

King Fahd University of  
Petroleum & Minerals  
Computer Engineering Dept

---

COE 587 – Performance Evaluation And  
Analysis

Term 131

Dr. Ashraf S. Hasan Mahmoud

Rm 22-420

Ext. 1724

Email: ashraf AT kfupm DOT edu DOT sa

9/2/2013

Dr. Ashraf S. Hasan Mahmoud

1

---

**Slides are based on the  
textbook:**

**R. Jain, "Art of Computer  
Systems Performance  
Analysis," Wiley, 1991,  
ISBN:0471503363**

**Book website:**

**<http://www.cse.wustl.edu/~jain/books/perfbook.htm>**

9/2/2013

Dr. Ashraf S. Hasan Mahmoud

2

## Overview of Performance Evaluation

---

- **Examples of “performance” related problems:**
  - **Specifying performance requirements**
  - **Evaluating design alternatives**
  - **Comparing two or more systems**
  - **Determine the optimal value of a parameter**
  - **Finding performance bottleneck**
  - **Characterizing the load of the system**
  - **Determine the number and sizes of components**
  - **Predicting the performance for future loads**
  - **Etc.**

9/2/2013

Dr. Ashraf S. Hasan Mahmoud

3

## Terminologies

---

- **System: any collection of hardware or software, and firmware components**
  - **E.g. CPU, Network link, etc.**
- **Metrics: Criteria used to evaluate the performance of the system**
  - **E.g. total delay in a queue, blocking rate, etc.**
- **Workloads: Requests made the users of the system**
  - **E.g. customer arrival rate**

9/2/2013

Dr. Ashraf S. Hasan Mahmoud

4

## Topics of the Textbook/Course

---

- 1. An Overview of Performance Evaluation**
- 2. Measurement Techniques and Tools**
- 3. Probability Theory and Statistics**
- 4. Experimental Design And Analysis**
- 5. Simulation (with emphasis on discrete-time event driven simulations)**
- 6. Queueing Models**

9/2/2013

Dr. Ashraf S. Hasan Mahmoud

5

## Chapter 1: An Overview of Performance Evaluation

---

### **Example 1.1:**

**What performance metrics should be used to compare the performance of the following systems?**

- (a) Two disk drives**
  - (b) Two transaction processing systems**
  - (c) Two packet retransmission algorithms**
- ++++

- **Focus of part 1 of textbook**
  - **Select appropriate evaluation techniques, performance metrics, and workloads for a system.**

9/2/2013

Dr. Ashraf S. Hasan Mahmoud

6

# Measurement Techniques and Tools

---

## **Example 1.2:**

**What type of monitor (software or hardware) would be more suitable for measuring the following quantities:**

- (a) Number of instructions executed by a processor**
  - (b) Degree of multiprogramming on a timesharing system**
  - (c) Response time of packets on a network**
- ++++

- **Focus of part 2 of textbook**
  - **Conduct performance measurements correctly.**

9/2/2013

Dr. Ashraf S. Hasan Mahmoud

7

# Probability Theory and Statistics

---

## **Example 1.3:**

**The number of packets lost on two links was measured for four file sizes as shown in Table.**

**Which link is better?**

**Table: Packets lost on two links.**

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

++++

- **Focus of part 3 of textbook**
  - **Use of proper statistical techniques to compare several alternatives.**

9/2/2013

Dr. Ashraf S. Hasan Mahmoud

8

# Experimental Design And Analysis

---

## **Example 1.4:**

**The Performance of a system depends on the following three factors:**

- (a) Garbage collection technique used: G1, G2, or none.**
- (b) Type of workload: editing, computing, or artificial intelligence (AI).**
- (c) Type of CPU: C1, C2, or C3.**

++++

- **Focus of part 4 of textbook**
  - **To design measurements and simulation experiments to provide the most information with the least effort.**

9/2/2013

Dr. Ashraf S. Hasan Mahmoud

9

# Simulation

---

## **Example 1.5:**

**In order to compare the performance of two cache replacement algorithms:**

- (a) What type of simulation model should be used?**
- (b) How long should the simulation be run?**
- (c) What can be done to get the same accuracy with shorter run?**
- (d) How can one decide if the random-number generator in the simulation is a good generator?**

++++

- **Focus of part 5 of textbook**
  - **Perform simulations correctly**

9/2/2013

Dr. Ashraf S. Hasan Mahmoud

10

## Queueing Models

### **Example 1.5:**

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 queueing model for the system, determine the following:

- (a) System utilization
- (b) Average service time per query
- (c) Number of queries completed during observation time
- (d) Etc.

