Computer Systems Performance Analysis and Benchmarking (37-235)

Analytic Modeling

Simulation

Measurements / Benchmarking

Lecture by:

Prof. Thomas Stricker

Assignments/Projects:

Christian Kurmann

Textbook:

Raj Jain, “The Art of Computer Systems Performance Analysis”, 1991 Wiley & Sons, New York

Topic of Today:

- **Workload Characterization**
- **Monitors**
- **Logging**

Workload Characterization

- services requested from SUT
- resource demands of users
- users (workload unit)
    - humans
    - programs
    - appliances
- workload components
    - applications
    - sites
    - user sessions (logged)
- workload parameters
    - transaction types
    - packet sizes
    - source/dest of packets
    - instructions

Ways to specify workloads:

- Averaging
- Specifying dispersion (C.O.V)
- Single parameter histogram
- Multi parameter histogram
- Principle component analysis
- Markov models
- Clustering

Examples

Averages/C.O.Vs

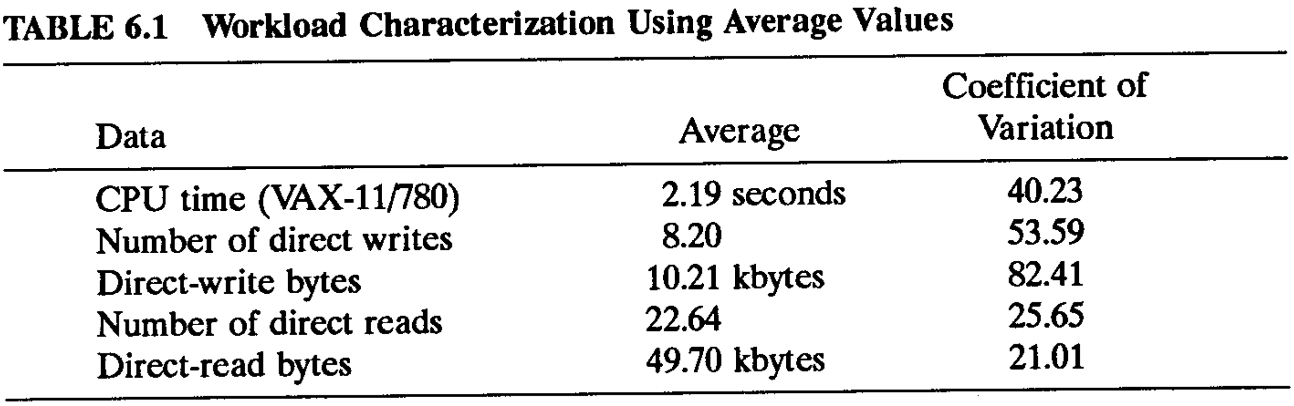


TABLE 6.1 Workload Characterization Using Average Values

Data	Average	Coefficient of Variation
CPU time (VAX-11/780)	2.19 seconds	40.23
Number of direct writes	8.20	53.59
Direct-write bytes	10.21 kbytes	82.41
Number of direct reads	22.64	25.65
Direct-read bytes	49.70 kbytes	21.01

