Common Mistakes and How to Avoid Them

Raj Jain
Washington University in Saint Louis
Saint Louis, MO 63130
Jain@cse.wustl.edu

These slides are available on-line at:

http://www.cse.wustl.edu/~jain/cse567-17/

Overview

❑ Common Mistakes in Evaluation
❑ Checklist for Avoiding Common Mistakes
❑ A Systematic Approach to Performance Evaluation
❑ Case Study: Remote Pipes vs RPC

Common Mistakes in Evaluation

1. No Goals
   ➢ No general purpose model
   ➢ Goals $\Rightarrow$ Techniques, Metrics, Workload
   ➢ Not trivial
2. Biased Goals
   ➢ ``To show that OUR system is better than THEIRS''
   ➢ Analysts = Jury
3. Unsystematic Approach
4. Analysis Without Understanding the Problem
5. Incorrect Performance Metrics
6. Unrepresentative Workload
7. Wrong Evaluation Technique

Common Mistakes (Cont)

8. Overlook Important Parameters
9. Ignore Significant Factors
10. Inappropriate Experimental Design
11. Inappropriate Level of Detail
12. No Analysis
13. Erroneous Analysis
14. No Sensitivity Analysis
15. Ignoring Errors in Input
16. Improper Treatment of Outliers
17. Assuming No Change in the Future
18. Ignoring Variability
19. Too Complex Analysis

Common Mistakes (Cont)

20. Improper Presentation of Results
21. Ignoring Social Aspects
22. Omitting Assumptions and Limitations

Checklist for Avoiding Common Mistakes

1. Is the system correctly defined and the goals clearly stated?
2. Are the goals stated in an unbiased manner?
3. Have all the steps of the analysis followed systematically?
4. Is the problem clearly understood before analyzing it?
5. Are the performance metrics relevant for this problem?
6. Is the workload correct for this problem?
7. Is the evaluation technique appropriate?
8. Is the list of parameters that affect performance complete?
9. Have all parameters that affect performance been chosen as factors to be varied?

Checklist (Cont)

10. Is the experimental design efficient in terms of time and results?
11. Is the level of detail proper?
12. Is the measured data presented with analysis and interpretation?
13. Is the analysis statistically correct?
14. Has the sensitivity analysis been done?
15. Would errors in the input cause an insignificant change in the results?
16. Have the outliers in the input or output been treated properly?
17. Have the future changes in the system and workload been modeled?
18. Has the variance of input been taken into account?

Checklist (Cont)

19. Has the variance of the results been analyzed?
20. Is the analysis easy to explain?
21. Is the presentation style suitable for its audience?
22. Have the results been presented graphically as much as possible?
23. Are the assumptions and limitations of the analysis clearly documented?

A Systematic Approach to Performance Evaluation

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

Repeat

Case Study: Remote Pipes vs RPC

❑ **System Definition:**
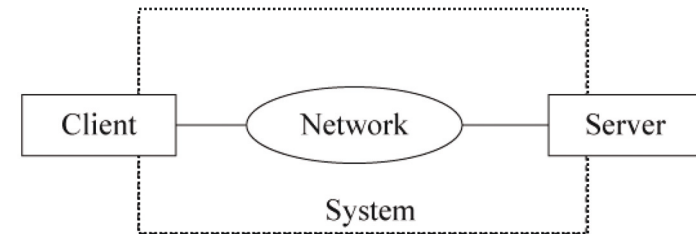
❑ **Services:** Small data transfer or large data transfer.

Case Study (Cont)

❑ **Metrics:**
   ❑ No errors and failures. Correct operation only.
   ❑ Rate, Time, Resource per service.
   ❑ Resource = Client, Server, Network

This leads to:
➢ Elapsed time per call.
➢ Maximum call rate per unit of time, or equivalently, the time required to complete a block of n successive calls.
➢ Local CPU time per call.
➢ Remote CPU time per call.
➢ Number of bytes sent on the link per call.

Case Study (Cont)

❑ **System Parameters:**
➢ Speed of the local CPU.
➢ Speed of the remote CPU.
➢ Speed of the network.
➢ Operating system overhead for interfacing with the channels.
➢ Operating system overhead for interfacing with the networks.
➢ Reliability of the network affecting the number of retransmissions required.

Case Study (Cont)

❑ **Workload parameters:**

- ➢ Time between successive calls.
- ➢ Number and sizes of the call parameters.
- ➢ Number and sizes of the results.
- ➢ Type of channel.
- ➢ Other loads on the local and remote CPUs.
- ➢ Other loads on the network.

Case Study (Cont)

❑ **Factors:**

- ➢ Type of channel: Remote pipes and remote procedure calls
- ➢ Size of the Network: Short distance and long distance
- ➢ Sizes of the call parameters: small and large.
- ➢ Number n of consecutive calls=Block size: 1, 2, 4, 8, 16, 32, …, 512, and 1024.

Note:

- ➢ Fixed: type of CPUs and operating systems.
- ➢ Ignore retransmissions due to network errors
- ➢ Measure under no other load on the hosts and the network.

Case Study (Cont)

❑ **Evaluation Technique:**

- ➢ Prototypes implemented ⇒ Measurements.
- ➢ Use analytical modeling for validation.

❑ **Workload:**

- ➢ Synthetic program generating the specified types of channel requests.
- ➢ Null channel requests
  ⇒ Resources used in monitoring and logging.