++++

- **Focus of part 6 of textbook**
  - Use simple queueing models to analyze the performance of systems

## The Art of Performance Evaluation – cont'd

### **Example 1.7:**

The throughputs of two systems A and B were measured in transactions per second. The results are shown in Table.

**Table: Throughput in transactions per second.**

System	Workload 1	Workload 2
A	20	10
B	10	20

**Which system is better?**

## The Art of Performance Evaluation – cont'd

- Possible answers:

1. Use the average of the two workloads

**Table: Throughput in transactions per second.**

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

2. Normalize with respect to performance of system B

**Table: Throughput in transactions per second.**

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

3. Normalize with respect to performance of system A

**Table: Throughput in transactions per second.**

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

9/2/2013

Dr. Ashraf S. Hasan Mahmoud

13

## Chapter 2: Common Mistakes and How to Avoid Them

- Common mistakes in performance evaluation
- Systematic approach to performance evaluation

9/2/2013

Dr. Ashraf S. Hasan Mahmoud

14

Common Mistakes In Performance Evaluation (Box 2.1 in textbook – page 22)

---

- 1. No goals or biased goals**
- 2. Unsystematic approach**
- 3. Analysis without understanding the problem**
- 4. Incorrect performance metrics**
- 5. Unrepresentative workload**
- 6. Wrong evaluation techniques**
- 7. Overlooking important parameters**
- 8. Ignoring significant factors (parameters versus factors)**
- 9. In appropriate experimental design**
- 10. Inappropriate level of detail**
- 11. No analysis or erroneous analysis**

9/2/2013

Dr. Ashraf S. Hasan Mahmoud

15

Common Mistakes In Performance Evaluation (Box 2.1 in textbook – page 22) – cont'd

---

- 12. No sensitivity analysis (relative importance of parameters)**
- 13. Ignoring errors in input**
- 14. Improper treatment of outliers (what are outliers?)**
- 15. Assuming no change in the future**
- 16. Ignoring variability (to determine variability is a more difficult task compared to determining the mean)**
- 17. Too complex analysis – a simple analysis leading to the same conclusion is preferable**
- 18. Improper presentation of results**
- 19. Ignoring social aspects (writing and speaking skills versus substantive skills)**
- 20. Omitting assumptions and limitations**

9/2/2013

Dr. Ashraf S. Hasan Mahmoud

16



## A Systematic Approach to Performance Evaluation (Box 2.2 in textbook – page 26)

---

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 experiment
9. Analyze and interpret data
10. Present results

What is the difference between parameter and factor?

How is a “workload” related to parameter and/or factor?

9/2/2013

Dr. Ashraf S. Hasan Mahmoud

17

## Case Study 2.1: Remote Procedure Call Project

---

- **Project Statement:**
  - Consider the problem of comparing remote pipes with remote procedure calls. In a procedure call, the calling program is blocked, control is passed to the called procedure along with a few parameters, and when the procedure is complete, the results as well as the control returns to the calling program. A remote procedure call is an extension of this concept to distributed computer systems. A program on one computer system calls a procedure object on another system. The calling program waits until the procedure is complete and the result is returned. Remote pipes are also procedure like objects, but when called, the caller is not blocked. The execution of the pipe occurs concurrently with the continued execution of the caller. The results, if any, are later returned asynchronously.
- **Plan:**
  - Define the system, services, metrics, parameters, etc.
  - Refer to textbook for details.

