A Systematic Approach to Performance Evaluation

Prof. Santosh Kumar
Dept. of Computer Science
University of Memphis
Fall 2008

Some Terminology

- **System:** Any collection of hardware, software, or both
- **Model:** Mathematical representation of a concept, phenomenon, or system
- **Metrics:** The criteria used to evaluate the performance of a system
- **Workload:** The requests made by the users of a system

Terminology (contd.)

- **Parameters:** System and workload characteristics that affect system performance
- **Factors:** Parameters that are varied in a study especially those that depend on users
- **Outliers:** Values (in a set of measurement data) that are too high or too low as compared to the majority

Major Steps

1. State Goals
2. List Services and Outcomes
3. Select Appropriate Metrics
4. List the Parameters
5. Select Evaluation Techniques
6. Select Workload
7. Design Experiment (s)
8. Analyze
9. Present the Results

State Goals

- Identify the goal of study
  - Not trivial, but
  - Will affect every decision or choice you make down the road
- Clearly define the system
  - Where you draw the boundary will
    - Dictate the choice of model
    - Affect choice of metrics and workload

List Services and Outcomes

- Identify the services offered by the system
- For each service, identify all possible outcomes
- What's the point
  - These will help in the selection of appropriate metrics

Select Appropriate Metrics

- These are the criteria for performance evaluation
- Look for these must have properties
  - Specific
  - Measurable
  - Acceptable
  - Realizable
  - Thorough
- Examples?

Select Appropriate Metrics

- These are the criteria for performance evaluation
- Desired Properties
  - Specific
  - Measurable
  - Acceptable
  - Realizable
  - Thorough
- Prefer those that
  - Have low variability,
  - Are non-redundant, and
  - Are complete

- Examples?

Examples

- Successful Service Rate – *Throughput*
- Frequency of correct results – *Reliability*
- Being available when needed – *Availability*
- Service users fairly – *fairness*
- Efficiency of resource usage - *Utilization*
- How to measure these?

A Classification

- Higher is better
  - Examples?
- Lower is better
  - Examples?
- Nominal is Best
  - Examples?

Criteria for Metric Set Selection

- Low-variability
  - Helps reduce the number of runs needed
  - Advice: Avoid ratios of two variables
- Non-redundancy
  - Helps make results less confusing and reduce the effort
  - Try to find a relationship between metrics
    - If a simple relationship exists, keep only one
- Completeness

Debate on Metrics

- Metric for measuring fairness?

- Another example:
  - Objective: Hide sources of information in sensor networks
  - Metrics for evaluation?

Common Metrics

- Response Time
  - Turnaround time, reaction time
  - Stretch factor
    - Response time at a particular load divided by response time at minimum load
- Throughput
  - Nominal capacity: Under ideal workload
  - Usable capacity: With acceptable response time
- Efficiency: usable capacity/nominal capacity
- Utilization: busy time/elapsed time

Metrics - Summary

- Metrics chosen should be measurable
  - Can assign a numerical value to it
- Acceptable,
- Easy to work with (i.e. can measure it easily)
- Avoid redundancy
- Pay attention to the unit used
- Sanity check – Check the boundary conditions (i.e. best system, ideal workload, etc.) to see if the metric is sensible

Major Steps

1. State Goals
2. List Services and Outcomes
3. Select Appropriate Metrics
4. List the Parameters
5. Select Evaluation Techniques
6. Select Workload
7. Design Experiment (s)
8. Analyze
9. Present the Results

List the Parameters

- Identify all system and workload parameters
  - System parameters
    - Characteristics of the system that affect system performance
  - Workload parameters
    - Characteristics of usage (or workload) that affect system performance
- Categorize them according to their effects on system performance
- Determine the range of their variation or expected variation
- Decide on one or at most a couple to vary while keeping others fixed

Select Evaluation Technique(s)

- Three Techniques
  - Measurement
  - Simulation
  - Analytical Modeling

Measurement, Simulation, or Analysis?

