# MONTE - A RISK ANALYSIS PROGRAM

Richard Fleming, AT&T Bell Laboratories - The City College of N.Y.

## 1. ABSTRACT

This paper describes procedures which can be used to establish a relationship between the service that a computer center provides, and the expenses associated with providing a particular level of service. These procedures are embedded in a Monte Carlo simulation model (MONTE). The MONTE program can be used by investigators to investigate the financial effect of service level changes.

## 2. INTRODUCTION

As the pace of technology increases, users become increasingly dependent upon their information systems. They see larger portions of their budgets allocated to computer expenses and they ask if the services they are receiving are worth what they are being charged. Computer center managers, on the other hand, are often uncertain whether it is better to lower costs by cutting services, or increase performance levels by spending more.

Unless the relationship between service levels and costs/expenses is known it is impossible to allocate resources efficiently. Managers need to baseline the quality of the service that they currently provide and understand the relationship between that baseline and their current expenses. Users need to determine the relationship between the service that they are currently receiving, and the costs that they incur. Both groups need to determine how much money is lost when computer service degrades.

This paper describes procedures which can be used to answer these types of questions. These procedures establish a relationship between the service that a computer center provides, and the expenses associated with providing a particular level of service. They are embedded in a Monte Carlo simulation model (MONTE). The MONTE modeling program can be used by investigators to investigate the financial effect of service level changes.

## 3. OVERVIEW

The Monte Carlo method can be described as a method of studying an artificial stochastic model of a physical or mathematical process, particularly one based upon probabilistic assumptions. The physical system is replaced by an abstract mathematical model. This model is comprised of independent variables and a series of equations that calculate the dependent variables. The output from the model is information.

The MONTE model takes independent variables (Performance Data, Salary Costs, Equipment Expenses, etc.) and calculates a number of dependent variables (User Cost, Computer Center Expenses, etc.). The analyst can use this information to track service levels over time, or evaluate the cost/benefit effect of changes in service levels.

MONTE can be characterized as an **intuitive, distribution based, simulation model** that uses **device independent, cost/based** variables. The model was designed to be easy to understand, use and modify. The following sections discuss these characteristics in more detail and explain why they were chosen.

### 3.1 INTUITIVE MODEL

Processes can usually be modeled in many different ways. MONTE was developed as an easy to understand cost/expense based "intuitive" model. The variables are based on performance measurements, user costs, and computer center expenses. These values are readily available, well defined and commonly used by system professionals to track service levels.

The calculations that MONTE uses are simple arithmetic equations which can be easily verified or customized to meet specific requirements. A conscious attempt was made to free MONTE from the mathematical jargon that is too often used in analytical models.

### 3.2 DISTRIBUTION BASED

MONTE was constructed so that data can be sampled from frequency distributions. It can therefore be considered "distribution based." Important information is gained when the program can make use of the range of data values, as opposed to only some "average" number. Most conventional models will simply plug the average batch turn-around time into the equations. This represents only one point on what is, in reality, a continuous curve of possible combinations of values, each with a probability associated with it. Single values become even more misleading if the distribution is not normal in a mathematical sense. MONTE is not affected to the same degree by these problems because it samples from either the actual data, or whichever distribution matches the real data most closely.

### 3.3 SIMULATION BASED

There are two basic ways of modeling physical processes; analytical and simulation. The analytical approach uses mathematical techniques to construct and solve equations that describe the process in question. Using the equations of queuing theory, for example, an analyst can determine exactly the average numbers of batch jobs in a queue if the number of servers and the average service time is known.

Solving sets of equations that model real-world processes is a difficult undertaking when the number of variables is large and the data values are not "well behaved" in a mathematical sense. It is often easier to obtain solutions to these types of problems by simulating the behavior in question. A simulation approach involves setting up a model whose equations establish a high degree of correlation between the input/output of the model, and the input/output of the physical system under investigation. A model in which the solutions to a problem are obtained in such a manner, particularly if the simulation involves random number sampling , is commonly known as "Monte Carlo Simulation."

### 3.4 DEVICE INDEPENDENT

The variables that MONTE uses (interactive response time, batch-turn around, etc.) are device independent. Service levels can be measured on anything from a PC to a supercomputer. MONTE can be used to evaluate the effect that moving users between systems will have on overall service levels. Because of this characteristic it is possible to combine the performance of many different machines into one overall computer center service level indicator.

