Workload Characterization Techniques

Adapted by Prof. Gheith Abandah

Overview

❑ Terminology

❑ Components and Parameter Selection

❑ Workload Characterization Techniques: Averaging, Single Parameter Histograms, Multi-parameter Histograms, Markov Models, Clustering

❑ Clustering Method: Minimum Spanning Tree, Nearest Centroid

❑ Problems with Clustering

Terminology

❑ **User** = Entity that makes the service request

❑ **Workload components**:

➢ Applications

➢ Sites

➢ User Sessions

❑ **Workload parameters** or workload features: Measured quantities, service requests, or resource demands.
For example: transaction types, instructions, packet sizes, source-destinations of a packet, and page reference pattern.

Components and Parameter Selection

❑ The workload component should be at the SUT interface.

❑ Each component should represent as homogeneous a group as possible. Combining very different users into a site workload may not be meaningful.

❑ Domain of the control affects the component:
Example: mail system designer are more interested in determining a typical mail session than a typical user session.

❑ Do not use parameters that depend upon the system, e.g., the elapsed time, CPU time.

Components (Cont)

❑ Characteristics of service requests:

➢ Arrival Time

➢ Type of request or the resource demanded

➢ Duration of the request

➢ Quantity of the resource demanded, for example, pages of memory

❑ Exclude those parameters that have little impact.

Workload Characterization Techniques

1. Averaging
2. Single-Parameter Histograms
3. Multi-parameter Histograms
4. Markov Models
5. Clustering

Averaging

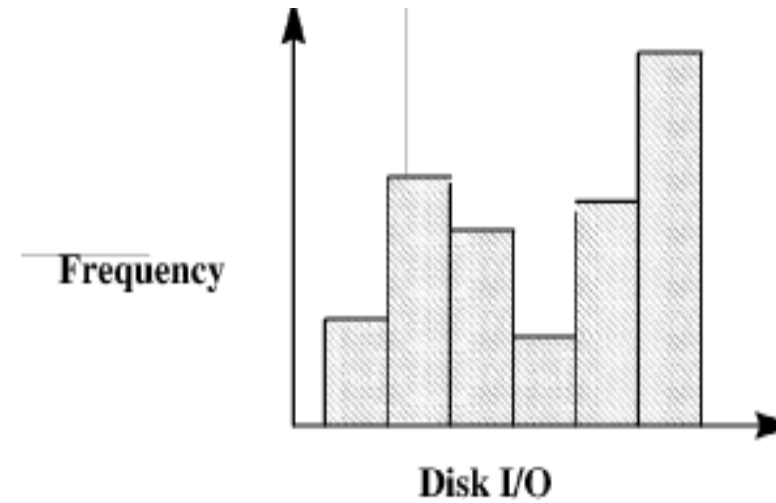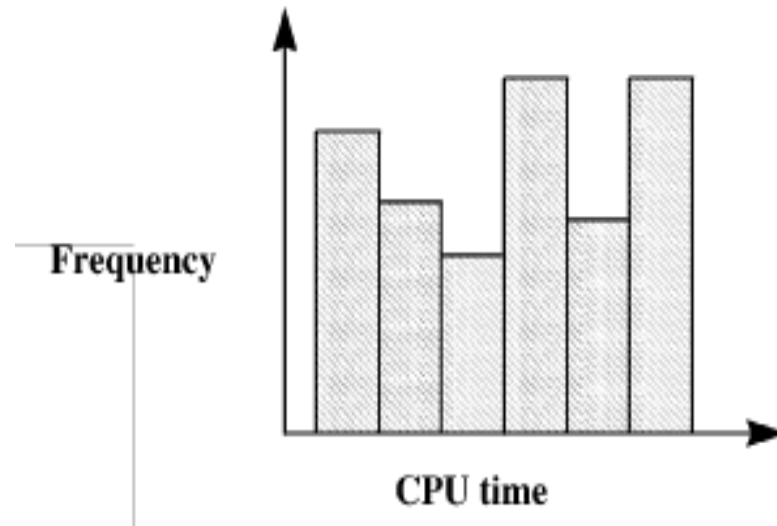❑ Mean $$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