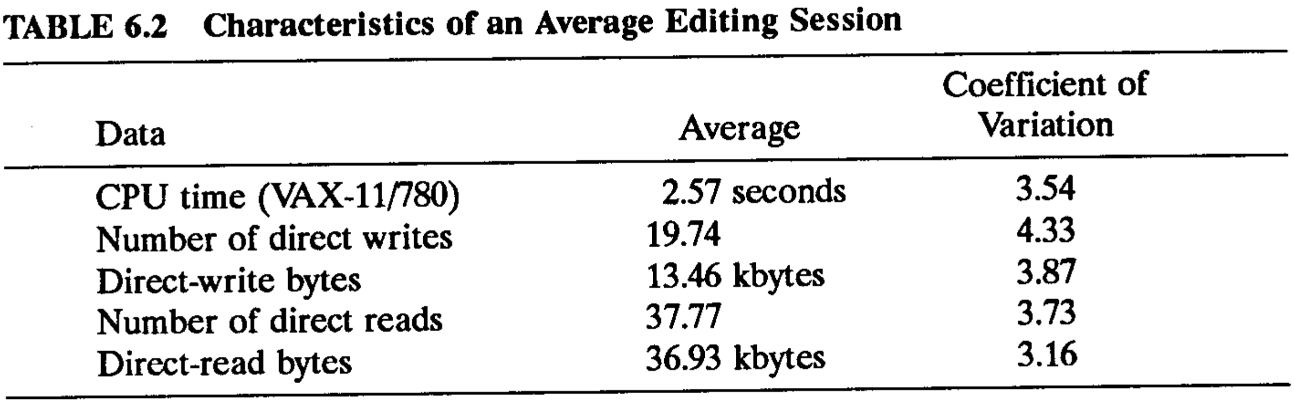


TABLE 6.2 Characteristics of an Average Editing Session

Data	Average	Coefficient of Variation
CPU time (VAX-11/780)	2.57 seconds	3.54
Number of direct writes	19.74	4.33
Direct-write bytes	13.46 kbytes	3.87
Number of direct reads	37.77	3.73
Direct-read bytes	36.93 kbytes	3.16

Histograms

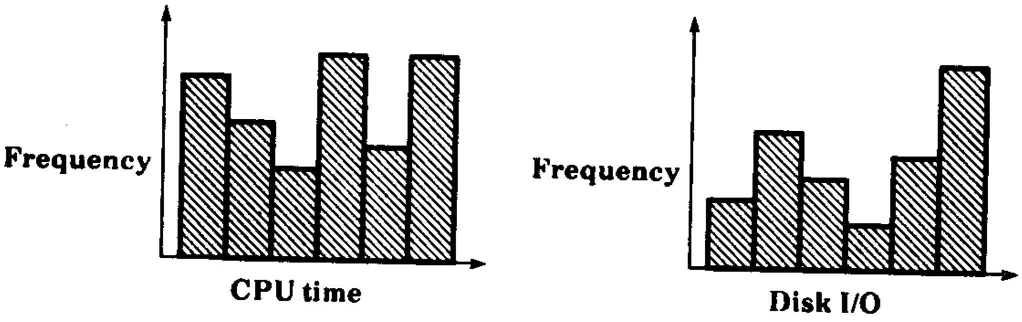


FIGURE 6.1 Single-parameter histograms of CPU time and disk I/O.

