EEC 686/785
Modeling & Performance Evaluation of Computer Systems

Lecture 1

Wenbing Zhao
Department of Electrical and Computer Engineering
Cleveland State University
wenbing@ieee.org
http://academic.csuohio.edu/zhao_w/teaching/EEC685-F05/eec685.htm
(based on Dr. Raj jain's lecture notes)

Outline

- Syllabus
- Introduction to performance evaluation

Why Studying Performance Evaluation?

- Performance is a key criterion in the design, procurement, and use of computer systems
- The goal of computer systems professionals is to get the highest performance for a given cost
- To achieve that goal, computer systems professions need at least a basic knowledge of performance evaluation terminology and techniques

Course Objective

- You will learn basic techniques that help solve a majority of day-to-day problems, such as:
  - Specifying performance requirements
  - Evaluating design alternatives
  - Comparing two or more systems
  - System tuning – determining optimal value of a parameter
  - Bottleneck identification
  - Workload characterization
  - Capacity planning – determining number and sizes of components
  - Forecasting – predicting performance at future loads

Course Outline

- Overview of performance evaluation
- Measurement techniques and tools
- Probability theory and statistics
- Experimental design and analysis
- Simulation
- Queueing models

Course Projects

- Select a system, for example:
  - Database, language compiler, text editor, processor
- Do:
  - Performance measurements
  - Analyze the collected data
  - Simulate
  - Analytically model the subsystem
- Prefer: Team of 2
- Goal: provide an insight (or information) not obvious before the project

Grading Policy

- Grade components & relative weights:
  - Assignments: 20% (individual work)
  - Course project: 20% (team work)
  - Midterm exams: 40% (20% each)
  - Final exam: 20%
  - Exams are all closed-book closed-notes, except that you can bring one page of formulas and definitions (US Letter size or smaller)

Grading Policy

- **Do not cheat!**
  - Do not copy other student's homework, exams or project
  - Do not copy someone else's work found on the Internet
    - Including project description and report
    - You can quote a sentence or two, but put those in quote and give reference
    - Cite your references

Reference Texts

- *The Art of Computer Systems Performance Analysis: Techniques for Experimental Design, Measurement, Simulation, and Modeling*, by Raj Jain, John Wiley & Sons, 1991
- *Computer Systems Evaluation and Prediction* , by Paul J. Fortier, and Howard E. Michael, Digital Press, 2003

Miscellaneous

- Instructor:
  - Dr. Wenbing Zhao
    - E-mail: wenbing@ieee.org
    - Lecture hours:
      - MW 4:00-5:50pm
    - Office hours: *TTh 4:00-6:00pm and by appointment*
- Course Web site:
  - http://academic.csuohio.edu/zhao_w/teaching/EEC685-F05/eec685.htm
  - Lecture nodes and homework will be posted

Basic Terms

- **System**: any collection of hardware, software, and firmware
- **Metrics**: the criteria used to evaluate performance of the system and components
- **Workloads**: the requests made by the users of the system

Basic Terms

- **Parameters**: system and workload characteristics that affect the performance of the system
  - Example system parameters:
    - Quantum size for CPU allocation
    - Working set size for memory allocation
  - Example workload parameters:
    - Number of users, request arrival patterns, priority, etc.

Basic Terms

- **Factors**: parameters that are varied in the study
- **Outliers**: values that are too high or too low compared to a majority of values in a set
  - It may or may not be caused by a real system phenomenon

Basic Terms

- **Measurement**:
- **Simulation**: the technique of representing the real world by a computer program. A simulation should imitate the internal processes and not merely the results of the thing being simulated
- **Analytical modeling**: a modeling technique that uses tools based on mathematical models such as queueing theory and stochastic processes

The Art of Performance Evaluation

- Performance evaluation is an art. Like a work of art, successful evaluation cannot be produced mechanically
  - Every evaluation requires an intimate knowledge of the system being modeled and a careful selection of the methodology, workload, and tools
  - Defining the real problem and converting it to a form in which established tools and techniques can be used and where time and other constraints can be met is a major part of the analyst's art.

The Art of Performance Evaluation

- Given the same data, two analysts may interpret them different
- Example: Comparing two systems. The throughputs of two systems A and B in transactions per second is as follows:

System	Workload 1	Workload 2
A	20	10
B	10	20

The Art of Performance Evaluation

- **Three solutions**:
- *The two systems are equally good* – compare the average of throughput

System	Workload 1	Workload 2	Average
A	20	10	15
B	10	20	15

The Art of Performance Evaluation

- *System A is better than B* – compare the ratio with system B as the base

System	Workload 1	Workload 2	Average
A	2	0.5	1.25
B	1	1	1

The Art of Performance Evaluation

- *System B is better than A* – compare the ratio with system A as the base

System	Workload 1	Workload 2	Average
A	1	1	1
B	0.5	2	1.25

- Similar games can be played in: selection of workload, measuring the systems, presenting the results

