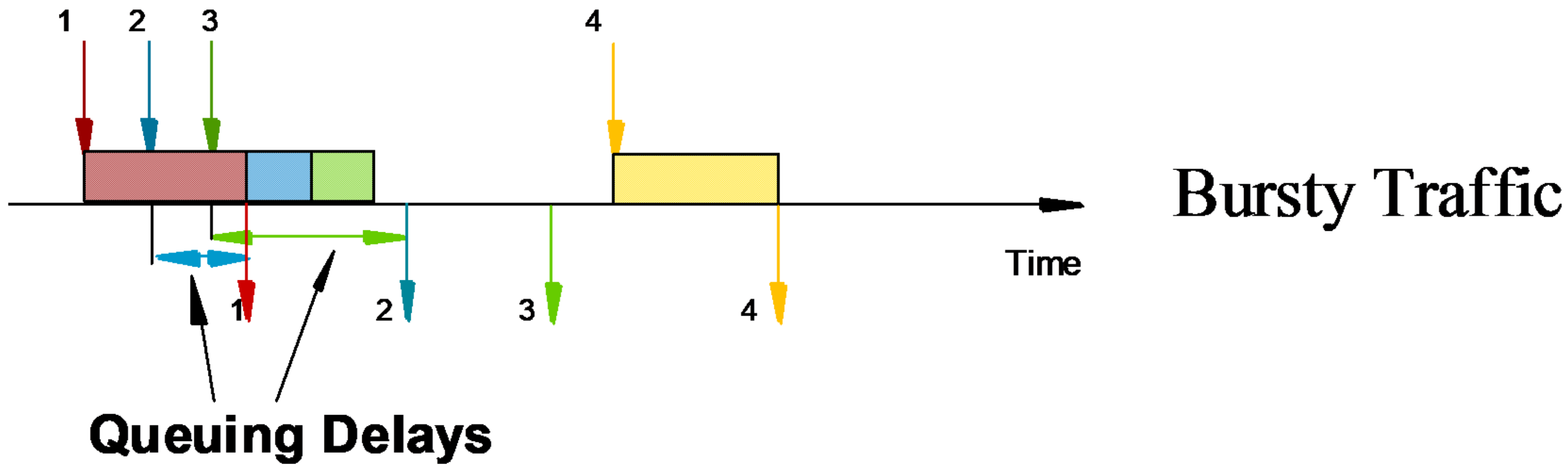
Regular Traffic



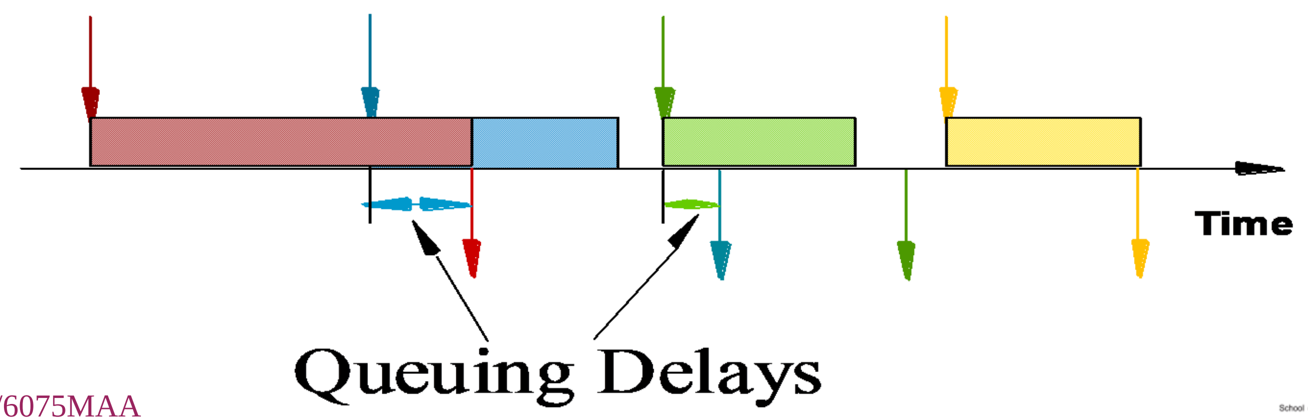
Irregular but Spaced Apart Traffic

Burstiness and Packet length Causes Interference

- Note that the departures are less bursty



- Regular arrivals but irregular packet lengths

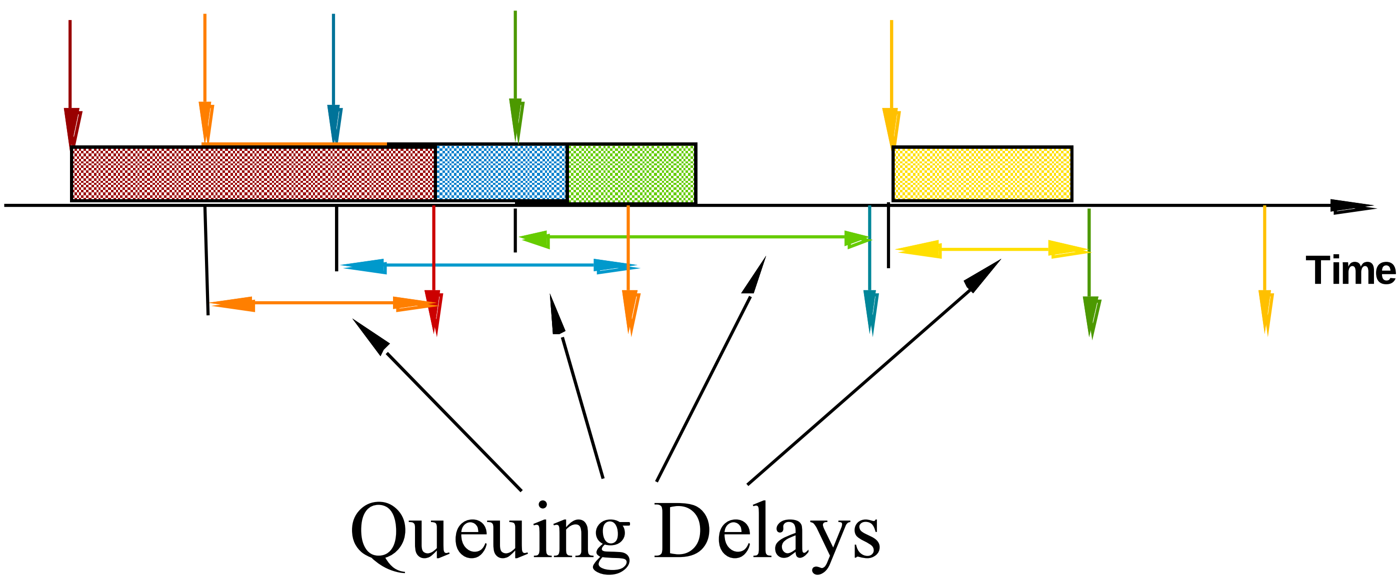


High Utilization also causes Interference

- As the work arrival rate:

  (packet arrival rate * packet length)