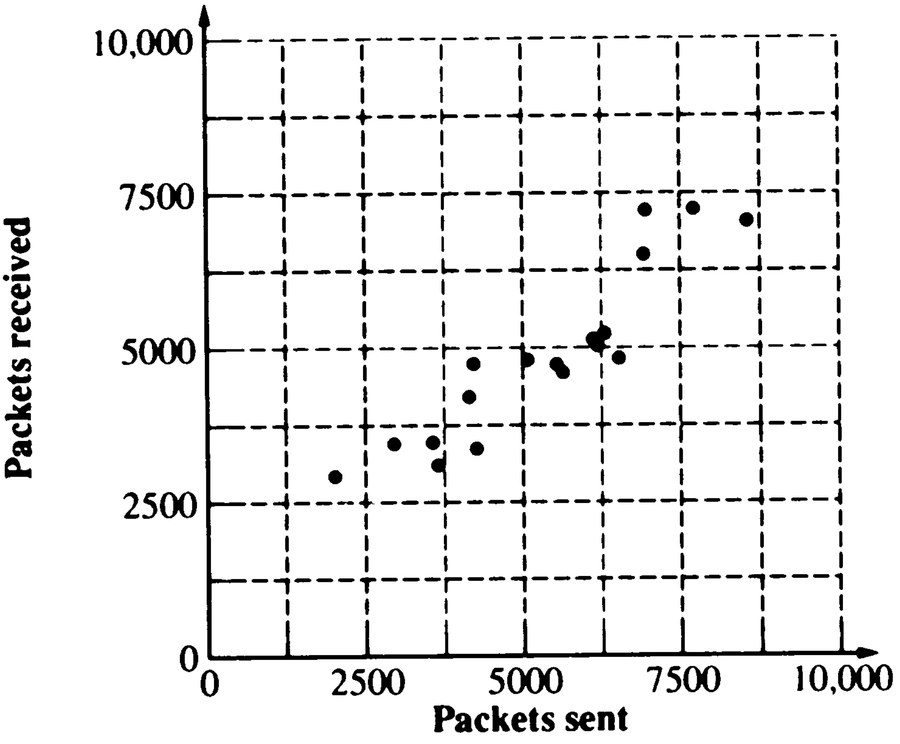


FIGURE 6.2 Two-parameter histogram.

Principle component analysis

Have many factors...but want a single one!

- Resource demands:
  - CPU, RAM, Disk...

  want just:
  - small, medium, large
- Determine some weights a_i and compute a principle factor:

$$y = \sum_{j=1}^{n} a_j x_j$$

- Most program query factors
  - Do it manually by guess.
  - Do it automatically by statistical criteria.

Example (book)