Such an approach also has the advantage of being stable over time. It can be used without recalibrating the basic indicator if machines or computer environments change. The two primary output variables from MONTE (User Costs and Computer Center Expenses) are calibrated in dollars.

### 3.5 COST/EXPENSE BASED

The output from MONTE is formatted on a "cost to the user" basis. The application manager can see exactly how much money can be lost if service levels degrade. The computer center manager, on the other hand, can use the model to determine how much money will need to be spent in order to achieve a specific service level objective. The effect of these expenditures will have either a positive or negative effect on the user's bottom line.

## 4. VARIABLES

A simulation model consists of two critical items, the equations that create the dependent variables to model the process, and the independent variables on which the equations are based.

When building a model the independent variables should be carefully chosen. They should contribute to the solution of the problem, and they should be easy to obtain. If theoretical distributions are used, their characteristics should closely match the characteristics of "real life" data points. As mentioned earlier, the service level variables which MONTE uses were selected to be device independent.

### 4.1 INDEPENDENT VARIABLES

The most important independent variables that are supplied to the MONTE program are shown in the following table. These factors were chosen because each can effect productivity, and most can be easily measured.

TABLE 1

| VARIABLES | CLASS |
|---|---|
| Number of Users on system | Load |
| User Command Count | Load |
| Number of Tape Mounts | Load |
| Number of Batch Jobs | Load |
| Number of Print Jobs | Load |
| Number of Help Requests | Load |
| Interactive Response Time | Performance |
| Batch Throughput Time | Performance |
| Network Availability | Performance |
| Print Turn-Around Time | Performance |
| Tape Mount Time | Performance |
| Help Desk Response Time | Performance |
| System Availability | Performance |
| System Expenses | Financial |
| Salary Information | Financial |

The variables in the MONTE program can be broadly divided into 3 classes:

1. **Performance**
2. **Load**
3. **Financial**

Depending on the objective of the run(s), the value of these variables will either remain constant across runs, or be varied so as to provide the analyst with a range of solutions. (See Section 5)

*4.1.1 USER SUPPLIED INDEPENDENT VARIABLES* It is often difficult for users and computer center managers to agree on the factors that constitute acceptable performance. Customer Service Representatives who are charged with establishing Service Levels are often faced with the problem of translating hardware measurements into values that can be understood from the customers perspective. Users are often understandably reluctant to undertake extensive surveys to obtain more precise information. A model based on unrealistic expectations or inaccurate information is of limited value. It may cause the manager to spend money improving performance in areas which really do not contribute significantly to an overall improvement in service. In other cases lack of information may mask serious service level deficiencies.

In an attempt to minimize these problems MONTE restricts the information required from users to the following elements:

TABLE 2

| USER SUPPLIED INDEPENDENT VARIABLES |
|---|
| 1. Salary Data |
| 2. User Groupings |
| 3. Group Efficiency Factors |

Salary Data should be available from the users' management.

The last two items, **User Groupings** and **Efficiency Factors** are discussed in more detail in the following sections. They are weighting factors which can be used to "tune" the model to reflect actual user activities.

*4.1.1.1 USER GROUPINGS* The following section describes minor weighting factors that are used to more accurately portray user behavior. The **User Grouping** and **Efficiency Factors** will be specified by the users and entered into the model on a percentage basis by the data analyst. These factors will vary as the behavior of the user population changes.

Interactive computer users, by the nature of their work, depend on computer systems which are available, and provide short interactive response times. Users performing interactive tasks such as editing, CAD/CAM design work, and program development are usually mentally absorbed in the task at hand, and rarely perform more than one task at a time. If the computer system is not available, either through telecommunications or systems problems, the time that the system is down is largely non-productive for the interactive user. Long interactive response times that cause the interactive user to wait also decrease their productivity in a non-linear manner.

The case is somewhat different for users who submit batch jobs that run in the background or print work. In these cases the computer user may be able to do other work while they are waiting for jobs to complete or output to become available. For these reasons, the MONTE model groups the user population into three segments on a percentage basis.

These three groups are described as follows.