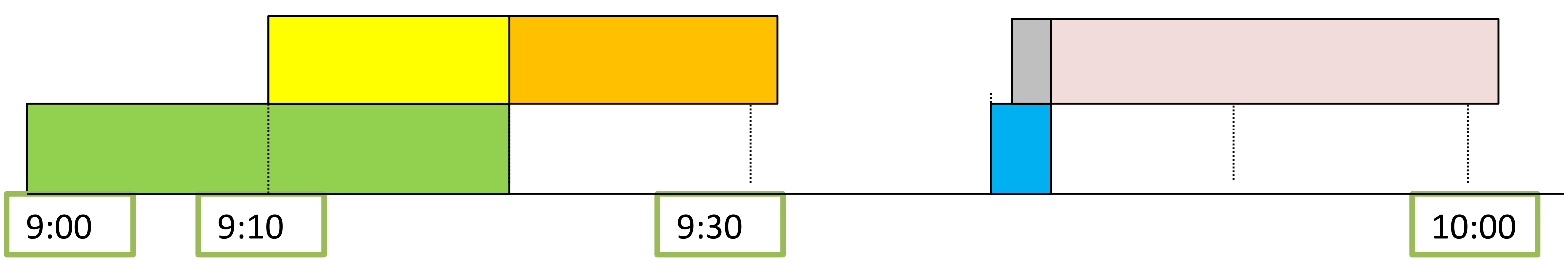
- increases, the opportunity for interference increases



Example: The simulation of Check-in (Tut)

	A	B	C	D	E	F	G	H	I	J
1	Simulation of the Check In Operation (4 Passengers)									
2		Arrival	Interarrival	Check In Service time						
3	Passenger No.	time	time	Check In Time	Start	End	Queue	Waiting time	Idle time	Events
4	1	9:00		20						
5	2	9:10		15						
6	3	9:40		5						
7	4	9:41		20						
8										
9										

