Systems performance/measurements—A quantitative base for management of computer systems

by PAUL MALICK

United Air Lines
San Francisco, California

INTRODUCTION

Systems Performance/Measurements is a new area of computer technology which provides a quantitative base for management of data processing systems. This function has generated high interest due to increasing emphasis on cost effective management of these data processing systems. The people, the tools used and results obtained are the subject of this paper.

SYSTEMS PERFORMANCE/MEASUREMENTS AT UAL

Airline industry problems

The airline industry has a unique set of problems to solve. They are:

• High Capital Investment
• Large Maintenance Requirement
• Highly Service Oriented
• Highly Regulated
• Highly Unionized
• Geographically Spread Out
• Highly Competitive Route Structure
• Stringent Security Requirements

Because of these problems, United Air Lines has been, and is, a pioneer in computing. The early systems were primarily batch; however, as the technology became available, the on-line systems were brought into being by working closely with vendors to develop the required systems. Newer, faster, more powerful and reliable systems are being developed as their need can be economically justified.

UAL computer organization

United has formed the Computer and Communication Services Division to effectively manage the large investment in resources, both machines and people.

Figure 1 shows the United Air Lines Computer Centers, equipment, major users and interaction.

United Air Lines has four computer centers, located in three cities, as well as a data communications network serving all cities throughout the airline's system and connected to other airlines. Non-pooled networks or links connect the four centers. These centers serve a diverse group of major users, including: Finance, Marketing, Maintenance and Operations. There are a large group of medium and smaller sized users including Corporate Planning, Food Services, Personnel, Industrial Engineering, etc.

Planning/measurements cycle

Effective management of a large capital investment in computer systems requires a quantitative base which must relate to the division planning. Figure 2 depicts the planning/measurements cycle which relates system performance to division planning.

Resource measurements indicate where we are now. The measurements are used as:

• Input to Planning
• The basis of chargeback to users for resources used.

The resource areas measured are: central site hardware, application software, systems software and network/remote hardware.

The plan indicates where we are going and when we will get there. Plans can be separated into three time periods as follows:

• Near term—days, weeks
• Short term—up to a year
• Long term—greater than a year

The near term problem is usually solved using measurements directly. Modeling and simulation using measurements as input are often required to provide the quantitative base of the plan for the short and long term.

A very important part of the planning process is to compare the planned vs. actual and feed the result back into the plan. The objectives of the feedback are to highlight mistakes so they can be corrected and serve as a basis for improving the planning mechanism.
The Systems Performance Measurement groups at United Air Lines provide most of the quantitative base for management of the computer systems. The groups answer the following basic questions:

1. How much of and how well are we utilizing the current resources, quantitatively?
2. What impact will hardware, software and applications changes and/or additions have on systems resource utilization, quantitatively?

The answering of these questions requires the use of the appropriate tools. In the case of question number one, the tools would primarily be software and hardware monitors. The tools for answering question two are a combination of software and hardware monitoring as well as simulation and modeling.

The results of the use of the tools along with analysis, are systems performance reports and recommendations. The report with the largest impact is the “Planning and Capacity Report” for each resource center. This report is used by division management as the prime quantitative input to planning.

The personnel in the Systems Performance area are all experienced computer people and tend toward the senior level. Each tends to have a systems viewpoint, be highly analytical, and work well with various people.

Planning and capacity reports

A planning and capacity report for a particular computer center contains the following:

1. Resource utilization, by resource, for each of the next five years. The on-line system resource utilization is broken out by user application.
2. Prediction of system bind points by resource and year.

All resource utilizations include changes and additions of hardware, software and applications at the time predicted.

RESOURCES MANAGED AND TOOLS USED

Several tools are used to provide the quantitative base for management of resources. The tools are measurement, simulation and modeling. The tools provide data for answering the two basic systems performance questions.

Resources managed

Figure 3 depicts the five computer resource areas to manage. These areas are managed as a system as opposed to separate entities. It should be noted, an on-line system contains all five areas while a batch system does not have a data communications network to manage. It is interesting to note that an on-line system is primarily self-scheduled and a batch system is primarily manually scheduled.