❑ **Experimental Design:**

- ➢ A full factorial experimental design with $2^3 \times 11 = 88$ experiments will be used.
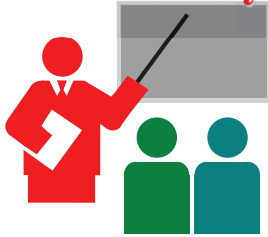
Case Study (Cont)

❑ **Data Analysis:**

- ➢ Analysis of Variance (ANOVA) for the first three factors
- ➢ Regression for number n of successive calls.

❑ **Data Presentation:**

- ➢ The final results will be plotted as a function of the block size n.

Summary

❑ The analysis technique, metrics, workloads depend upon the goal of the study

❑ Metrics are based on services provided by the system

❑ System and workload parameters determine the right set of experiments

❑ Correct analysis and presentation of results is important

Homework 2

❑ Read chapters 2

❑ Submit answers to
  ➢ Exercise 2.2 assuming the system is a personal computer
  ➢ The solution should be limited to 3 pages.
  ➢ Submit by email to jain@wustl.edu with subject "CSE567M Homework 2"

❑ Due: Next Monday

Common Mistakes in Homework 2

❑ Not defining the system

❑ List of metrics not based on services

❑ Mixing system and workload parameters

❑ Factors not in the list of parameters

❑ System parameters not in the list of factors

❑ Levels of factors not specified or specific

❑ Workload is the input not the method

Google Search Modifiers

❑ filetype:pdf, doc, ppt, pptx
❑ site:wustl.com
❑ intitle:trend
❑ inurl:trend
❑ allintitle:Networking Trends
❑ Allinurl:
❑ " " ⇒ Exact Phrase
❑ OR
❑ AND
❑ + ⇒ Must include
❑ - ⇒ Not include
❑ ~X ⇒ X or similar
❑ * ⇒ Wildcard

Google Search (Cont)

- Google search, http://en.wikipedia.org/wiki/Google_Search
- How to search Google, http://www.wikihow.com/Search-Google
- Google Guide Quick reference: Google advance operators cheat sheet, http://www.googleguide.com/advanced_operators_reference.html
- Search Tips & Tricks –Inside Search – Google, http://www.google.com/insidesearch/tipstricks/all.html
- 12 Quick tips to search Google like an expert, http://blog.hubspot.com/blog/tabid/6307/bid/1264/12-Quick-Tips-To-Search-Google-Like-An-Expert.aspx
- Basic search help - web search help, http://www.google.com/support/websearch/bin/answer.py?hl=en&answer=134479&rd=1
- More search help - web search help, http://www.google.com/support/websearch/bin/answer.py?hl=en&answer=136861&topic=1221265
- Search results options, http://www.google.com/support/websearch/bin/answer.py?hl=en&answer=142143&topic=1221265
- Search preferences, http://www.google.com/support/websearch/bin/answer.py?hl=en&answer=35892&rd=1

Project Homework 1

- Search web pages, books, and journal articles from IEEE XPlorer, ACM Digital Library, MOBIUS, Safari books, ILLIAD at Olin Library for **one** of the following topics:
  - Computer Systems Performance Analysis
  - Computer Systems Modeling
  - Computer Systems Simulation
  - Experimental Design
  - Queueing Theory
  - Long Range Dependence
- On the web try the following search points:
  - http://library.wustl.edu/findart.html
  - http://library.wustl.edu/fulltext/
  - http://mobius.umsystem.edu/screens/opacmenu.html
  - http://scholar.google.com
  - http://books.google.com
  - http://dl.acm.org

Project Homework 1 (Cont)

  - http://ieeexplore.ieee.org/Xplore/home.jsp
  - http://searchnetworking.techtarget.com/
- Ignore all entries dated 2012 or before. List others in the following format (up to 5 each):
  - Author, "Title," publisher, year, ISBN, where available. (for 5 books)
  - "Title," URL [One line description] (for 5 web pages)
  - Author, "Title," complete bibliographic data including vol., no., year, pp. (for 5 technical/magazine articles)
  - Title, publisher, URL (for 5 journals/magazines/periodicals)
- For Books (1st item above) and Journals (4th item above) find items that are devoted primarily to your topic and not any item that have the topic as a subset.
- Serially number the references and submit electronically to maede.zolanvari@wustl.edu. The mail should have a subject field of "CSE567M Project Homework 1" (Please note the subject carefully)
- Make a list of other interesting search points and share with the class.20

Common Mistakes in Project Homework #1

- Listing older books
- Listing books/Magazines/journals that have little to do with the topic – may show up in search engines because of a minor mention of the topic or words
- Web Pages – No one line descriptions
- Incomplete bibliographic data for journal articles. Need volume, issue, year, pages.
- Missing journals. Need names of journals dealing with the topic chosen.

Scan This to Download These Slides

Raj Jain
http://rajjain.com

Related Modules

CSE567M: Computer Systems Analysis (Spring 2013),
https://www.youtube.com/playlist?list=PLjGG94etKypJEKjNAa1n_1X0bWWNyZcof

CSE473S: Introduction to Computer Networks (Fall 2011),
https://www.youtube.com/playlist?list=PLjGG94etKypJWOSPMh8Azcgy5e_10TiD

Wireless and Mobile Networking (Spring 2016),
https://www.youtube.com/playlist?list=PLjGG94etKypKeb0nzyN9tSs_HCd5c4wXF

CSE571S: Network Security (Fall 2011),
https://www.youtube.com/playlist?list=PLjGG94etKypKvzfVtutHcPFJXumyyg93u

Video Podcasts of Prof. Raj Jain's Lectures,
https://www.youtube.com/channel/UCN4-5wzNP9-ruOzQMs-8NUw