9/2/2013

Dr. Ashraf S. Hasan Mahmoud

18

## Chapter 3: Selection of Techniques and Metrics

---

- **Topics:**
  - **Selecting an evaluation technique**
  - **Selecting performance metrics**
  - **Commonly used performance metrics**
  - **Utility classification of performance metrics**
  - **Setting performance requirements**

9/2/2013

Dr. Ashraf S. Hasan Mahmoud

19

## Selecting An Evaluation Technique

---

- **There are three techniques for performance evaluation**

**Table: Criteria for selecting an evaluation technique.**

Criterion	Analytical Modeling	Simulations	Measurement
1. Stage	Any	Any	Prototype
2. Time required	Small	Medium	Varies
3. Tools	Analysts	Computer languages	Instrumentation
4. Accuracy*	Low	Moderate	Varies
5. Trade-off evaluation	Easy	Moderate	Difficult
6. Cost	Small	Medium	High
7. Saleability	Low	Medium	High

\* In all cases, results may be misleading or wrong.

9/2/2013

Dr. Ashraf S. Hasan Mahmoud

20

## Selecting An Evaluation Technique - Considerations

- **Generally you want the evaluation to be performed by TWO or MORE techniques**
  - Typically – analysis followed by simulation
- **Until validated, all evaluation results are suspect**
- **Validation rules:**
  - **Simulation results MUST be validated by analysis or measurements**
  - **Analysis results MUST be validated by simulation or measurements**
  - **Measurement results MUST be validated by simulation or modeling**
- **Results should not be counterintuitive**

9/2/2013

Dr. Ashraf S. Hasan Mahmoud

21

## Selecting Performance Metrics

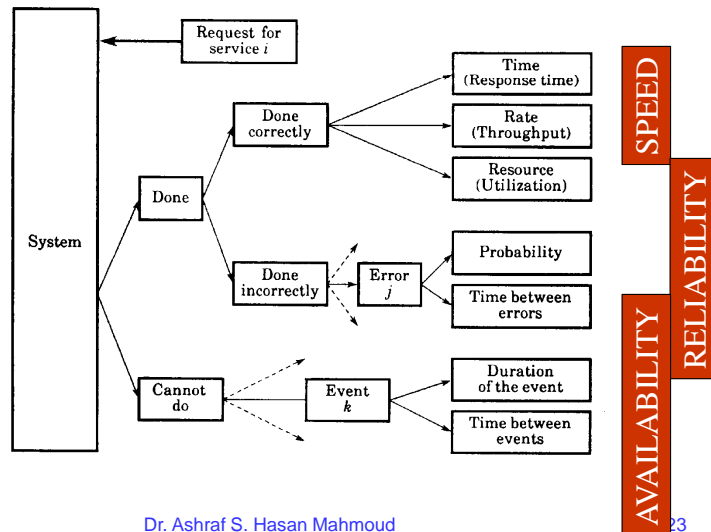
- **List the services for system**
- **For every service numerate performance related metrics**
- **Example: Gateway forwarding packets**
  - **Service – forwarding packets**
  - **Service performed correctly → time for each packet, rate of forwarding, resources consumed**  
→ (time-rate-resource) = (responsiveness, productivity, utilization)
    - Resource with highest utilization is the **BOTTLENECK**
  - **Service performed incorrectly → Error**
    - Classification of errors (prob of each class occurring)
  - **Service not performed → system down, fail, or unavailable**
    - Classification of failures

9/2/2013

Dr. Ashraf S. Hasan Mahmoud

22

## Selecting Performance Metrics – cont'd



9/2/2013

Dr. Ashraf S. Hasan Mahmoud

23

## Selecting Performance Metrics – Considerations

- **Individual versus global**
  - Example – user metrics versus network metrics
  - Maybe conflicting
- **Low-variability – helps reduce the # of repetitions required to obtain a given level of statistical confidence (chapter 12)**
  - Metrics that are ratios of two variables have a larger variability than either of the two variables – try to avoid!!
- **Nonredundancy – two variables convey essentially the same information**
  - Example – average waiting time in a queue and average number of request in a queue – related
- **Completeness – All possible outcomes of system should be reflected in the set of metrics**

9/2/2013

Dr. Ashraf S. Hasan Mahmoud

24

## Jain's Fairness Index

- **Example: network, a multiuser system, etc.**
- **Fairness: Added metric – Variability of throughput across users**
- **For a given set of user throughputs ( $x_1, x_2, \dots, x_n$ ), the fairness index is calculated as**

$$f(x_1, x_2, \dots, x_n) = \frac{\left(\sum_{i=1}^n x_i\right)^2}{n \sum_{i=1}^n x_i^2}$$

