Overview of Performance Evaluation

- Introduction
- Common mistakes and how to avoid them
- Selection of techniques and metrics
- Some real examples

Introduction to Performance Evaluation

Introduction

- What is meant by computer system performance analysis:
  - Analysis with following objectives:
    - To get highest performance at a given cost
    - To use suitable tools and techniques
  - Use computer systems expertise
    - Understanding of architecture, OS, and software
- A performance analyst or engineer wears two hats:
  - Computer systems expert
    - To understand the performance requirements of their systems
  - Mathematician
    - To compare different alternatives to find one that meets the objectives

What Specifically Needs to be Learned?

- Specifying performance requirements
- Evaluating design alternatives
- Comparing two or more systems
- Determining the optimal value of a parameter (system tuning)
- Identifying performance bottleneck in a system
- Characterizing the load on a system (workload characterization)
- Determining the number and sizes of components (capacity planning)
- Predicting the performance at future loads (forecasting)

Definitions

- System
  - Consists of hardware, OS, and software
- Metrics
  - Quantitative measures to evaluate the performance of various components of a system
- Workload
  - Work demanded by users from a system

Overview of Selected Topics

- Selection of appropriate evaluation techniques, metrics, and workloads
- Performance measurement techniques and tools
- Use of proper statistical techniques to compare alternatives
- Experiment design for measurement- or simulation-based evaluation
- Simulation
- Use of simulation to evaluate system performance

Example: Selection of Techniques, Metrics, and Workloads for a System

- What performance metrics should be used to compare the performance of the following system:
  - Two disk drives
  - Two transaction processing systems
  - Two packet retransmission algorithms

Performance Measurement Techniques and Tools

- Types of workloads
- Popular benchmarks
- The art of workload selection
- Workload characterization techniques
- Monitors
- Accounting logs
- Monitoring distributed systems
- Load drivers
- The art of data presentation

Example: Measurement-Based Evaluation

- Which type of monitor (software or hardware) would be more suitable for measuring each of the following quantities:
  - Number of instructions executed by a processor
  - Degree of multiprogramming on a time-sharing system
  - Response time of packets on a network

Statistical Techniques

- Probability and statistics concepts
- Important distributions
- Summarizing measured data by a single number
- Summarizing the variability of measured data
- Graphical methods to determine the distribution of measured data
- Simple statistics
- Confidence interval
- Comparing two alternatives
- Measures of relationship
- Regression models

Example: Statistical Techniques

The number of packets lost on two links was measured for four file sizes as shown below:

File size	Link A	Link B
1000	5	10
1200	7	3
1300	3	0
50	0	1

Which link is better?

Experimental Design and Analysis

- Introduction to experimental design
- 2^k factorial design
- 2^k factorial design with replications
- General full factorial designs with k factors

Example: Experiment Design

- Performance of a system depends on following three factors:
  - Garbage collection technique used: G1, G2, or none
  - Type of workload: editing, computing, or artificial intelligence (AI)
  - Type of CPU: C1, C2, or C3
- How many experiments are needed? How does one estimate the performance impact of each factor?

Simulation

- Introduction of simulation
- Types of simulations
- Model verification and validation
- Analysis of simulation results
- Random number generation
- Testing random number generators
- Random variate generation
- Commonly used distributions

Example: Simulation

- In order to compare the performance of two cache replacement algorithms:
  - What type of simulation model should be used?
  - How long should the simulation be run?
  - What can be done to get the same accuracy with a shorter run?
  - How can one decide if the random-number generator in the simulation is a good generator?

Queuing Theory

- Introduction to queuing theory
- Analysis of a single queue
- Queuing networks
- Operational laws
- Mean value analysis and related techniques

Example: Queuing Theory

- The average response time of a database system is 3 seconds. During a 1-minute observation interval, the idle time on the system was 10 seconds. Using a queuing model for the system, determine the following:
  - System utilization
  - Average service time per query
  - Number of queries completed during the observation interval
  - Average number of jobs in the system
  - Probability of number of jobs in the system being greater than 10
  - 90-percentile response time
  - 90-percentile waiting time