- State: Number of customers in system
- Events: Pn arrived, Pn out, Pn in service
  - Passenger Number: 1, 2, 3, 4
  - Arrival times: 9:00, 9:10, 9:40, 9:41
  - Service times: 20, 15, 5, 20



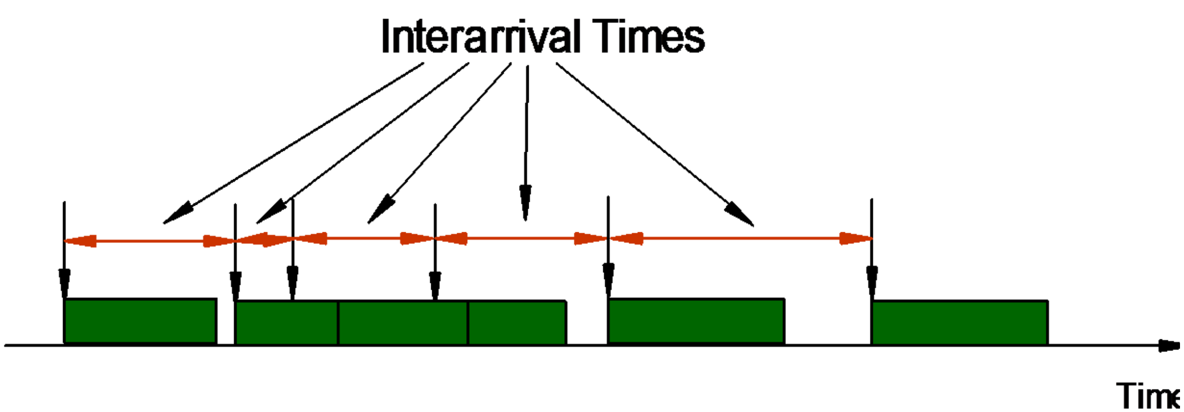
Capability of Queueing Models

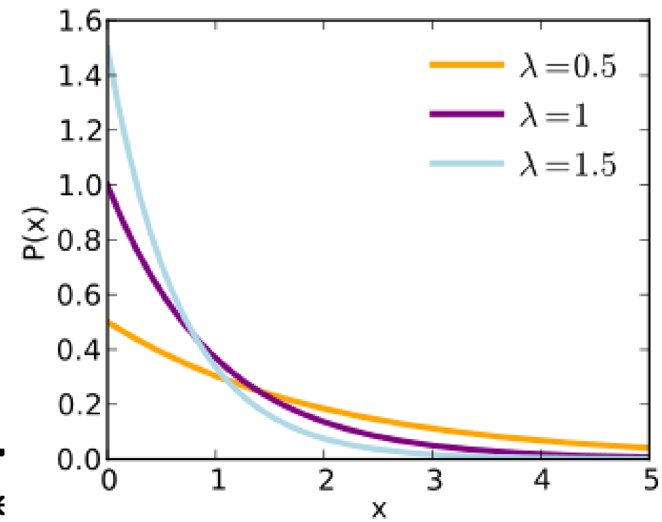
- The Queueing Models can answer questions like:

  - Distribution of the number of customers in queue.
  - Distribution of the number of customers in system.
  - Mean waiting time in the queue.
  - Mean system response time
    (waiting time in queue + service time).
  - Mean utilisation of the servers.

Poisson Process with Rate λ

- Interarrival times are independent and exponentially distributed
- Models well the accumulated traffic of many independent sources
- The average interarrival time is 1/ l (secs/packet), so l is the arrival rate (packets/sec)





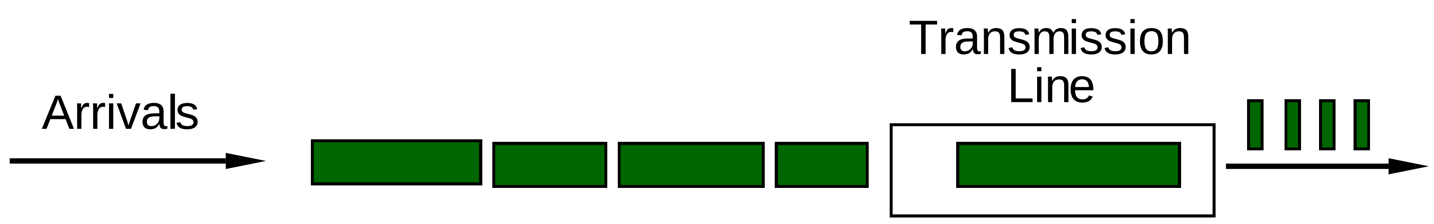
Service Centre Disciplines

- Common queue examples
  - First-come-first-served(FCFS)
  - Last-come-first-served(LCFS or LIFO)
  - Shortest Queue (SQ)
  - Priority Queuing (PQ)

