The Art of Workload Selection

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/cse567-08/
Overview

- Services Exercised
  - Example: Timesharing Systems
  - Example: Networks
  - Example: Magnetic Tape Backup System
- Level of Detail
- Representativeness
- Timeliness
- Other Considerations in Workload Selection
The Art of Workload Selection

Considerations:

- Services exercised
- Level of detail
- Loading level
- Impact of other components
- Timeliness
Services Exercised

- **SUT** = System Under Test
- **CUS** = Component Under Study
Services Exercised (Cont)

- Do not confuse SUT w CUS
- Metrics depend upon SUT: MIPS is ok for two CPUs but not for two timesharing systems.
- Workload: depends upon the system.
- Examples:
  - CPU: instructions
  - System: Transactions
  - Transactions not good for CPU and vice versa
  - Two systems identical except for CPU
    - Comparing Systems: Use transactions
    - Comparing CPUs: Use instructions
  - Multiple services: Exercise as complete a set of services as possible.
Example: Timesharing Systems

- Applications
  ⇒ Application benchmark
- Operating System
  ⇒ Synthetic Program
- Central Processing Unit
  ⇒ Instruction Mixes
- Arithmetic Logical Unit
  ⇒ Addition instruction
Example: Networks

Mail, file transfer, virtual terminal, etc.

Applications

Data Compression, etc.

Presentation

Dialogs

Session

Messages

Transport

Packets

Network

Frames

Datalink

Bits

Physical
Example: Magnetic Tape Backup System

- **Backup System:**
  - **Services:** Backup files, backup changed files, restore files, list backed-up files.
  - **Factors:** File-system size, batch or background process, incremental or full backups.
  - **Metrics:** Backup time, restore time.
  - **Workload:** A computer system with files to be backed up. Vary frequency of backups.

- **Tape Data System:**
  - **Services:** Read/write to the tape, read tape label, auto load tapes.
  - **Factors:** Type of tape drive.
  - **Metrics:** Speed, reliability, time between failures.
  - **Workload:** A synthetic program generating representative tape I/O requests.
Magnetic Tape System (Cont)

- **Tape Drives:**
  - Services: Read record, write record, rewind, find record, move to end of tape, move to beginning of tape.
  - Factors: Cartridge or reel tapes, drive size.
  - Metrics: Time for each type of service, for example, time to read record and to write record, speed (requests/time), noise, power dissipation.
  - Workload: A synthetic program exerciser generating various types of requests in a representative manner.

- **Read/Write Subsystem:**
  - Services: Read data, write data (as digital signals).
  - Factors: Data-encoding technique, implementation technology (CMOS, TTL, and so forth).
  - Metrics: Coding density, I/O bandwidth (bits per second).
Magnetic Tape System (Cont)

- Workload: Read/write data streams with varying patterns of bits.

- Read/Write Heads:
  - Services: Read signal, write signal (electrical signals).
  - Factors: Composition, inter-head spacing, gap sizing, number of heads in parallel.
  - Metrics: Magnetic field strength, hysteresis.
  - Workload: Read/write currents of various amplitudes, tapes moving at various speeds.
Level of Detail

- Most frequent request:
  - Examples: Addition Instruction, Debit-Credit, Kernels
  - Valid if one service is much more frequent than others

- Frequency of request types
  - Examples: Instruction mixes
  - Context sensitivity $\Rightarrow$ Use set of services
  - History-sensitive mechanisms (caching) $\Rightarrow$ Context sensitivity

- Time-stamped sequence of requests
  - May be too detailed
  - Not convenient for analytical modeling
  - May require exact reproduction of component behavior
Level of Detail (Cont)

- Average resource demand
  - Used for analytical modeling
  - Grouped similar services in classes

- Distribution of resource demands
  - Used if variance is large
  - Used if the distribution impacts the performance

- Workload used in simulation and analytical modeling:
  - Non executable: Used in analytical/simulation modeling
  - Executable workload: can be executed directly on a system
Representativeness

The test workload and real workload should have the same:

- Elapsed Time
- Resource Demands
- Resource Usage Profile: Sequence and the amounts in which different resources are used.
Timeliness

- Users are a moving target.
- New systems $\implies$ new workloads
- Users tend to optimize the demand.
- Fast multiplication $\implies$ Higher frequency of multiplication instructions.
- Important to monitor user behavior on an ongoing basis.
Other Considerations in Workload Selection

- Loading Level: A workload may exercise a system to its:
  - Full capacity (best case)
  - Beyond its capacity (worst case)
  - At the load level observed in real workload (typical case).
  - For procurement purposes ⇒ Typical
  - For design ⇒ best to worst, all cases

- Impact of External Components:
  - Do not use a workload that makes external component a bottleneck ⇒ All alternatives in the system give equally good performance.

- Repeatability
Services exercised determine the workload
Level of detail of the workload should match that of the model being used
Workload should be representative of the real systems usage in recent past
Loading level, impact of external components, and repeatability or other criteria in workload selection
Exercise 5.1

What metric and workload would you choose to compare:

a. Two systems with similar functionality: IBM PC versus MAC
b. Two systems for very different applications: PC versus Workstations
c. Two systems with identical functionality: IBM PC versus Dell PC
d. Two versions of the same operating systems: Windows 98 vs Windows XP
e. Two hardware components: Two floppy drives
f. Two languages: C vs. Pascal

One metric and one workload is sufficient
Exercise 5.2

- Select an area of computer systems, for example, databases, networks, processors, and so on. Prepare a table identifying increasing levels of services, components, factors, and workloads.
Homework

- Read chapters 4 and 5
- Submit answer to Exercise 5.1