The central processing unit executes instructions, thus the most important parameters are:

- CPU activity (which is broken into):
  - operating system
  - application or problem program
  - Average Instruction Execution Time (AIET)

The memory module holds programs and data. The prime parameters are:

- Memory Utilization (by whom)
- Memory Contention (multi-processor systems only)
- High Speed Buffer Hits/Misses

The Input/Output channels transfer data to and from peripheral devices and memory. The prime parameters are:

- Channel Utilization (peripherals)
- Channel Utilization (computer-to-computer)

The peripheral devices input, output and/or store data. The major parameters are:

- Control Unit Utilization
- Device Utilization
- Average Access Time
- Number of Accesses
The data communications network provides a data path to and from devices which are remote from the computer. The network can be envisioned as a peripheral device which is more complicated than any other. The prime parameters are:

- Data Communications Line Loading
- Number of Messages
- Number of Characters Per Message

There are various system parameters which are important because they provide data which show quantitatively the relationships between the parts. These are:

- Number of file accesses
- Queue levels
- Job/Step Log
- Number of concurrently active channels
- Number of jobs or transactions concurrently active.
- Amount of time instruction execution is overlapped with channel activity.
- Job transaction elapsed time
- Remote device response time

System optimization is concerned with the “balance” of all the parameters identified.

**Measurements**

**Basic questions answered by measurements**

There are three basic questions answered by measurements:

a. What percent of the resources are being used now?
b. Who is using what percent of the resources?
c. What are the specific device characteristics for the system being measured (ALET, average access time, etc.)?

The answer to the first question indicates the capacity of the resource being used, thus the reserve capacity can be calculated. The answer to the second question indicates who is consuming the resources. The data obtained by answering the third question is used as input to the simulation/modeling process. The measurement tools used are hardware and software monitoring.

**Hardware monitoring**

**Unique characteristics**

A hardware monitor has a set of unique characteristics as a measurement tool:

a. It is transparent to (no-load on) the system being measured.
b. It can capture 100 percent of the samples.
c. It captures data not available by any other method (such as Average Access Time, Average Instruction Execution Time, etc.)

The first characteristic indicates that no distortion is placed upon the system measurements. The second characteristic denotes the degree of confidence that the measurements taken are representative of the period measured. The third characteristic means that the key data required for modeling is available and accurate.

**Choosing a hardware monitor**

United Air Lines owns a triplex Compress D-7900 hardware monitor system. The choice of this particular system was made based upon the following prime characteristics.

a. Number of Counters—determines how many functions or events can be accumulated separately (example: monitoring all storage activity).
b. Plugboard Logic Capacity—determines how much signal combination and translation can take place simultaneously (example: core utilization).
c. Maximum Counting Rate—the counters should count at a rate equal to the speed of the signal being measured so 100 percent of the signals can be captured. Note: this can sometimes be accomplished using plugboard logic. However, the logic is then unavailable for other purposes and plugboard logic is a critical resource.
d. Built-in Signal Simulator—is used to check out plugboard logic before the monitor is connected to the system. Without this capability, checkout of plugboard logic is difficult, time-consuming and often inaccurate. This feature is a must!
e. Modular, Reliable and shipped easily—the system must be capable of being broken into smaller elements for shipping. Modularity helps this situation greatly. Reliable units are a must to develop confidence in the results.

**Using a hardware monitor**

The use of a hardware monitor follows a very definite pattern which is shown in Figure 4, Hardware Monitor Use Flow.

The first step is to set up a test plan. The question to ask is: “What are the results I want to obtain?” Once the ob-
jectie is established, the necessary effort can be expended to plan toward and reach that objective.

Researching the probe points has been necessary in many cases because the probe points are not available in any probe point library. This usually doesn't reflect probe point library inadequacy, only that the problem probably hasn't been solved before. Most of the standard measurement probe points are in the library, however. It is important to document the test plan, probe points and logic so they can be used again with a minimum of effort.

Validation of the resulting data with whatever other data available is necessary to develop confidence in the results. A direct comparison with software measurements is ideal. However, when these are not available, trends and relationships with other measurements or simulation can be used.

Writing a final report solves several problems. First, an historical record is available to indicate the results whenever they are needed. Second, the report provides the starting point for the next time the tests are run. Finally, organizing the results into report form often uncovers problems with the tests which can either be corrected immediately or the next time the test is run.