- Service types from a queuing theory standpoint
  - Single server (one queue - one transmission line)
  - Multiple server (one queue - several transmission lines)
  - Priority server (several queues with hard priorities - one transmission line)
  - Shared server (several queues with soft priorities - one transmission line)

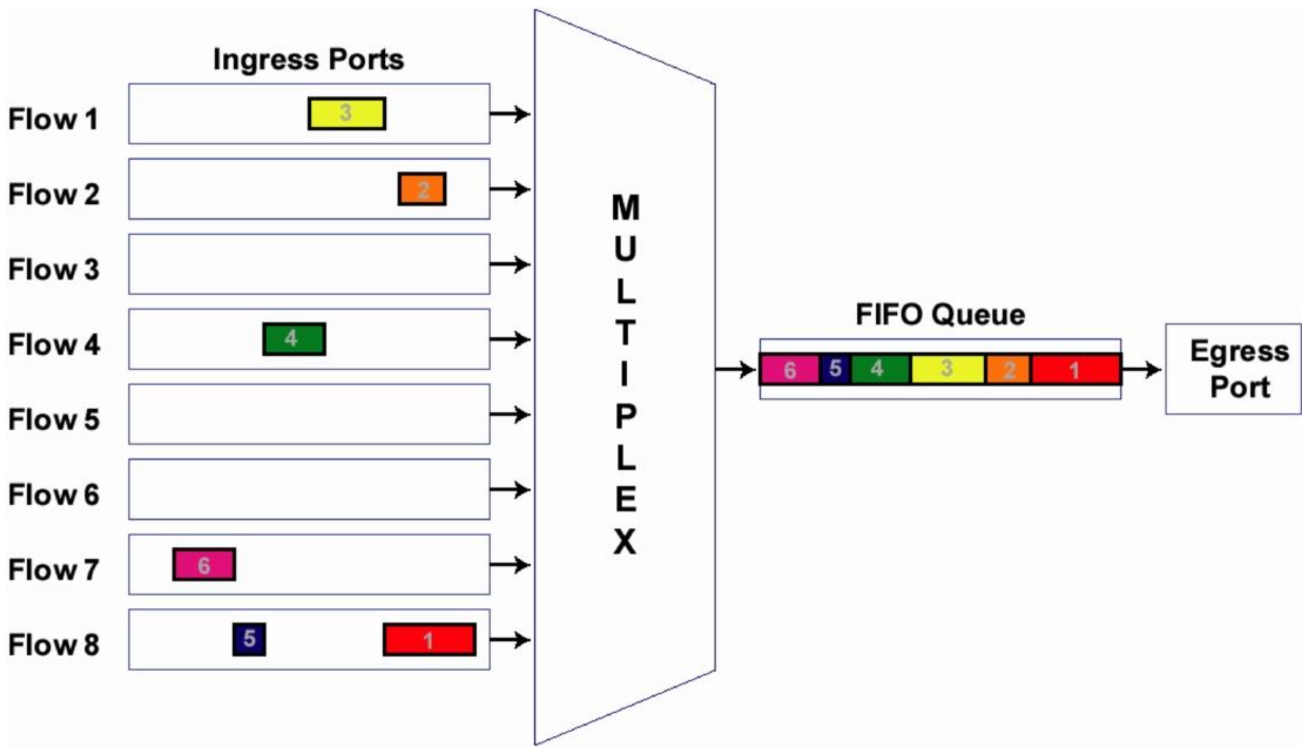
Single Server FIFO

- Single transmission line serving packets on a FIFO (First-In-First-Out) basis
- Each packet must wait for all packets found in the system to complete transmission, before starting transmission
  - Departure Time = Arrival Time + Workload Found in the System +Transmission time
- Packets arriving to a full buffer are dropped