1. **GROUP** A- The percentage of users who submit work but do not care how long it takes to get such work back. Users submitting background work or print jobs will fall into this category. There is no cost associated with a performance

degradation for Group A users.

2. **GROUP B** - The percentage of users who submit work and cannot do any other work until their jobs return. CAD/CAM designers and users submitting compiles or link edits can often be considered Group B. Users who work in such a single-stream mode place heavy demands on a computer facility for excellent turnaround and short response times because they do not have another assignment to turn to while waiting for computed results. Most interactive users should fall into the Group B category.

   Costs due to degraded performance are fully weighted for Group B users.

3. **GROUP C** - The portion of the user community assigned to GROUP C would contain users who submit their work, care about getting their output in a reasonable time, but can do other jobs while they are waiting. In theory Group C users would seem to be the most efficient workers because the degree of efficiency goes up as the ability to interleave work increases. For a GROUP B user all of the time waiting for a task to complete must be assigned as a cost. Since a GROUP C user can work productively on another task during the time that would normally be wasted by the GROUP B user, the cost associated with waiting will be reduced. In reality however, there is some loss from the ideal when users attempt to perform concurrent tasks. The MONTE model includes a user specified **Efficiency Factor** which modifies the user cost and takes this effect into consideration.

   MONTE uses the percentage values assigned to each of the different groups when User Costs are assigned for Batch and Print jobs. All interactive work is charged at the full rate.

*4.1.2 SERVICE LEVEL INDEPENDENT VARIABLES* A great deal of research has been done studying the effect of using various service level metrics. MONTE follows the traditional approach of using computer performance information in the model. The independent service level variables that MONTE uses to calculate System Expenses and User Costs are shown below.

TABLE 3

| SERVICE LEVEL INDEPENDENT VARIABLES |
|---|
| 1. System Availability |
| 2. Network Availability |
| 3. Interactive Response Time |
| 4. Batch Turn-Around Time |
| 5. Print Time |
| 6. Tape Mount Time |
| 7. Help Desk Turn-Around Time |

#### 4.2 DEPENDENT VARIABLES

Using the values of the independent variables, and the equations built within the model portion of the program, MONTE derives a large number of dependent variables. The most important of these are shown in the following table. Remember that EXPENSES are associated with the supplier of services (the data center). COSTS are associated with the users of computer services.

TABLE 4

| DEPENDENT VARIABLES |
|---|
| 1. Total Cost |
| 2. User Cost |
| 3. System Expenses |
| 4. Availability Cost |
| 5. Print Cost |
| 6. Interactive Cost |
| 7. Batch Job Cost |
| 8. Tape Mount Cost |
| 9. Help Desk Cost |

*4.2.1 SYSTEM EXPENSES* System Expenses include all of the expenses associated with supporting the computer facilities at a given level of service. Initial values for the system expenses can be obtained from data center management for current levels of service.

Total System Expenses consist of the following line items.

TABLE 5

| SYSTEM EXPENSE VARIABLES |
|---|
| 1. Loaded Salaries |
| 2. Materials/Supplies |
| 3. Clerical Services |
| 4. Equipment Rental |
| 5. Software Licenses |
| 6. Rent |
| 7. Data Communications |
| 8. Other Data Communications |

*4.2.2 USER COSTS* Computer service means many things to many people. MONTE uses the independent variables shown in TABLE 3 to generate user costs for the following items:

TABLE 6

| DEPENDENT USER COST VARIABLES |
|---|
| 1. Interactive Command Cost |
| 2. Batch Job Cost |
| 3. Print Job Cost |
| 4. Tape Mount Cost |

MONTE calculates the cost that users incur when service levels decline because of poor system performance. Using the independent variables (Response time, Availability, Salary, etc.) MONTE calculates data points for each dependent variable. The equations assume that only on a perfect system will there be no user costs. Such a system would be capable of providing instant interactive response time, instant print and batch turn-around times, and 100% availability at all times. Obviously such a theoretically "perfect" system is not physically possible. Anything less than perfect system performance however, will be charged as a user cost.

*4.2.2.1 AVAILABILITY USER COSTS* Availability is a critical performance indicator. The MONTE model accepts data on the availability of both the computer system, and the network through which the user accesses the computer. The percentage of users on the network is used to pro-rate the cost of availability numbers.