**Measurements obtained using the hardware monitor**

Hardware Monitor measurements obtained include the following:

- CPU Utilization
- Channel Utilization
- Core Utilization
- Memory Contention (on a multi-processor system)
- High Speed Buffer Hits
- Computer to Computer Channel Utilization
- Average Instruction Execution Time (AIEF)
- Peripheral & Peripheral Control Unit Utilization (Disk, Drum, Printer, Tape, etc)
- Drum & Disk Average Access Time
- Data Communications Network Characters transferred.

These measurements were obtained on various United Air Lines computer systems including: IBM 360/50, 65, 370/155 and UNIVAC 1108; plus an IBM 2969 (Programmed Terminal Interface). It should be noted that most of these measurements would be of equal interest for either an on-line or batch system. Only the values and their relationships change.

Hardware monitor measurements represent a composite of all individual times or events. Thus, another measurement method is normally required to reconstruct the components which make up the total. This method is software monitoring.

**Software monitoring**

**Unique characteristics**

A software monitor has a set of unique characteristics as a measurement tool:

a. The counter mechanism is connected directly into the system software thus the counter and collection mechanism bias the results.

b. Data collection is done on a sampling basis because of the bias mentioned above.

c. It captures only data available internally in the computer.

The last characteristic means that this measurement tool can be used to capture transaction, program, job and file data. Software monitors are normally continuously resident in the computer. Thus, they can be turned on at will for diagnostic or other purposes without delay. It should be noted the sample rate must be carefully chosen to balance the mechanism bias on the results with the validity of the sample.

**Software monitors used**

The following software monitors are used at United Air Lines.

a. Systems Management Facility (SMF)—Is a standard IBM OS system measurement package. In addition to standard data reduction, Mark IV file management package (from Informatics) is used to easily reduce the data for special purposes.

b. MFKEYS—Is a data collection and reduction package developed by UAL used to capture CPU and drum subsystem utilization as well as key system parameters from the operational computer system.

c. Apollo Data Collection and Reduction—Is a UAL developed package used to measure the Apollo reservation system utilization. The measurements include CPU, File, Message Mix, Program and transaction activities.

d. Compool Snap—Is a UAL developed data collection and reduction package for the operational computer system which is used to capture message and program hand-off data.
e. PTI Data Collection—Is a vendor developed data collection and reduction program used to collect core utilization, data communication line and message data in the IBM 2969 Programmed Terminal Interface system for the Apollo reservations system.

Each software data collection package is used for a specific area as identified. It is necessary to have tools to quantify the activity in each resource area so a full picture can be developed.

**Measurements obtained using software monitors**

The following measurements have been obtained using software measurements:

- CPU Utilization (total, by job or transaction)
- Channel Utilization
- Core Utilization
- Peripheral Unit Utilization
- File accesses by File (Data Set) by Job and Step
- Job and Job Step Log, Start, End
- Number of Operating System Requests
- Queue Levels
- Number of Transactions by Transaction Type
- Program/Job/Step Execution Time
- Number of Messages
- Length of Message

These measurements were obtained on various United Air Lines computer systems as previously noted under hardware monitoring.

**Simulation/modeling**

**Basic questions answered by simulation & modeling**

Simulation and Modeling are the tools used to predict the future in terms of resources. The basic questions answered are:

1. What impact will various changes and/or additions have on the system quantitatively?
2. When and at what level will the system saturate?
3. What are the system bind points?

The answers to these questions become the basis for a Planning and Capacity Report. It should be noted that correct predictions are made on the basis of an accurate picture of the present. This accurate picture of the present is based upon actual measurements.

**Modeling description**

There are two parts to the modeling process. They are the model mechanism and the model input. The model mechanism represents the system being modeled. The total computer system model consists of the system software (operating system), the system hardware and their interaction. The dynamic interaction of these parts are simulated by the use of various algorithms and probability distributions within a logical flow of events.

The model inputs provide the data base to be operated upon and the stimulus for the model mechanism. For example, the total system model inputs are the descriptions of applications, programs, files and messages along with the time of day and frequency the jobs or transactions are run. Validation of the model mechanism and inputs are necessary both individually and as a whole. The objective is to increase the confidence in the results so they can become accepted and relied upon. The results are only as good as that which goes into the modeling process. Validation, where possible, should be with actual data.

Models have variable inaccuracies because:

- The model is not a complete representation of the system (the system itself is).
- The mechanism may have inaccuracies.
- The inputs may be incomplete and/or incorrect.

Thus, a model output which is correct to ±5 percent is often considered very accurate.