❑ Mode (for categorical variables): Most frequent value
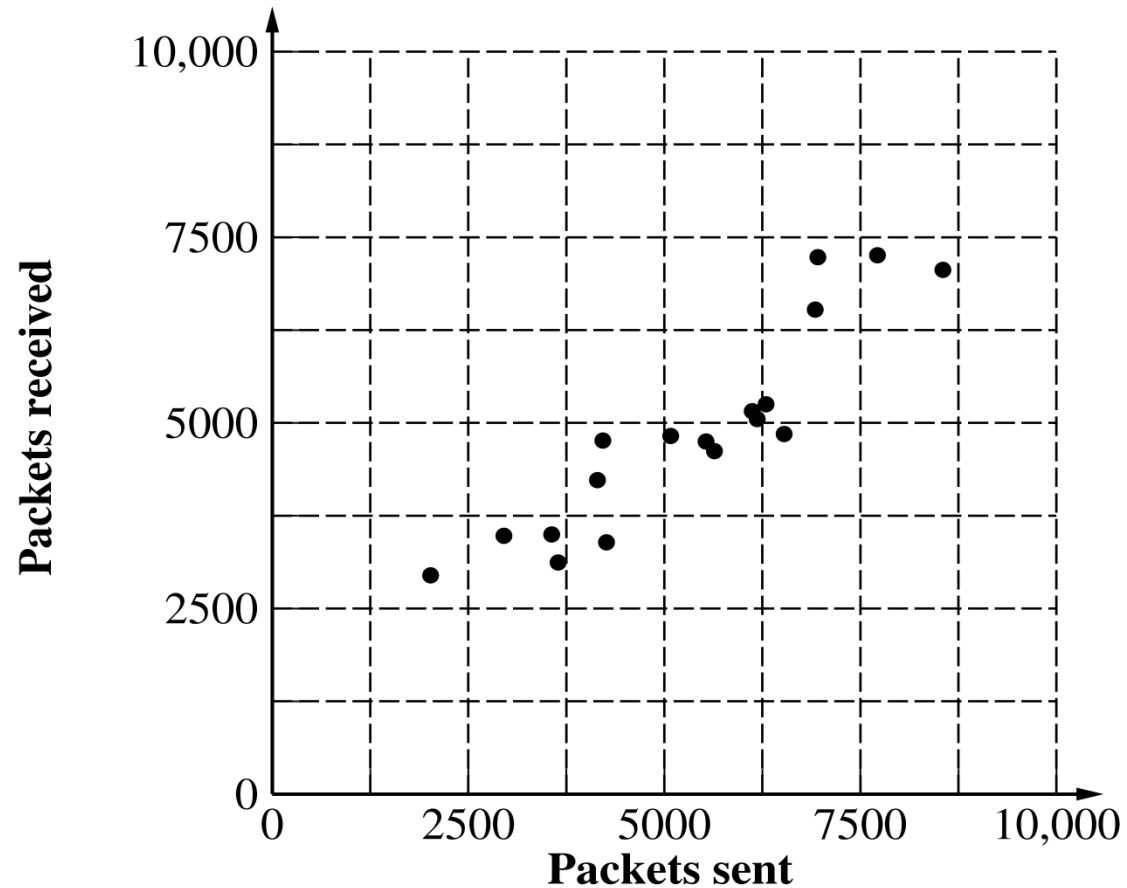
❑ Median: 50-percentile

Specifying Dispersion

❑ Standard deviation s:

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2$$

❑ Coefficient Of Variation: $s/\bar{x}$

❑ Minimum and Maximum

Single Parameter Histograms

❑ Use only if the variance is high.

❑ Ignores correlation among parameters.

