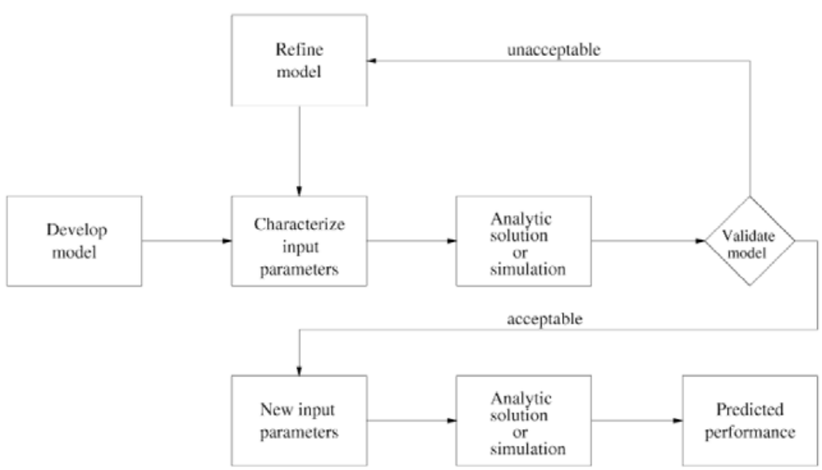
Performance Modeling

Introduction

Steps in Performance Modeling



Model Development

- Need to have a good understanding of how the system works
  - if possible, consult with those who have detailed knowledge of the system
  - Seek expert opinion on system components that need to be carefully modeled, e.g., bottleneck
  - Where appropriate, use modeling studies on similar systems as reference

Model Development

- Understand the objective of the performance study
- Identify the key entities and their attributes
- Select the type of queuing models to be used
  - e.g., single server queue, single service facility with multiple servers, queuing network
- For those entities that correspond to resources, identify their resource management schemes
  - e.g., queuing disciplines, buffer management schemes

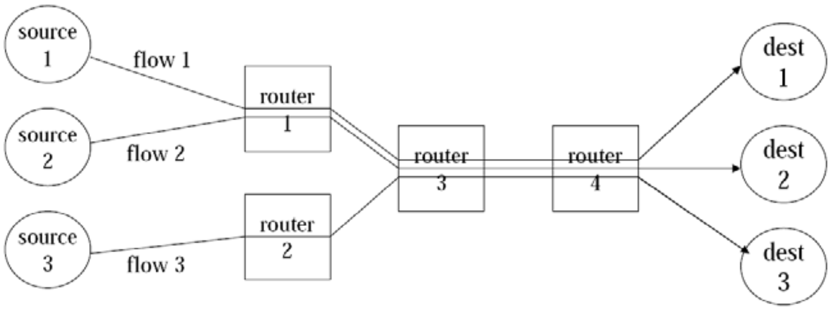
Model Development

- Identify the input parameters and the performance measures
  - input parameters may include some of the attributes (e.g., interarrival time) and/or resource management schemes
  - selection of input parameters and performance measures should be consistent with the objective of the study

Example 1 –Channel Scheduling

- In a communication networks, messages sent from source to destination are organized as "flows"
- At a router, the routing algorithm recognizes flows when determining the outgoing channel for each incoming message
- There is finite buffer space for each outgoing channel

Example 1 –Channel Scheduling



Objective 1

- Suppose objective is to compare the delay performance at a given channel for different combinations of channel scheduling algorithms and buffer allocation schemes, under different load levels

Model Development

- Focus on the channel from router 1 to router 32 flows (flows 1 and 2) are transmitted over this channel
- Assumptions
  - no error in message transmissions
  - a message is lost if buffer space at outgoing channel is not available
  - control messages are not included in the model

Entities and Attributes

Entities	Attributes
channel	capacity
buffer	size
message belonging to flow 1	length
message belonging to flow 2	length
system	interarrival time of flow 1
	interarrival time of flow 2

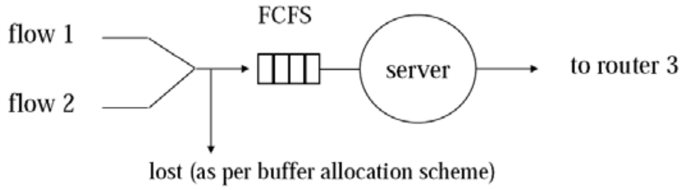
Queueing Model

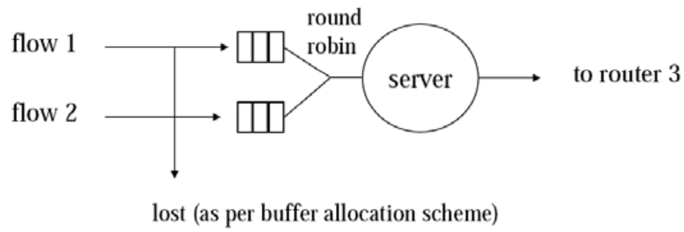
- Type of model -single server queue
- Server, which represents the communication channel
- Two classes of customers, one for each flow
- Finite waiting room, which models the finite buffer space