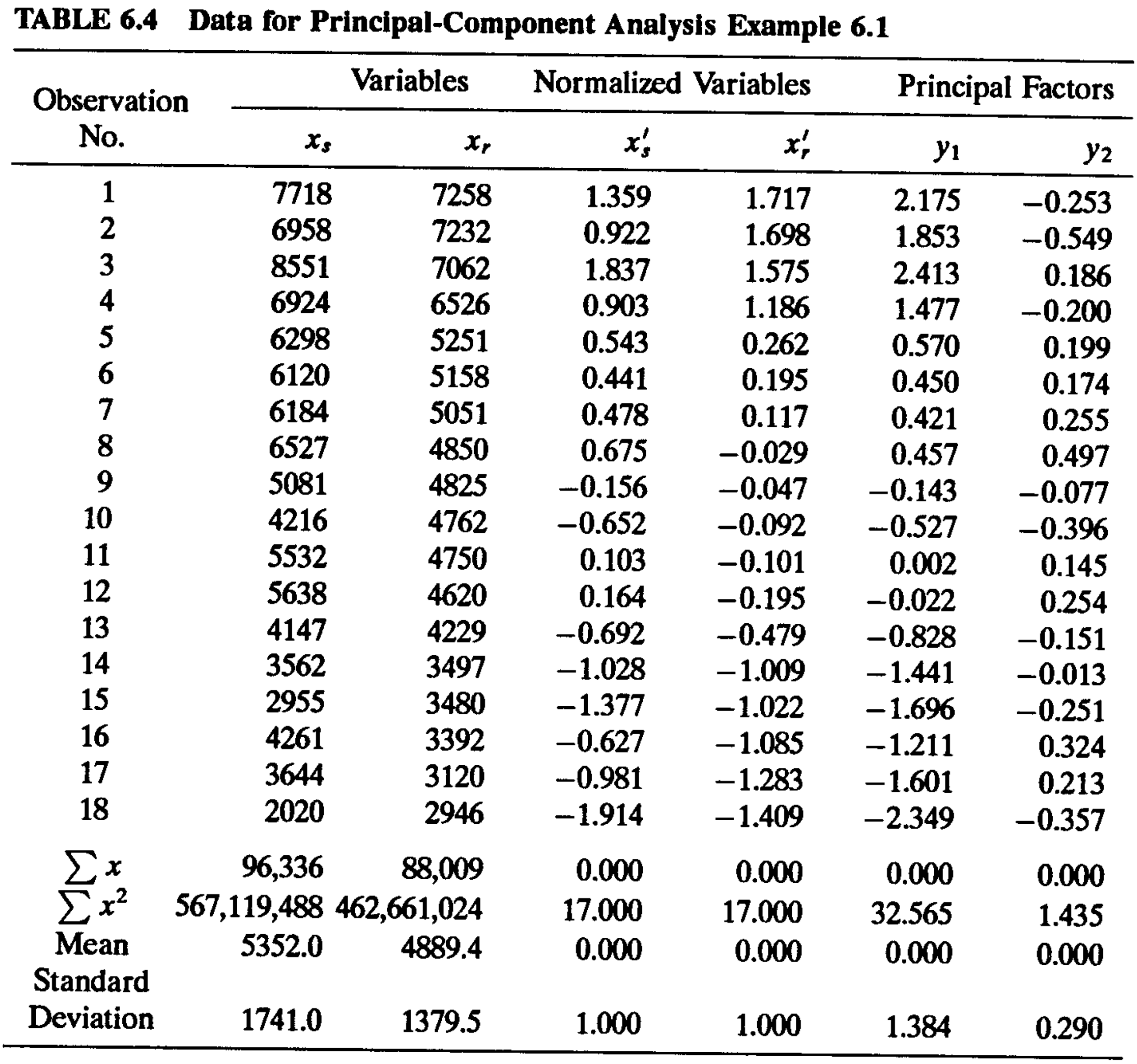


TABLE 6.4 Data for Principal-Component Analysis Example 6.1

Observation No.	Variables		Normalized Variables		Principal Factors	
	x_s	x_r	x'_s	x'_r	y_1	y_2
1	7718	7258	1.359	1.717	2.175	−0.253
2	6958	7232	0.922	1.698	1.853	−0.549
3	8551	7062	1.837	1.575	2.413	0.186
4	6924	6526	0.903	1.186	1.477	−0.200
5	6298	5251	0.543	0.262	0.570	0.199
6	6120	5158	0.441	0.195	0.450	0.174
7	6184	5051	0.478	0.117	0.421	0.255
8	6527	4850	0.675	−0.029	0.457	0.497
9	5081	4825	−0.156	−0.047	−0.143	−0.077
10	4216	4762	−0.652	−0.092	−0.527	−0.396
11	5532	4750	0.103	−0.101	0.002	0.145
12	5638	4620	0.164	−0.195	−0.022	0.254
13	4147	4229	−0.692	−0.479	−0.828	−0.151
14	3562	3497	−1.028	−1.009	−1.441	−0.013
15	2955	3480	−1.377	−1.022	−1.696	−0.251
16	4261	3392	−0.627	−1.085	−1.211	0.324
17	3644	3120	−0.981	−1.283	−1.601	0.213
18	2020	2946	−1.914	−1.409	−2.349	−0.357
$\sum x$	96,336	88,009	0.000	0.000	0.000	0.000
$\sum x^2$	567,119,488	462,661,024	17.000	17.000	32.565	1.435
Mean	5352.0	4889.4	0.000	0.000	0.000	0.000
Standard Deviation	1741.0	1379.5	1.000	1.000	1.384	0.290

1. *Compute the mean and standard deviations of the variables*:

$$\bar{x}_s = \frac{1}{n}\sum_{i=1}^{n} x_{si} = \frac{96{,}336}{18} = 5352.0$$

$$\bar{x}_r = \frac{1}{n}\sum_{i=1}^{n} x_{ri} = \frac{88{,}009}{18} = 4889.4$$

$$\begin{aligned} s_{x_s}^2 &= \frac{1}{n-1}\sum_{i=1}^{n}(x_{si} - \bar{x}_s)^2 \\ &= \frac{1}{n-1}\left[\left(\sum_{i=1}^{n} x_{si}^2\right) - n\bar{x}_s^2\right] \\ &= \frac{567{,}119{,}488 - 18 \times 5352^2}{17} = 1741.0 \end{aligned}$$