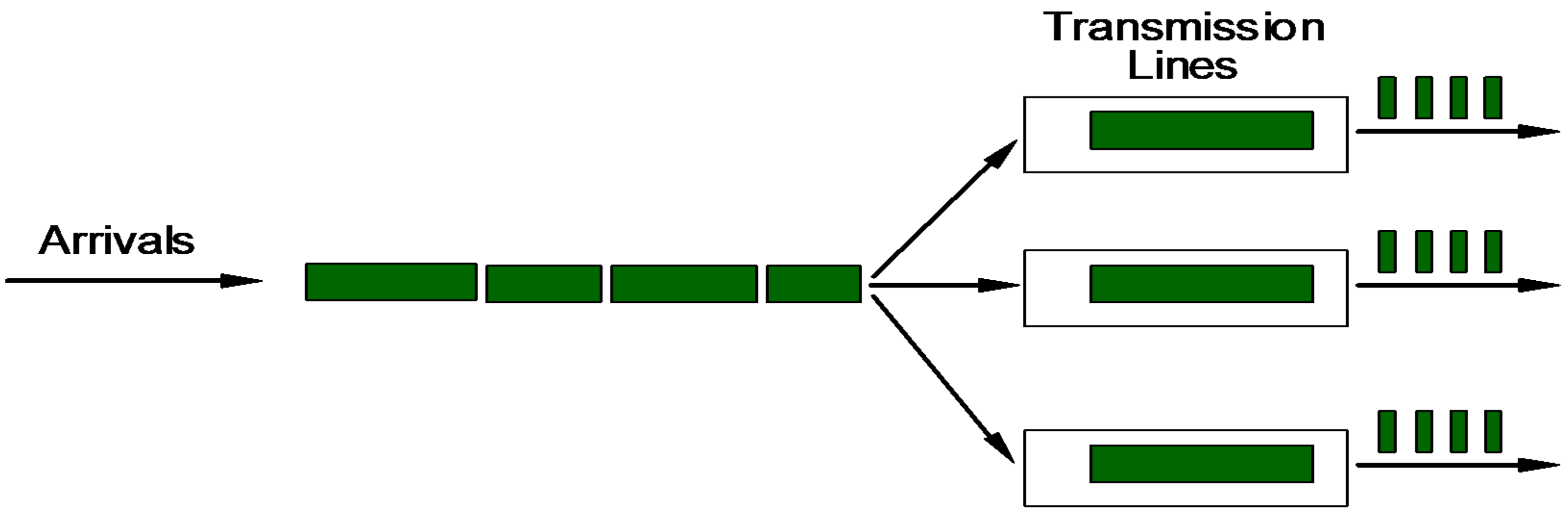
FIFO Queue

- Packets are placed on outbound link to egress device in FIFO order
  - Device (router, switch) multiplexes different flows arriving on various ingress ports onto an output buffer forming a FIFO queue



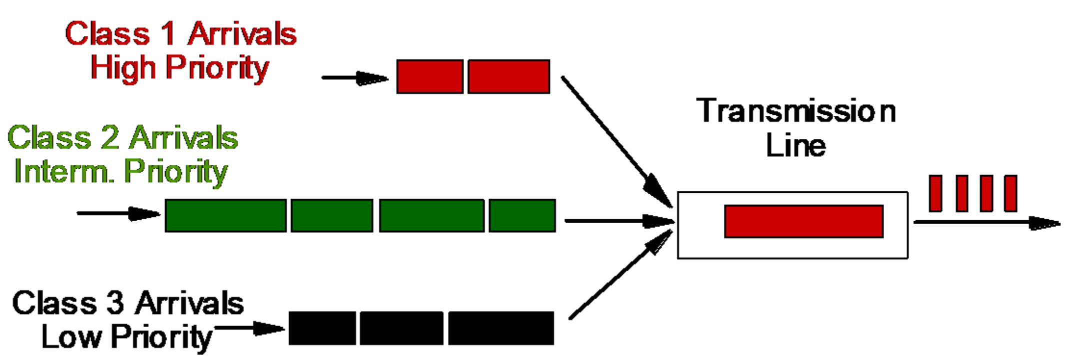
Multiple Servers

- Multiple packets are transmitted simultaneously on multiple lines/servers
- Head of the line service: packets wait in a FIFO queue, and when a server becomes free, the first packet goes into service