- **Note that for  $x_i \geq 0$ ,  $f \in [0,1]$** 
  - **$f = 0 \rightarrow$  no fairness at all**
  - **$f = 1 \rightarrow$  100% fair**

## Jain's Fairness Index - Example

- **Khalaf, M.: Resource Management for OFDMA Systems Using Ant Colony-based Optimization Technique, Master Thesis submitted to Deanship of Graduate Studies, King Fahd University of Petroleum and Minerals (Jan 2012).**

# Jain's Fairness Index - Example - cont'd

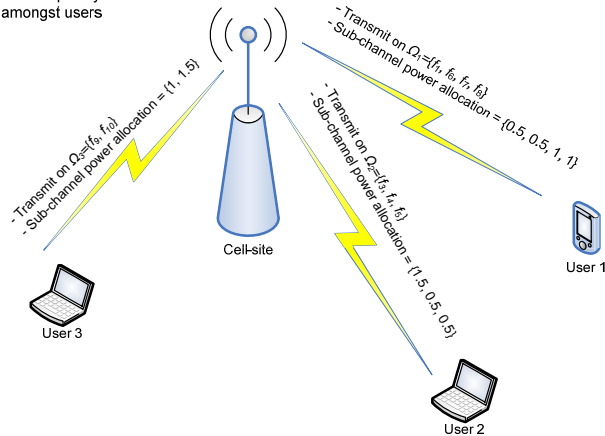
## System Model

### RRM procedure - resources:

- Available sub-channels =  $\{f_1, f_2, \dots, f_{10}\}$
- Total cell-site power budget 10 Watts

### RRM procedure - Task:

- Distribute frequency sub-channels and power amongst users



9/2/2013

Dr. Ashraf S. Hasan Mahmoud

27

# Jain's Fairness Index - Example - Mathematical Model

- Some algorithm is used to obtain  $\Omega_k$  (set of allocated subchannels for kth user) for  $k = 1, 2, \dots, K$  such that  $\Omega_k$  are disjoint for all  $k$ .

- Compute  $p_{k,n}$  that fulfills the following optimization problem is:

$$\max_{p_{k,n}} \sum_{k=1}^K \sum_{n \in \Omega_k} \frac{1}{N} \log_2 (1 + p_{k,n} H_{k,n})$$

Convex optimization problem with some constraints

- Subject to

$$\text{Power constraint} \implies \sum_{k=1}^K \sum_{n \in \Omega_k} p_{k,n} \leq P_{\text{total}} \quad p_{k,n} \geq 0 \quad \text{for all } k, n$$

$$\text{Proportional Fairness} \implies \frac{R_1}{\gamma_1} = \frac{R_2}{\gamma_2} = \dots = \frac{R_K}{\gamma_K}$$

Constraint of interest! How to evaluate compliance?

9/2/2013

VTC 2010 Spring, Taipei - Taiwan

28

## Jain's Fairness Index - Example - cont'd

- **Constraint of interest**  $\frac{R_1}{\gamma_1} = \frac{R_2}{\gamma_2} = \dots = \frac{R_K}{\gamma_K}$

- **Once a solution is obtained, it is not necessary that it satisfies the above constraint 100% - How close?**

- **Use Jain's Fairness Index** 
$$F = \frac{\left( \sum_{k=1}^K R_k / \gamma_k \right)^2}{K \sum_{k=1}^K (R_k / \gamma_k)^2}$$

- **A value for F = 1 → 100% satisfied**
- **A value for F = 0 → not satisfied at all**