Similarly,

$$s_{x_r}^2 = \frac{462{,}661{,}024 - 18 \times 4889.4^2}{17} = 1379.5$$

2. *Normalize the variables to zero mean and unit standard deviation.* The normalized values x'_s and x'_r are given by

$$x'_s = \frac{x_s - \bar{x}_s}{s_{x_s}} = \frac{x_s - 5352}{1741}$$

$$x'_r = \frac{x_r - \bar{x}_r}{s_{x_r}} = \frac{x_r - 4889}{1380}$$

   The normalized values are shown in the fourth and fifth columns of Table 6.4.

3. *Compute the correlation among the variables*:

$$R_{x_s,x_r} = \frac{(1/n)\sum_{i=1}^{n}(x_{si} - \bar{x}_s)(x_{ri} - \bar{x}_r)}{s_{x_s}s_{x_r}} = 0.916$$

4. *Prepare the correlation matrix*:

$$\mathbf{C} = \begin{bmatrix} 1.000 & 0.916 \\ 0.916 & 1.000 \end{bmatrix}$$

5. *Compute the eigenvalues of the correlation matrix.* This is done by solving the characteristic equation. Using I to denote an identity matrix,

$$|\lambda I - C| = \begin{vmatrix} \lambda - 1 & -0.916 \\ -0.916 & \lambda - 1 \end{vmatrix} = 0$$

   or

$$(\lambda - 1)^2 - 0.916^2 = 0$$

   The eigenvalues are 1.916 and 0.084.

6. *Compute the eigenvectors of the correlation matrix.* The eigenvector $\mathbf{q}_1$ corresponding to $\lambda_1 = 1.916$ is defined by the following relationship:

$$\mathbf{C}\mathbf{q}_1 = \lambda_1\mathbf{q}_1$$

   or

$$\begin{bmatrix} 1.000 & 0.916 \\ 0.916 & 1.000 \end{bmatrix} \times \begin{bmatrix} q_{11} \\ q_{21} \end{bmatrix} = 1.916 \begin{bmatrix} q_{11} \\ q_{21} \end{bmatrix}$$

   or

$$q_{11} = q_{21}$$

   Restricting the length of the eigenvector to 1, the following vector is the first eigenvector:

$$\mathbf{q}_1 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}$$

Similarly, the second eigenvector is $$\mathbf{q}_2 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{bmatrix}$$

7. *Obtain principal factors by multiplying the eigenvectors by the normalized vectors*:

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} \frac{x_s - 5352}{1741} \\ \frac{x_r - 4889}{1380} \end{bmatrix}$$

8. *Compute the values of the principal factors.* These are shown in the last two columns of Table 6.4.
9. *Compute the sum and sum of squares of the principal factors.* The sum must be zero. The sum of squares give the percentage of variation explained. In this case, the sums of squares are 32.565 and 1.435. Thus, the first factor explains 32.565/(32.565 + 1.435), or 95.7%, of the variation. The second factor explains only 4.3% of the variation and can thus be ignored.
10. *Plot the values of principal factors.* The results are shown in Figure 6.3. Notice that most of the variation is along the first principal factor. The variation along the second factor is negligible.

