System Performance Evaluation

Introduction

Objectives of a Performance Study

- Predict system performance for a given set of inputs
  - e.g., predict the mean response time of a Web server when the number of users is increased
- Evaluate alternative system design strategies
  - e.g., investigate the comparative performance of alternative channel scheduling algorithms (used in routers)
- Performance debugging and tuning

Performance Evaluation Activities

- May differ at different stages of system development
- System is operational
  - measure system behavior with a view to improve performance
  - develop validated model that can be used for performance prediction and capacity planning
- System in planning stage
  - use high level models to obtain performance estimates for alternative system configurations

Approaches to Performance Evaluation

- Performance measurement
  - obtain measurement data by observing the events and activities on an existing system
- Performance modeling
  - represent the system by a model and manipulate the model to obtain information about system performance

Performance Measurement

- Measure the performance directly on a system
- Need to characterize the workload placed on the system during measurement
- Generally provide the most valid results
- Nevertheless, not very flexible
  - difficult (or even impossible) to vary some parameters of the system to obtain performance results

Performance Modeling

- Model
  - an abstraction of the system obtained by making a set of assumptions about how the system works
  - capture the essential characteristics of the system
- Reasons of using models
  - experimenting with the real system may be
    - too costly
    - too risky, or
    - too disruptive to system operation
  - system may only be in the planning stage

Performance Modeling

- Workload characterization
  - capture the resource demands and intensity of the load brought to the system
- Performance measure
  - the measure of interest, such as system throughput, mean response time, number of transactions per second
- Solution methods
  - analytic modeling
  - simulation

Analytic Modeling

- Mathematical methods are used to obtain solutions to the performance measures of interest
- Numerical results are easy to compute if a simple analytic solution is available
- Useful approach when one only needs rough estimates of performance measures
- Solutions to complex models may be difficult to obtain
- In some cases, numerical results may be difficult to compute even though an analytic solution is available

Simulation

- Develop a simulation program that implements the model
- Run the simulation program and use the data collected to estimate the performance measures of interest
- A system can be studied at an arbitrary level of detail
- It may be costly to develop and run the simulation program

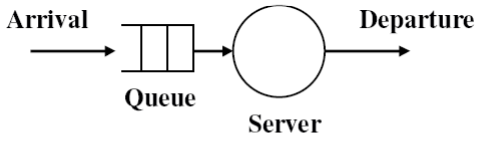
Stochastic Model

- Model contains some random input components which are characterized by probability distributions, e.g., time between arrivals to a system by exponential distribution
- Output is also random, and provides probability distributions of the performance measures of interest

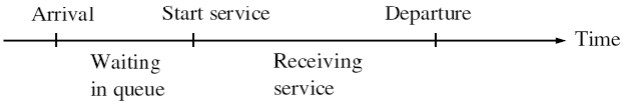
Queueing Model

- Popular model for computer systems and computer networks
- Single server queue: models a component of overall system, such as CPU, disk, communication channel
- Network of queues: models system components and their interaction

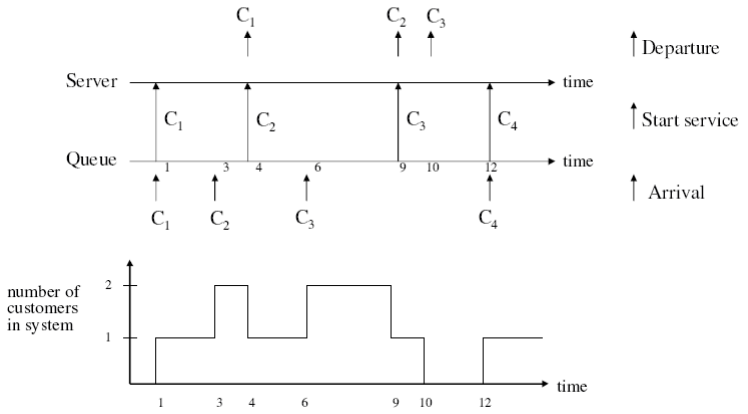
Single Server Queue Example (1) (Infinite population model)



❍ Behaviour of each customer



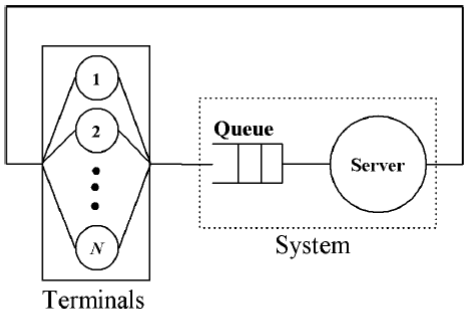
Timing Diagram



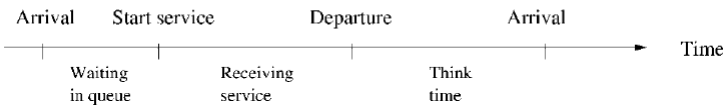
- Input parameters
  - interarrival time
  - service time
- Performance measures of interest
  - mean response time
  - mean number of customers in system
  - utilization factor

Single Server Queue Example (2)(Finite population model)

- Interactive system model

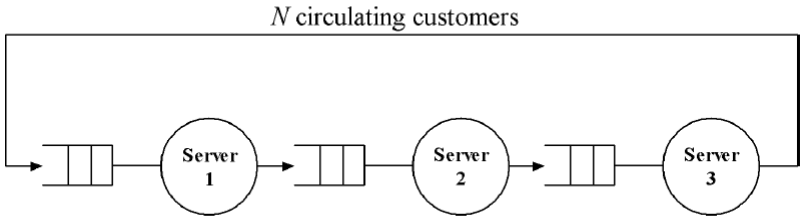


- Behavior of each user



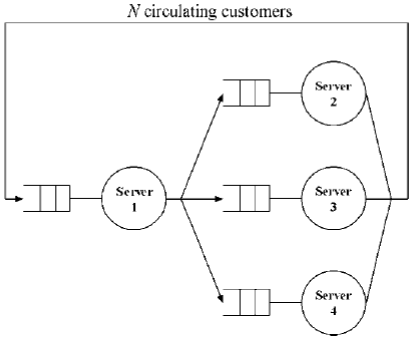
Queuing Network Model Example (1)

- Virtual circuit in a data network



Queuing Network Model Example (2)

- Central server model



Model Validation

- To develop an acceptable level of confidence that the performance measures predicted by the model are applicable to the system
  - Note: the most valid model may not be the most cost-effective model

Performance Prediction

- Run the model with new input parameters or workload to obtain predicted performance under new operating conditions