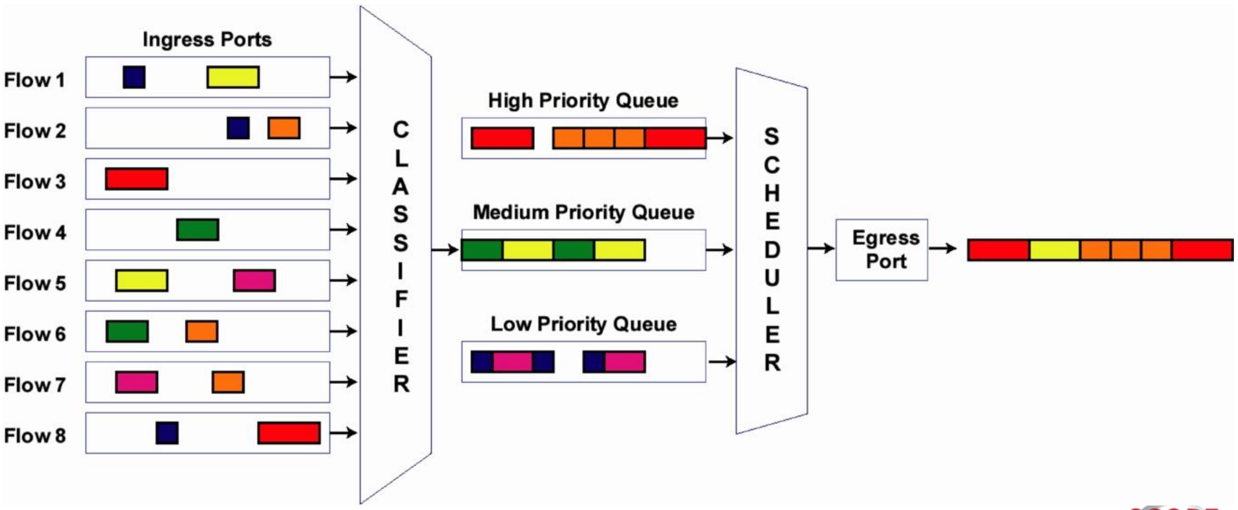
Priority Servers

- Packets form priority classes (each may have several flows)
- There is a separate FIFO queue for each priority class
- Packets of lower priority start transmission only if no higher priority packet is waiting
- Priority types:
  - Non-preemptive (high priority packet must wait for a lower priority packet found under transmission upon arrival)
  - Preemptive (high priority packet does not have to wait …)



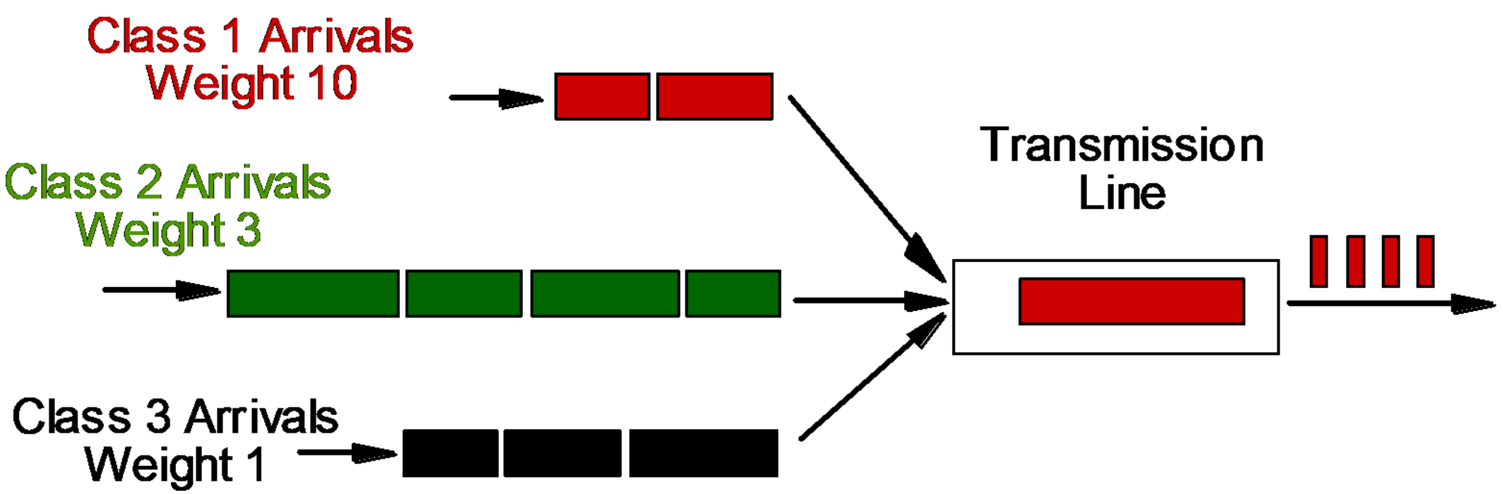
Priority Queuing

- Packets are classified into separate queues
  - E.g., based on source/destination IP address, source/destination TCP port, etc.
- All packets in a higher priority queue are served before a lower priority queue is served
  - Typically in routers, if a higher priority packet arrives while a lower priority packet is being transmitted, it waits until the lower priority packet completes