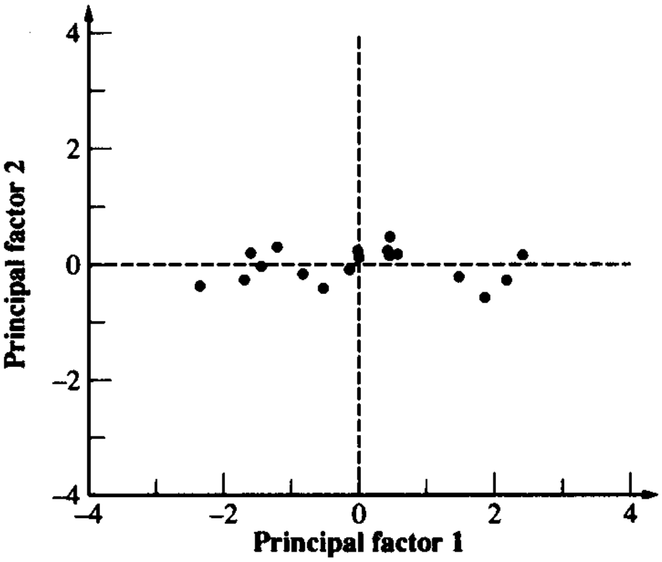


FIGURE 6.3 Packets sent and received data plotted along the principal-component axes.

Markov Models

Limited memory behavior

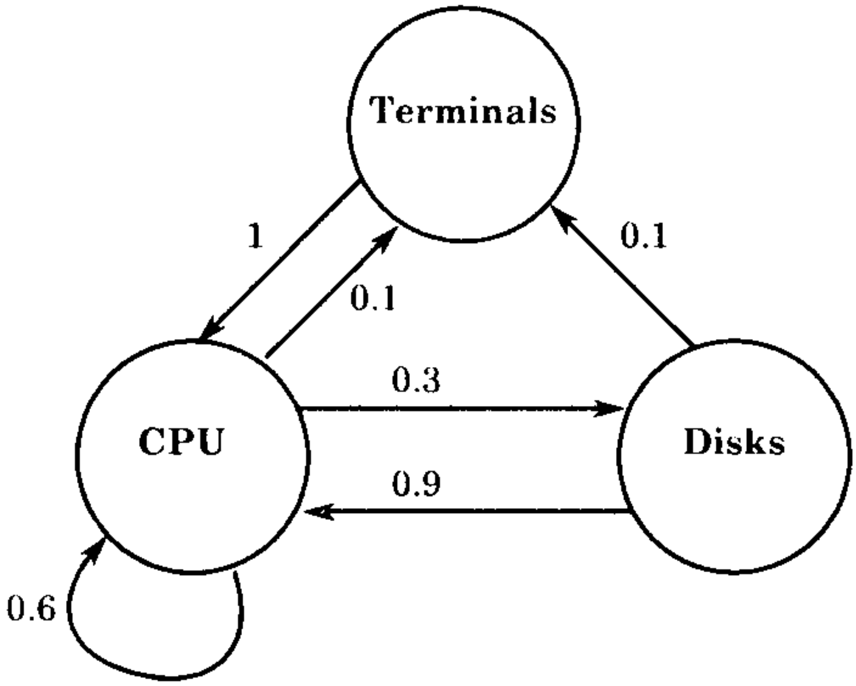
P(Transition to next state) =

f(current state) ≠ f (all previous state)

Example

TABLE 6.5 Transition Probability Matrix

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



Monitors

Used by:

- systems programmer (optimization)
- systems manager (find bottleneck)
- systems manager (tune system)
- systems analyst (characterize work)
- systems analyst (find model params.)

Terms:

- Event = change in system state
- Trace = log of events incl. time
- Overhead = perturbation by monitor
- Domain = activities observed
- Input rate = max frequency of event
- Resolution = coarseness of information
- Input width = number of bit sampled

Monitor Classifications

1. Dimension

- Event driven
- Sampling
- Hybrid

2. Dimension

- On-line
- Batch

Methods:

- Software Monitors
- Hardware Monitors

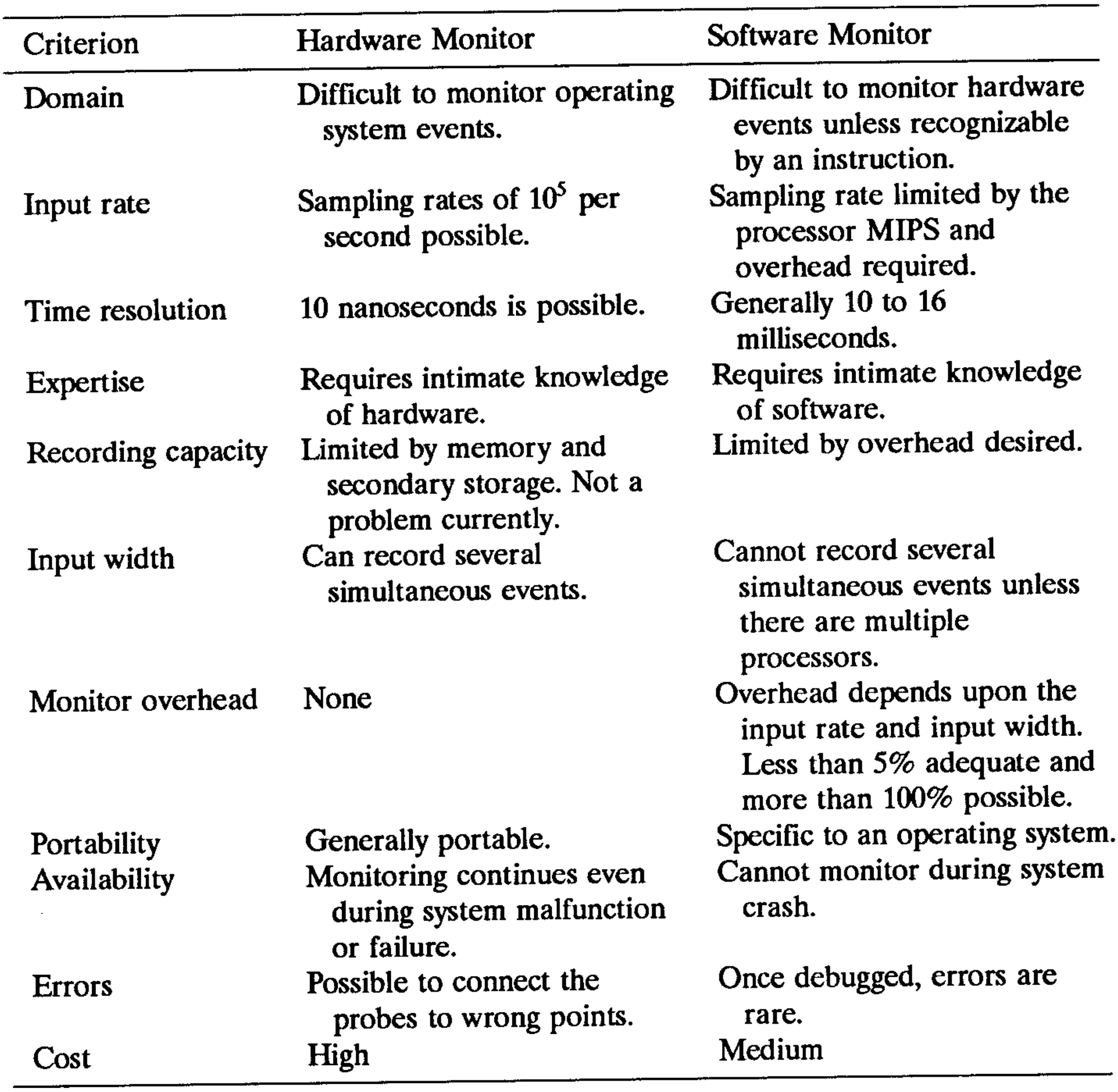
Design Issues (Software)

- Activation Mechanism
- Trap instruction / Trace mode
- Timer Interrupt
- Buffer size / Number of buffers
- Buffer overflow
- Data compression/analysis on the fly
- On/Off switch / Language
- Priority / Monitoring abnormal events

Design Issues (Hardware)

- Probes / Counters
- Logic Elements / Comparators
- Mapping Hardware
- Timer
- Tape/Disks