9/2/2013

Dr. Ashraf S. Hasan Mahmoud

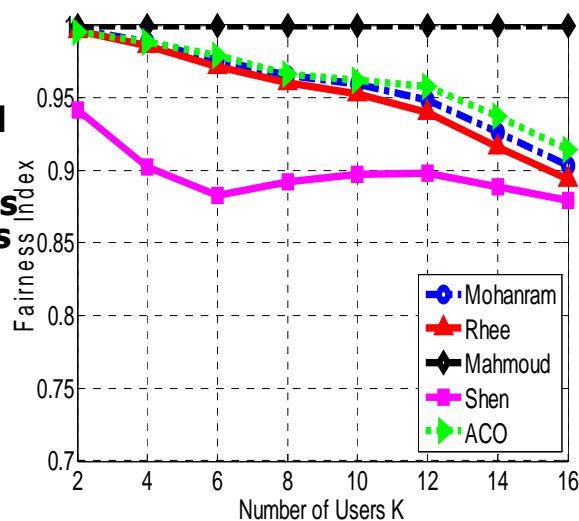
29

## Jain's Fairness Index - Example - Sample Results - cont'd

- **Several algorithm (solutions) are obtained**

- **Note fairness index ranges from 0 to 1**

- **Workload - number of users**



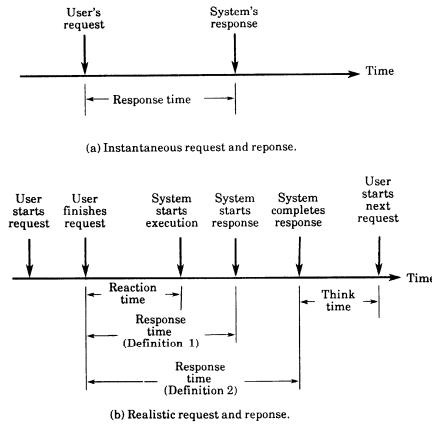
9/2/2013

Dr. Ashraf S. Hasan Mahmoud

30

## Commonly Used Performance Metrics

- **Response time** – interval between user's request and system response
- **Turnaround time** – time between submission till the full completion
- **Reaction time** – Time till system starts execution
- **Stretch factor** – ratio of response time at a particular load to that at the minimum load
- **Throughput** – rate of service for requests
  - MIPS/MFLOPS
  - Packet/sec or b/s
  - Transaction/sec



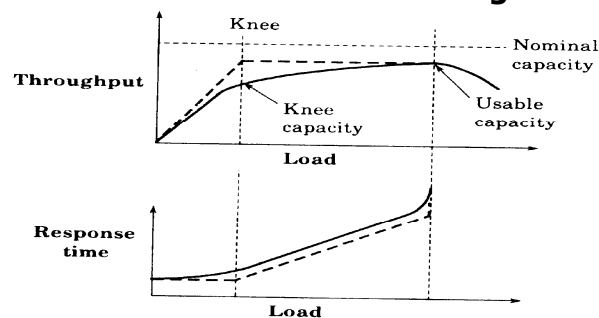
9/2/2013

Dr. Ashraf S. Hasan Mahmoud

31

## Commonly Used Performance Metrics – Capacity

- **Nominal capacity** – maximum achievable throughput under ideal workload conditions
  - Called "bandwidth" for computer networks
- **Usable capacity** – maximum throughput without exceeding a specified response time
- **Knee Capacity** – Throughput for system when response time starts to increase significantly



9/2/2013

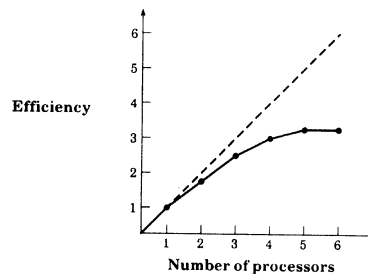
Dr. Ashraf S. Hasan Mahmoud

32



## Commonly Used Performance Metrics – Efficiency/Utilization

- **Efficiency** – maximum achievable throughput (usable capacity) to nominal capacity.
- **Utilization** – fraction of time the resource is busy serving requests
- **Idle time** – period the resource is not used



9/2/2013

Dr. Ashraf S. Hasan Mahmoud

33

## Commonly Used Performance Metrics – Reliability

- **Reliability** – the probability of error OR mean time between errors (error-free seconds)
- **Availability** – the fraction of the time the system is available to service users
  - **Downtime** – duration the system is not available
  - **Uptime** – duration the system is available
  - **Mean Time To Failures (MTBF)** – mean uptime