Shared Servers

- Again we have multiple classes/queues, but they are served with a “soft” priority scheme
- Round-robin
- Weighted fair queuing



Single queue System – M/M/1 System

- Nomenclature: M stands for "Memoryless" (a property of the exponential distribution)
  - M/M/1 stands for Poisson arrival process (which is memoryless)
  - M/M/1 stands for exponentially distributed transmission times
- Assumptions:
  - Arrival process is Poisson with rate λ packets/sec
  - Packet transmission times are exponentially distributed with mean $1/\mu$
  - One server
  - Independent interarrival times and packet transmission times
- Transmission time is proportional to packet length
- Note $1/\mu$ is secs/packet so μ is packets/sec (packet transmission rate of the queue)
- Utilization factor: $\rho = \lambda/\mu$ (stable system if $\rho < 1$)

Delay Calculation

- Let

  Q = Average time spent waiting in queue

  W = Average packet delay (transmission plus queuing)

  Note that $W = 1/\mu + Q$

- Also by Little's law

  $L = \lambda W$ and $L_q = \lambda Q$

where

  L_q = Average number waiting in queue

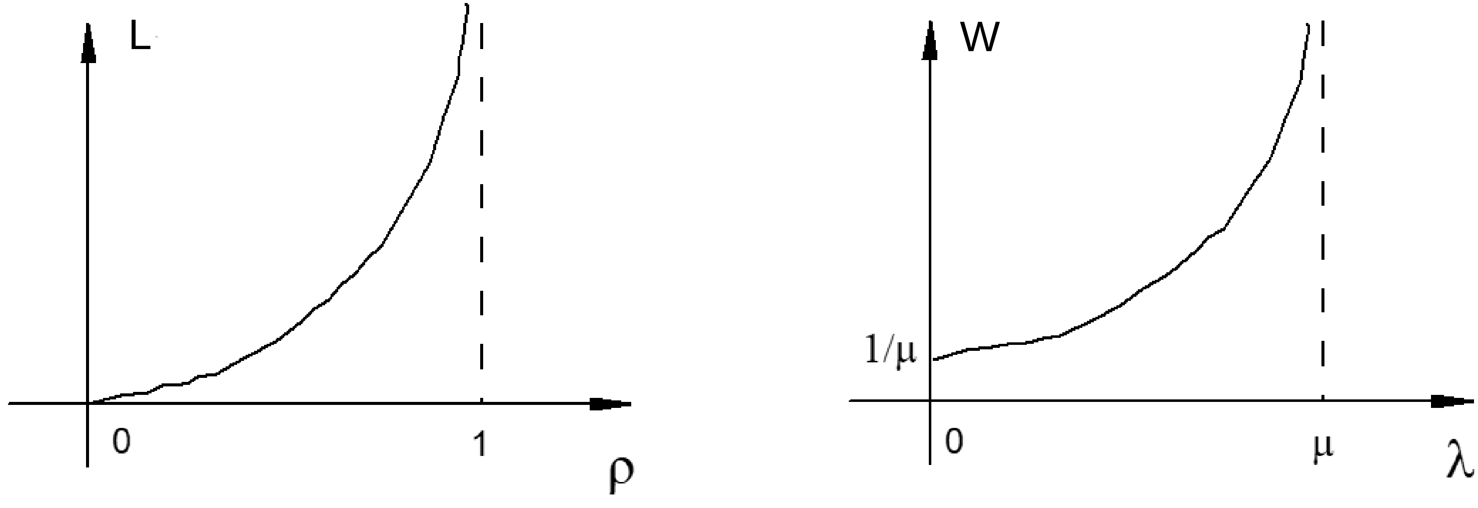
- These quantities can be calculated with formulas derived by Markov chain analysis