Modeling is a time and energy consuming task. Thus, it is not always justified economically.

**Models used**

Where possible, existing models are used either directly or with modifications. Normally, however, an existing model has not been available. Therefore, United Air Lines has made a considerable investment in building models.

The following are the primary models used at United Air Lines:

a. Systems Analysis Machine (SAM)—Is a model developed by Applied Data Research, primarily for IBM batch systems; however, any system may be modeled. The vendor has made a large addition to the model for United Air Lines, which allows modeling of a job stream without detailed descriptions of each job and job step. This feature is a must to obtain data for preparation of a Planning and Capacity Report. The outputs from the model are CPU, channel, device and program activity statistics. Much of the detailed testing and validation of the model addition was performed by United Air Lines.

b. Operations System Planning Model—is a model, developed by United Air Lines, of the Operational Computer System which is used to obtain the basic Planning and Capacity Report data including CPU, channel and message activity. This model is written in GPSS.

c. United Systems Simulation Model (USSM)—Is a model developed by United Air Lines in GPSS, which is used for obtaining the basic Apollo Reservations System Planning and Capacity Report data. Thus, it predicts CPU, channel and message activity.
The following special purpose models have been developed:

- Apollo Network Model
- Operational Computer Network Model
- Average Instruction Execution Time (AIET) Model
- IBM 3330 Disk Model
- Programmed Terminal Interface System Model

Each model serves or has served a specific purpose. The network models are used for data communications network management. The AIET model was used to verify the operational computer system average instruction execution time. AIET is a key input to any system model and the Hardware Monitor was not available to capture the actual value at the time. The IBM 3330 Disk Model was developed to simulate the operation of 3330's on the Apollo reservation computer system.

Special purpose models are developed as the need arises to quantify the dynamic impact of system changes before they are approved and implemented.

The magnitude of the expenditures, the possible impact upon service to our users and the risk of implementation are all factors in any decision to build a particular model.

**USE OF TOOLS**

The most significant and important use of the Systems Performance tools has been to provide a quantitative base for decision-making. Decisions have been made, using the quantitative base, on major changes to each of the four computing centers.

**Operational computer system**

The decision was made to postpone the acquisition of one million dollars of mass storage drums for a minimum of two years based upon actual measurements of average access time. The actual average access time was found, with the hardware monitor, to be 30 percent less than the manufacturer's specification. The difference was due to the amount of storage used and the data placement on the drums for the UAL application mix.

**Marketing (reservations) computer system**

The decision was made to upgrade to the largest capacity computer available based upon predicted resource utilization. The predictions showed the life of the new computer to be approximately five years. Thus, a costly interim conversion was avoided.

A decision was made, also, to sign contracts for computer services with several outside firms due to the availability of predicted excess capacity for the length of the contracts. Thus, revenue is being produced by using the excess capacity.

**Financial computer system**

The decision was made to not upgrade from a pair of 370/155 to a pair of 370/168 computers to meet Summer of 1973 peak resource requirements. Measurements indicated the peak CPU load build up was very high. Using these measurements as a base, it was predicted that the peak load would be at or near system saturation, causing great disservice to our users. System optimization and offloading of jobs to the back-up computer for the marketing system reduced the peak CPU activity by ten percentage points below the predicted peak. Thus, considerable investment in hardware and manpower were postponed for at least a year.

**Maintenance computer system**

The decision was made to convert this computer center to primarily a remote job entry (RJE) system, based upon predicted resource requirements and the availability of adequate capacity in the back-up computer for the marketing system. Expansion of this system will not be necessary for at least a year.

It is felt, at United Air Lines, the Systems Performance groups continuously pay for themselves by providing a quantitative base to:

- Make correct decisions
- Avoid costly mistakes

Direct cost savings, such as cost avoidance or postponement, can be calculated. Indirect cost savings, such as lost time due to mistakes, are difficult to measure.

It must be remembered that people must ultimately make the final decisions, but a quantitative base can, if used correctly, greatly assist the process.

**ACKNOWLEDGMENTS**

I wish to acknowledge the contributions of my fellow Systems Performance people at United Air Lines, whose efforts produced many of the tools and techniques shown in this paper. I also wish to thank the persons who reviewed this paper and made various helpful suggestions and, finally, I wish to thank United Air Lines for the opportunity to prepare and present this paper.

**REFERENCES**