Multi-parameter Histograms

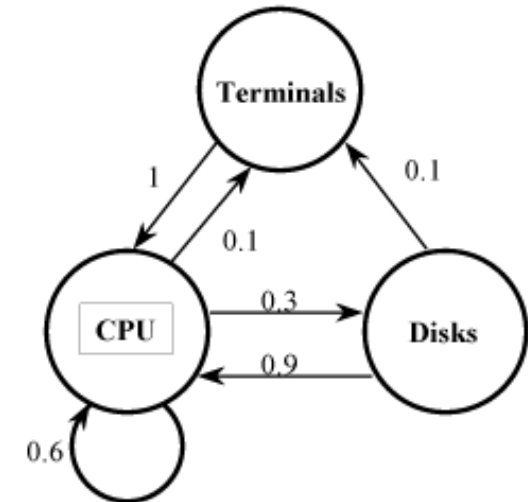❑ Difficult to plot joint histograms for more than two parameters.
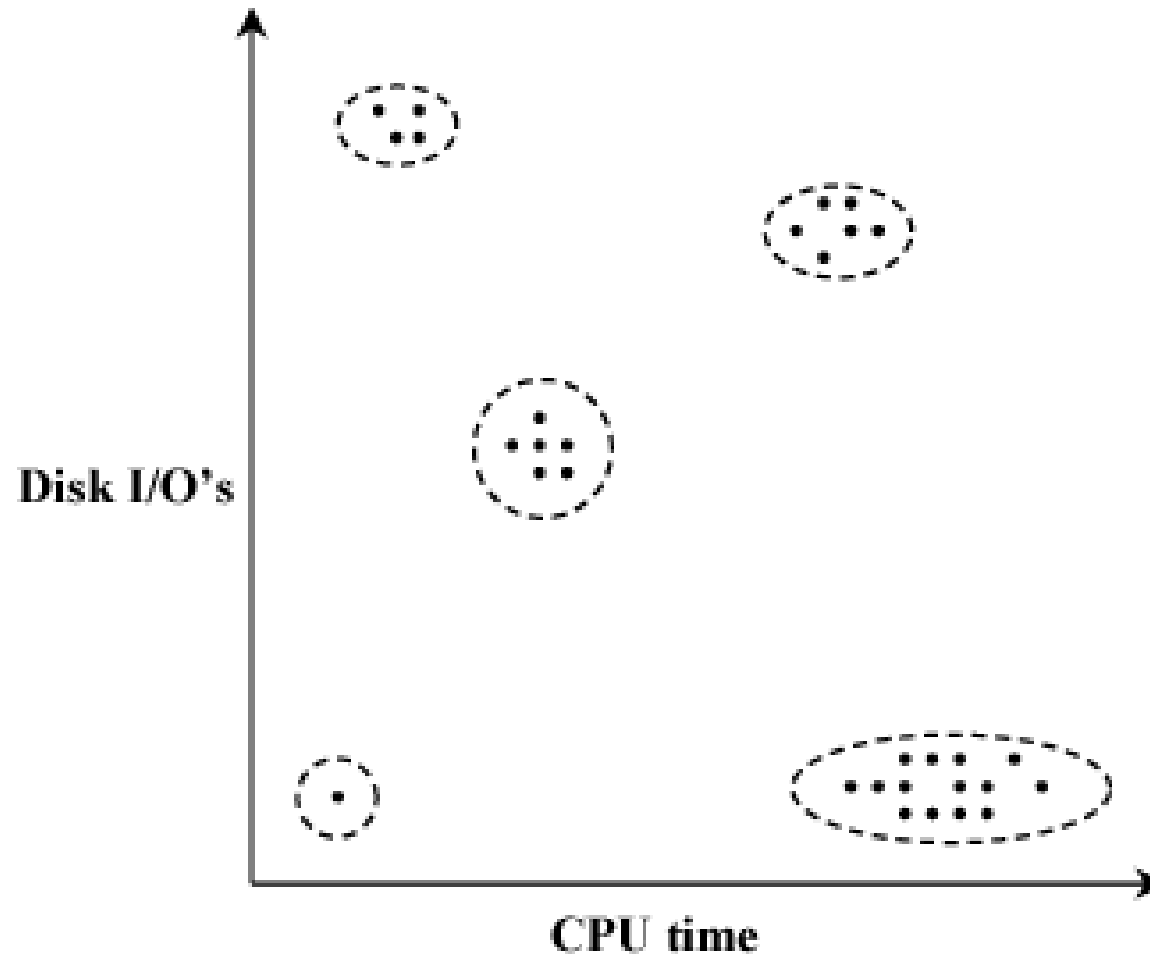
Markov Models

❑ Markov

⟹ the next request depends only on the last request

❑ Described by a transition matrix:

| From/To	CPU	Disk	Terminal
CPU	0.6	0.3	0.1
Disk	0.9	0	0.1
Terminal	1	0	0

Clustering

Clustering Steps

1. Take a sample, that is, a subset of workload components.
2. Select workload parameters.
3. Select a distance measure.
4. Remove outliers.
5. Scale all observations.
6. Perform clustering.
7. Interpret results.
8. Change parameters, or number of clusters, and repeat steps 3-7.
9. Select representative components from each cluster.

Summary
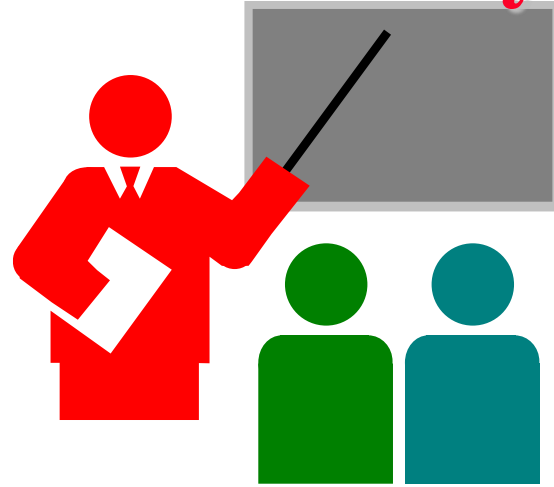
❑ Workload Characterization = Models of workloads

❑ Averaging, Single parameter histogram, multi-parameter histograms, …

❑ Principal component analysis consists of finding parameter combinations that explain the most variation

❑ Clustering: divide workloads in groups that can be represented by a single benchmark