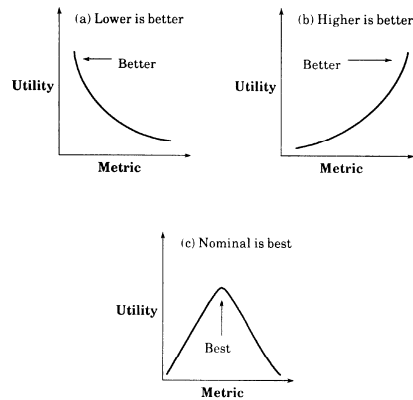
9/2/2013

Dr. Ashraf S. Hasan Mahmoud

34

## Utility Classification of Performance Metrics

- **Higher is better (HB) – e.g. Throughput**
- **Lower is better (LB) – e.g. Response time**
- **Nominal is best (NB) – e.g. Utilization**



9/2/2013

Dr. Ashraf S. Hasan Mahmoud

35

## Setting Performance Requirements

- **Example:**
  - **The system should be both processing and memory efficient. It should not create excessive overhead**
  - **There should be an extremely low probability that the network will duplicate a packet, deliver a packet to the wrong destination, or change data in a packet.**
- **Unacceptable requirements**
  - **Nonspecific**
  - **Nonmeasurable**
  - **Nonacceptable**
  - **Nonrealizable**
  - **Nonthorough**
- **Solution: requirements must be SMART → (S)pecific, (M)easurable, (A)cceptable, (R)ealizable, and (T)horough!**

9/2/2013

Dr. Ashraf S. Hasan Mahmoud

36

## Example Study – BGP Performance Evaluation

- **AlRefai, A.: BGP-based solutions for international ISP blocking. Master Thesis submitted to Deanship of Graduate Studies, King Fahd University of Petroleum and Minerals (2010)**

9/2/2013

Dr. Ashraf S. Hasan Mahmoud

37

## Example Study – BGP Performance Evaluation – cont'd

- **Example:**
- **AS12 – autonomous system of interest**
- **AS4 – Main IISP**
- **AS4 – Secondary IISP**
- **Comm between LAN\_West and LAN\_East**



Fig. 11 OPNET simulation model

9/2/2013

Dr. Ashraf S. Hasan Mahmoud

38

## Example Study – BGP Performance Evaluation – cont'd

- **Selected mitigation techniques**
    - AS-path shortening
    - More-specific prefixes
    - Use of communities
  - **Types of Traffic**
    - HTTP
    - FTP
    - VOIP
  - **Background load**
    - 20%
    - 50%
  - **Internet delay**
    - 0.1 sec
    - 5 sec
  - **Etc**
- **Performance metrics**
    - Network convergence time
    - Throughput at router/link X
    - Packet drop
    - Page response time (HTTP)
    - Etc.

9/2/2013

Dr. Ashraf S. Hasan Mahmoud

39

## Example Study – BGP Performance Evaluation – cont'd

- **Sample result**
- **Note the confidence intervals!**

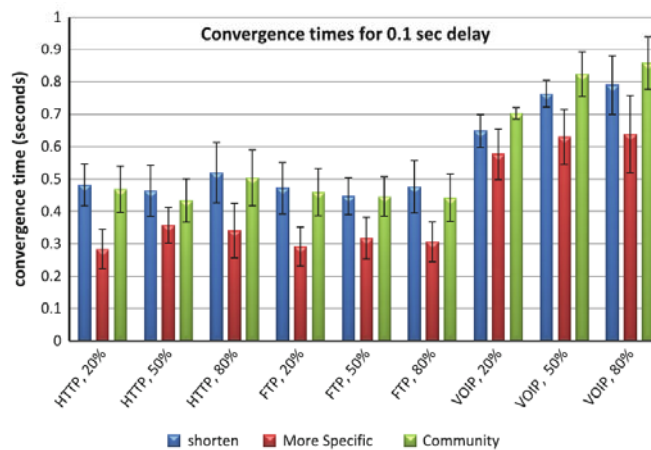


Fig. 12 Convergence time for BGP-based solutions assuming 100 msec average internet delay

9/2/2013

Dr. Ashraf S. Hasan Mahmoud

40