Resource Management Scheme

- Channel scheduling algorithms
  - first come, first served (FCFS)
  - fair queuing -a separate queue is maintained for each flow, and the two queues are serviced in a round robin fashion
- Buffer allocation scheme
  - complete sharing -buffers are allocated to messages in a FCFS manner, no distinction is made on the basis of flow ID
  - sharing with minimum allocation -each flow is guaranteed a minimum number of buffers; the remainder is allocated in a FCFS manner

Single Server Queue Model





Input Parameters and Performance Measures

- Input parameters
  - interarrival time of flow *i* messages, $i = 1, 2$
  - message length of flow *i*, $i = 1, 2$
  - channel capacity
  - buffer size
  - channel scheduling algorithm
  - buffer allocation strategy
- Performance measure
  - delay (waiting time + transmission time) experienced by flow *i* messages, $i = 1, 2$

Objective 2

- Suppose objective is to compare the end-to-end delay (within the network) for different combinations of channel scheduling algorithms and buffer allocation schemes, under different load levels

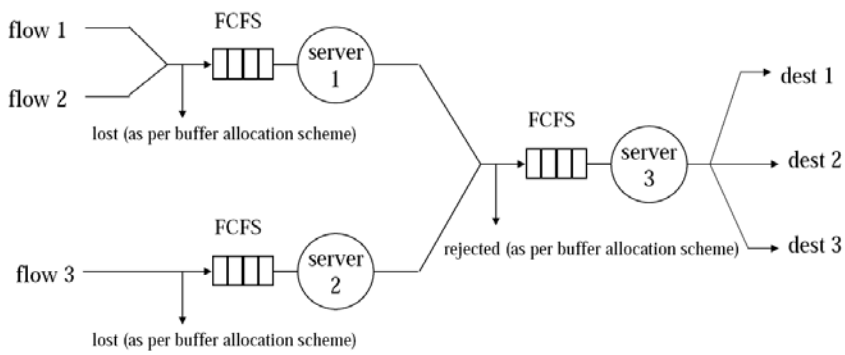
Entities and Attributes

Entities	Attributes
channel 1 (router 1 to 3)	capacity
channel 2 (router 2 to 3)	capacity
channel 3 (router 3 to 4)	capacity
buffer at channel *i*, *i* = 1, 2, 3	size
message belonging to flow *j*, *j* = 1, 2, 3	length routing frequency
system	interarrival time of flow *j*, *j* = 1, 2, 3

Queuing Model

- Type of model -queuing network
- Servers, each of them representing an outgoing channel (note: processing at routers is assumed to be negligible)
- Three classes of customers, one for each flow
- Finite waiting room at each server, which models the finite buffer space

Queueing Network Model



Example 2 –Computer Lab

- KAUST IT has installed *m* workstations in a lab
- Students may freely use these workstations for their school work
- It is anticipated that the number of students will increase in the near future
- The objective is to develop a queuing model that can be used to determine the required number of workstations

Model Assumptions

- Infinite population model -number of potential customers (i.e., students) is large
- An arriving customer who finds all workstations in use will
  - leave immediately with probability *p*, or
  - wait for an available workstation with probability *1-p*
- The queuing discipline is not well-defined
  - FCFS is a reasonable assumption (if one is interested in mean waiting time)

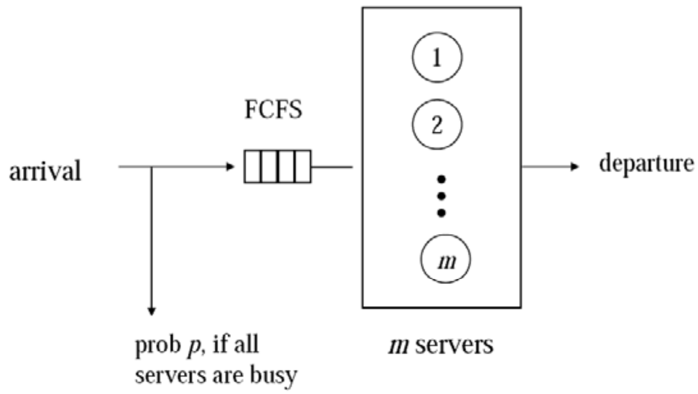
Entities and Attributes

Entities	Attributes
service facility	number of workstations
workstation	--
customer	service time
	p = prob. of leaving immediately if all workstations are busy at time of arrival
system	customer interarrival time

Queueing Model

- Type of model -single service facility with multiple servers (number of servers is *m*)
- Each server represents a workstation
- One class of customers -a customer can be served by any of the *m* servers
- FCFS discipline
- An arriving customer will leave immediately (with prob. *p*) if all servers are busy

Single Service Facility withMultiple Servers Model



Input Parameters and Performance Measures

- Input parameters
  - – customer interarrival time
  - – service time
  - – number of servers, *m*
  - – Probability, *p*
- Performance measures
  - – probability that an arriving customer finds all servers busy
  - – mean number of busy servers