- Can be a combination of two or all three
- Use the goal of study to guide your decision
- The resources and skills available may also be taken into account
- Remember, each of these techniques has its pros and cons
  - Let us look at some of them

Measurement

- (+) Provides realistic data
- (+) Can test the limits on load
- (-) System or a prototype should be working
- (-) The prototype may not represent the actual system
- (-) Not that easy to correlate cause and effect
- Challenges
  - Defining appropriate metrics
  - Using appropriate workload
  - Statistical tools to analyze the data

Simulation

- (+) Less expensive than building a prototype
- (+) Can test under more load scenarios
- (-) Synthetic since the model is not the actual system
- (-) Can not use simulation to make any guarantees on expected performance
- Challenges
  - Need to be careful when to use simulation
  - Need to get the model right
  - Need to represent results well (the graphical tools)
  - Need to learn simulation tools

Analytical Modeling

- (+) Can make strong guarantees on expected behavior
- (+) Can provide an insight in to cause and effect
- (+) Does not need to build a prototype
- (-) Performance prediction only as good as the model
- Challenges
  - Significant learning curve
  - Mathematically involved
  - Choosing the right model (the art work)

Bottom Line

- You can use measurement to demonstrate feasibility of an approach.
- You can use measurement or simulation to show an evidence that your algorithm or system performs better than competing approaches in certain situations.
- But, if you would like to claim any properties of your algorithm (or system), the only option is to use analysis and mathematically prove your claim.

When to Use What?

- It is good to be versed in all three

1. Can start with measurement or simulation to get a feel of the model or expected behavior
2. Start with a simple model
3. Perform an analysis to predict the performance and prove some behavioral properties
4. Observe the actual performance to determine the validity of your model and your analysis
5. Can use simulation for the previous step if a working system is not available/feasible
6. Go back to revise the model and analysis if significant inconsistency is observed and start with Step 4
7. Finally use simulation to verify your results for large scale data or for scenarios that can not be modeled with existing expertise and available time

Team Homework

- Submit a list of Questions you would put on an in-class Quiz
  - At least two questions that have yes/no answer
  - At least two multiple choice questions
  - At least two short-answer questions
  - No memorization questions
  - Design your questions to test understanding
  - Can propose application questions

Major Steps

1. State Goals
2. List Services and Outcomes
3. Select Appropriate Metrics
4. List the Parameters
5. Select Evaluation Techniques
6. Select Workload
7. Design Experiment (s)
8. Analyze
9. Present the Results

Select Workload

- What is a workload?
- How do you represent it?
  - Range of values
    - What should be the increment size?
  - Probability Distribution
    - Need to find a good model that approximates reality
    - May require measurement/statistical analysis
    - In simulation, use an appropriate random number generator to produce values
  - Trace from an actual system

Design Experiment (s)

- To provide maximum information with minimum effort
  - Field experiments can take enormous preparation time
  - Attempt to get several experiments done in one setup
  - Explore if you can use data collected by someone else
  - Also, explore if you can use remote labs
  - Finally, explore if you can use simulation without loosing significant validity
  - Modifying simulation code can be time consuming, as well
- In both simulation and measurement, repeat the same experiment (for a fixed workload and fixed parameter values) sufficient number of times for statistical validity
- Always keep the goal in mind

Major Steps

1. State Goals
2. List Services and Outcomes
3. Select Appropriate Metrics
4. List the Parameters
5. Select Evaluation Techniques
6. Select Workload
7. Design Experiment (s)
8. Analyze
9. Present the Results

Analyze

- In Analytical Modeling
  - Carry out mathematical derivations that prove expected system behavior
- In Measurement,
  - Statistically analyze the collected data
  - Summarize the results by computing statistical measures

Present the Results

- In Analytical Modeling
  - Clear statements of lemmas, and theorems
  - Description of an algorithm with a proof of its properties
  - Present numerical computation results
    - to show how to use the formulae, and
    - to show the effect of varying the parameters
  - Perform simulation/measurement to show the validity of the model and analysis

Present the Results (contd.)

- In Simulation and Measurement
  - Clear statement of the goals of experiment
  - A list of assumptions
  - The experiment set up
    - platforms, tools, units, range of values for parameters
  - Graphical presentation of results
    - The simpler it is to understand the graphs, the better it is

Present the Results (contd.)

- In all three, after presentation of results
  - Discuss implications for the users
  - Discuss how a user can use the results
  - Any additional applications that can benefit from your experiment
  - Present conclusions
    - What did you learn, e.g., surprises, new directions
  - Discuss limitations and future work

Major Steps

1. State Goals
2. List Services and Outcomes
3. Select Appropriate Metrics
4. List the Parameters
5. Select Evaluation Techniques
6. Select Workload
7. Execute
8. Analyze
9. Present the Results