Professional Organizations

- ACM SIGMETRICS
- ACM SIGSIM
- IEEE Computer Society: technical committee on simulation
- CMG: The Computer Measurement Group, In.
  - Practical uses of computer performance
  - Regional groups in the united states and abroad

Professional Organizations

- IFIP Working Group 7.3
  - IFIP = International Federation for Information Processing
  - Multinational federation of technical societies
  - Several technical committees (TCs) and working groups (WGs)
  - WG 7.3 is devoted to computer systems modeling

Professional Organizations

- SCS: The Society for Computer Simulation
  - Publishes simulation books, journals
  - Sponsors a number of conferences
- SIAM: The Society for Industrial and Applied Mathematics
  - Development of new mathematical techniques
  - Journals: SIAM journal on computing etc.
- ORSA: The Operations Research Society of America

Conferences

- ACM SIGMETRICS
- PERFORMANCE – IFIP Working Group 7.3 sponsored conferences
- ACM SIGSIM, IEEE and SCS conferences

Journals

- **Performance Evaluation**: published monthly by Elsevier Science Publishers
- **Simulation**: monthly by SCS
- **Transactions** of the Society for Computer Simulation
- **Simulation Digest**: by ACM SIGSIM and IEEE TC on Simulation
- See also computer systems journals
  - IEEE Transactions on Software Engineering
  - IEEE Transactions on Computers
  - ACM Transactions on Computer Systems

Common Mistakes in Performance Evaluation

- No goals
  - No general purpose model. Each model must be developed with a particular goal in mind
  - Techniques, metrics, workload all depend upon the goal
  - Not trivial
- Biased goals
  - To show that OUR system is better than THEIRS
  - The performance analyst's role is like that of a jury (must be objective)

Common Mistakes in Performance Evaluation

- Unsystematic approach
  - Selecting system parameters, factors, metrics, and workloads arbitrarily will lead to inaccurate conclusions
  - Should identify a complete set of goals, system parameters, factors, metrics, and workloads
- Analysis without understanding the problem
  - A large share of the analysis effort goes in to defining a problem
  - A problem well stated is half solved

Common Mistakes in Performance Evaluation

- Incorrect performance metrics
  - The choice of correct performance metrics depends upon the services provided by the system or subsystem being modeled
    - Comparing the MIPS of two different CPU architectures (e.g., RISC and CISC) is meaningless since the instructions on the two computers are equal
- Unrepresentative workload
  - The workload used to compare two system should be representative of the actual usage of the systems in the field

Common Mistakes in Performance Evaluation

- Wrong evaluation technique
  - There are three evaluation techniques: measurement, simulation, and analytical modeling
  - There are a number of factors that should be considered in selecting the right technique
- Overlook important parameters
  - The final outcome of the study depends heavily upon the parameters used
  - Overlooking one or more important parameters may render the results useless

Common Mistakes in Performance Evaluation

- Ignore significant factors
  - It is important to identify those parameters, which, if varied, will make a significant impact on the performance
  - Factors that are under the control of the end user and can be easily changed by the end user should be given preference over those that cannot be changed
  - For unknown parameters, a sensitivity analysis, which shows the effect of changing those parameters from their assumed values,, should be done to quantify the impact of the uncertainty

Common Mistakes in Performance Evaluation

- Inappropriate experimental design
  - Experimental design relates to the number of measurement or simulation experiments to be conducted and the parameter values used in each experiment
  - In naïve experimental design, each factor is changed one by one. This simple design may lead to wrong conclusions if the parameters interact with each other
  - Should use full factorial experimental designs and fractional factorial designs

Common Mistakes in Performance Evaluation

- Inappropriate level of detail
  - The level of detail used in modeling a system has a significant impact on the problem formulation
  - Avoid formulations that are either too narrow or too broad
- No analysis
  - It is not adequate to collect enormous amounts of data without appropriate analysis and interpretation of the results
- Erroneous analysis
  - Don't make the common mistakes such as taking average of ratios and too short simulation

Common Mistakes in Performance Evaluation

- No sensitivity analysis
  - The analysis results may be sensitive to the workload and system parameters
  - Without a sensitivity analysis, one cannot be sure if the conclusions would change if the analysis was done in a slightly different setting
- Ignoring errors in input
  - Often the parameters of interest cannot be measured. Instead, another variable that can be measured is used to estimate the parameter
  - Such situations introduce additional uncertainties in the input data. The analyst needs to adjust the level of confidence on the model output obtained from such data

Common Mistakes in Performance Evaluation

- Improper treatment of outliers
  - It requires careful understanding of the system being modeled to decide if an outlier is caused by a real system phenomenon
- Assuming no change in the future
  - The future workload and system behavior might change and therefore, the analyst should limit the amount of time into the future that predications are made
- Ignoring variability
  - It is common to analyze only the mean performance since determining the variability is difficult
  - However, the mean alone may be misleading if the variability is high