Software vs. Hardware



Criterion	Hardware Monitor	Software Monitor
Domain	Difficult to monitor operating system events.	Difficult to monitor hardware events unless recognizable by an instruction.
Input rate	Sampling rates of 10^5 per second possible.	Sampling rate limited by the processor MIPS and overhead required.
Time resolution	10 nanoseconds is possible.	Generally 10 to 16 milliseconds.
Expertise	Requires intimate knowledge of hardware.	Requires intimate knowledge of software.
Recording capacity	Limited by memory and secondary storage. Not a problem currently.	Limited by overhead desired.
Input width	Can record several simultaneous events.	Cannot record several simultaneous events unless there are multiple processors.
Monitor overhead	None	Overhead depends upon the input rate and input width. Less than 5% adequate and more than 100% possible.
Portability	Generally portable.	Specific to an operating system.
Availability	Monitoring continues even during system malfunction or failure.	Cannot monitor during system crash.
Errors	Possible to connect the probes to wrong points.	Once debugged, errors are rare.
Cost	High	Medium

Distributed Systems Monitor

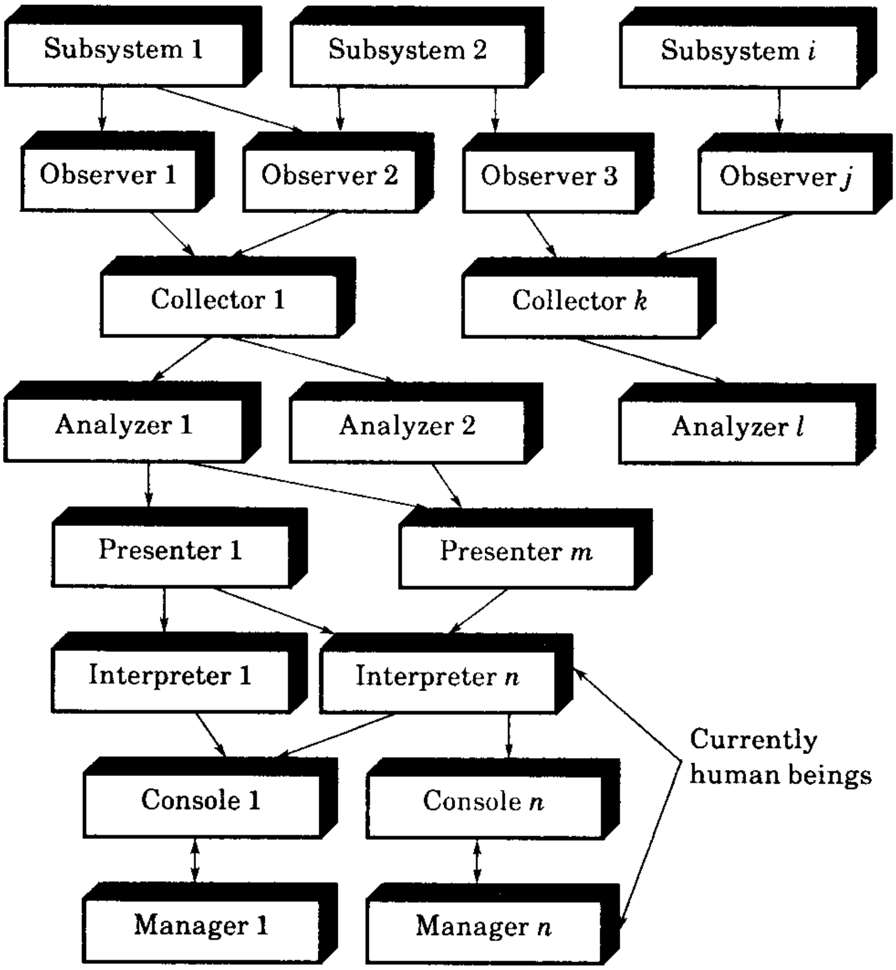


FIGURE 7.2 Components of a distributed-system monitor.