M/M/1 Results

- The analysis gives the steady-state probabilities of number of packets in queue or transmission
- P{n packets} = $\rho^n(1-\rho)$ where $\rho = \lambda / \mu$
- From this we can get the averages:

$$L = \rho /(1 - \rho)$$
$$W = L/ \lambda = \rho / \lambda(1 - \rho) = 1/(\mu - \lambda)$$

-



Example: How Delay Scales with Bandwidth

- Occupancy and delay formulas

  $L = \rho/(1 - \rho)$ $W = 1/(\mu - \lambda)$ $r = \lambda / \mu$

- Assume:
  - Traffic arrival rate λ is doubled
  - System transmission capacity μ is doubled

- Then:
  - Queue sizes stay at the same level (ρ stays the same)
  - Packet delay is cut in half (μ and λ are doubled)

- A conclusion: In high speed networks
  - propagation delay increases in importance relative to delay
  - buffer size and packet loss may still be a problem

M/M/m, M/M/∞ System

- Same as M/M/1, but it has m (or ∞) servers
- In M/M/m, the packet at the head of the queue moves to service when a server becomes free
- Qualitative result
  - Delay increases to ∞ as $r = \lambda / \mu$ approaches 1
- There are analytical formulas for the occupancy probabilities and average delay of these systems

Finite Buffer Systems: M/M/m/k

- The M/M/m/k system
  - Same as M/M/m, but there is buffer space for at most k packets. Packets arriving at a full buffer are dropped
- Formulas for average delay, steady-state occupancy probabilities, and loss probability
- The M/M/m/m system is used widely to size telephone or circuit switching systems

Characteristics of M/M/. Systems

- Advantage: Simple analytical formulas
- Disadvantages:
  - The Poisson assumption may be violated
  - The exponential transmission time distribution is an approximation at best
  - Interarrival and packet transmission times may be dependent (particularly in the network core)
  - Head-of-the-line assumption precludes heterogeneous input traffic with priorities (hard or soft)

Mathematical queue models (a/b/c):(d/e/f)

a = Arrivals distribution

b = Departures (service time) distribution

c = Number of parallel servers ($= 1, 2, \ldots, \infty$)

d = Queue discipline

e = Maximum number (finite or infinite) allowed in the system (in-queue plus in-service)

f = Size of the calling source (finite or infinite)

M = Markovian (or Poisson) arrivals or departures distribution (or equivalently exponential interarrival or service time distribution)

D = Constant (deterministic) time

E_k = Erlang or gamma distribution of time (or, equivalently, the sum of independent exponential distributions)

GI = General (generic) distribution of interarrival time

G = General (generic) distribution of service time

The queue discipline notation (symbol d) includes

$FCFS$ = First come, first served

$LCFS$ = Last come, first served

$SIRO$ = Service in random order

GD = General discipline (i.e., any type of discipline)

Examples

•(M/M/1) : (GD/ ∞ /∞)

*Poisson arrivals (λ) / departure (μ) single server, no limit on max number

•(M/D/4) : (FCFS/ 20 /∞)

*Poisson arrivals,
constant service time, and
4 parallel servers, and there is a limit of 20 customers on the entire systems
based on FCFS discipline

**Poisson arrival*: *arrival is a totally random fashion that cannot predict*

References

- Bazargan, M (2010) Airline Operations and Scheduling. 2nd edn. Ashgate

- Robinson, S. (2014) Simulation: The Practice of Model Development and Use. 2nd edn. Palgrave, Macmillan

- Taha, Hamdy A. (2010) Operations Research: An Introduction. 9th edn. Prentice Hall

References (Con't)

- Device Queuing Implementations
  - Vegesna, IP Quality of Service, Ciscopress.com, 2001
  - http://www.juniper.net/techcenter/techpapers/200020.pdf
- Probability and Queuing Models
  - Bertsekas and Tsitsiklis, Introduction to Probability, Athena Scientific, 2002, http://www.athenasc.com/probbook.html
  - Cohen, The Single Server Queue, North-Holland, 1992
  - Takagi, Queuing Analysis: A Foundation of Performance Evaluation. (3 Volumes), North-Holland, 1991
  - Gross and Harris, Fundamentals of Queuing Theory, Wiley, 1985
  - Cooper, Introduction to Queuing Theory, CEEPress, 1981
- OPNET Hybrid Simulation and Micro Simulation
  - See Case Studies papers in http://secure.opnet.com/services/muc/mtdlogis_cse_stdies_81.html