Common Mistakes in Performance Evaluation

- Too complex analysis
  - Given two analyses leading to the same conclusion, one that is simpler and easier to explain is obviously preferable
  - It is better to start with simple models of experiments, get some results or insights, and then introduce the complications
- Improper presentation of results
  - To convey or selling the analysis results, it requires the prudent use of words, pictures, and graphs to explain the results and analysis
  - The right metric to measure the performance of an analyst is not the number of analyses performed but the number of analyses that helped the decision makers

Common Mistakes in Performance Evaluation

35

- Ignoring social aspects
  - Successful presentation of the analysis results requires two types of skills: social and substantive
  - Writing and speaking are social skills while modeling and data analysis are substantive skills
- Omitting assumptions and limitations
  - Omitting assumptions and limitations may lead the user to apply the analysis to another context where the assumptions will not be valid

A Systematic Approach to Performance Evaluation

36

1. State goals and define the system
2. List services and outcomes
3. Select metrics
4. List parameters
5. Select factors to study
6. Select evaluation technique
7. Select workload
8. Design experiments
9. Analyze and interpret data
10. Present results

Step 1 – State Goals and Define the System

- The first step in performance evaluation is to state the goals of the study and define what constitutes the system by delineating system boundaries
- Given the same set of hardware and software, the definition of the system may vary depending upon the goals of the study

Step 1 – State Goals and Define the System

- Given two CPUs, for example,
  - If the goal is to estimate their impact on the response time of interactive users, the system would consist of the timesharing system, and the conclusions of the study may depend significantly on components external to the CPU
  - If the two CPUs are basically similar except for their ALUs and the goal is to decide which ALU should be chosen, the CPUs may be considered the system's and only the components inside the CPU may be considered part of the system
- The choice of system boundaries affects the performance metrics as well as workloads used to compare the systems

Step 2 – List Services and Outcomes

- Each system provides a set of services, for example:
  - A computer network allows its users to send packets to specified destinations
  - A database system responds to queries
- The second step in analyzing a system is to list these services

Step 2 – List Services and Outcomes

- When a user requests any of these services, there are a number of possible outcomes
- For example, a database system may answer a query
  - Correctly
  - Incorrectly (due to inconsistent updates)
  - Or not at all (due to deadlocks or some similar problems)
- A list of services and outcomes is useful later in selecting the right metrics and workload

Step 3 – Select Metrics

- Select criteria to compare the performance. These criteria are called metrics
- In general, the metrics are related to the speed, accuracy, and availability of services, for example:
  - The performance of a network is measured by the speed (throughput and delay), accuracy (error rte), and availability of the packets sent

Step 4 – List Parameters

- Make a list of all the parameters that affect performance
- The list can be divided into *system parameters* and *workload parameters*
- *System parameters* include both hardware and software parameters => generally do not vary among various installations of the system
- *Workload parameters* are characteristics of users' requests => vary from one installation to the next
- The list of parameters may not be complete during the first pass of the analysis. Additional parameters might be discovered after each pass

Step 5 – Select Factors to Study

- The list of parameters can be divided into two parts:
  - Those that will be varied during the evaluation
  - Those that will not
- The parameters to be varied are called factors and their values are called levels
- It is better to start with a short list of factors and a small number of levels for each factor and to extend the list in the next phase
- The parameters that are expected to have a high impact on the performance should be preferably selected as factors

Step 6 – Select Evaluation Technique

- The three broad techniques for performance evaluation are analytical modeling, simulation and measuring a real system
- The selection of the right technique depends upon the time and resources available to solve the problem and the desired level of accuracy

45

Step 7 – Select Workload

- The workload consists of a list of service requests to the system
- Depending upon the evaluation technique chosen, the workload may be expressed in different form:
  - For analytical modeling => as a probability of various requests
  - For simulation => a trace of requests measured on a real system
  - For measurement => user scripts to be executed on the systems
- It is essential that the workload be representative of the system usage in real life

46

Step 8 – Design Experiments

- Decide on a sequence of experiments that offer maximum information with minimal effort
- In practice, it is useful to conduct an experiment in two phases
  - In first phase, the number of factors may be large but the number of levels is small => to determine the relative effect of various factors
  - In second phase, the number of factors is reduced and the number of levels is increased

Step 9 – Analyze and Interpret Data

- Interpreting the results of an analysis is a key part of the analyst's art
- Analysis only produces results and not conclusions
- Outcomes of measurements and simulations are random quantities. In comparing two alternatives, it is necessary to take into account the variability of the results

Step 10 – Present Results

- The final step is to communicate the results to other members of the decision-making team.
- It is useful to present the results in graphic form and without statistical jargon
- At this point, the knowledge gained may prompt the analyst to go back and reconsider some of the decisions made in previous steps. The complete project might consists several cycles through the steps