Two separate distributions are used; one for the network, the other for the computer. The following table shows how availability costs are allocated for the network and the system.

TABLE 7
SYSTEM AND NETWORK AVAILABILITY

| SYSTEM | up | down | down | up |
|---|---|---|---|---|
| NETWORK | up | down | up | down |
| CHARGE | none | system | system | % users on network |

If the System is up for example, but the Network is down, MONTE calculates user cost by on a minute basis by multiplying the number of minutes down, by the number of users on the system at that time, by their salary. It then adjusts that figure by the percentage of users on the network. Users attached locally to the CPU can still perform useful work.

*4.2.2.2 INTERACTIVE USER COSTS* Interactive users issue commands and then wait for the system to respond. This is the time period that MONTE calculates as a USER COST. The time period from when the system responds, until the next command is entered, is usually considered to be user "think" time.

### 5. RESPONSIBILITIES OF THE DATA ANALYST

The data analyst is responsible for configuring and running the MONTE program. The validity of the model will largely be based on the decisions made by this person. Some of the decisions that the analyst must make will be described in this section.

### 5.1 ESTABLISHING OBJECTIVES

The way in which the analyst sets up the model is, to a large degree, determined by what the model is to be used for. The analyst must know the objective of the run(s). Although the basic program will not change, various CLASSES of data values should be held constant depending on this objective. In most cases it will be necessary to rerun the model as objectives change.

The following table suggests which classes of variables should remain constant as the type of analysis changes.

TABLE 8
CLASS ANALYSIS TABLE

| RUN-TYPE | FINANCIAL | PERFORMANCE | LOAD |
|---|---|---|---|
| Financial | Vary | Constant | Constant |
| Performance | Constant | Vary | Constant |
| Load | Constant | Constant | Vary |

If the purpose of the model is to measure and track service levels across time for example, it may be necessary to fix all variables except those that relate to performance and reliability (response time, availability, batch and print turn-around time). All other variables (# users, salaries, command counts, group user classes and efficiency factors) should remain constant so that service levels can be tracked over time.

### 5.2 ESTABLISHING A TIME FRAME

The analyst must determine the time frame upon which measurements should be based. The period should be large enough to ensure that the values for the independent variables are representative. If the time periods are too long however, it is likely that the characteristics of the system will have changed during the period being sampled.

### 5.3 CHOOSING THE DISTRIBUTION

It is the analysts' responsibility to screen the data before including values into the model. Scatter plots can be used to identify and eliminate outlying data points. Regression analysis can assist in choosing an appropriate theoretical distribution. Tightly constraining the standard deviations will obtain the practical result of selecting variables which are essentially constants.

*5.3.0.1 SAMPLING PROCEDURES* When building a simulation model, one must decide whether it is better to obtain the distribution values from the empirical data, or to represent a process by a theoretical distribution. This decision is somewhat dependent on how the model is to be used. There is an intrinsic appeal in sampling from the actual data points. If the analyst wishes to explore alternatives however, it is easier to change theoretical distribution parameters (mean, variance, etc.) than to modify the actual data points so that they produce the desired characteristics.

Before running the program the analyst needs to specify how many iterations of the program (program loops) will be sufficient. The more samples that are taken, the more certain the analyst will be that the distribution values will approach the parameters that are specified. As a practical matter however, convergence to acceptable levels occurs rapidly.

### 6. OPERATIONAL OVERVIEW

Once the analyst has decided on the characteristics of the model and entered that information the simulation can be run. In operation MONTE will read in constant variables and then loop through the distribution based variables selecting data points according the rules which the analyst has specified. Each iteration of the program will then substitute the selected value into the equations that "model" the process. Finally, the resulting data will be summarized and processed for output.

### 7. SAMPLE CALCULATION

Since MONTE was designed primarily to be a cost/expense based model, the output from the simulation is presented in these terms. The Service Level Independent Variables shown in TABLE 3 are very similar in name to the variables in TABLE 4. The difference is that the Dependent Variables are formulated on a cost/expense basis.

The following table shows a portion of the model that calculates the costs associated with degraded batch turn-around times. Each iteration of the program will generate a single data point for each of the variables based on the equations specified in the model.

TABLE 9
SAMPLE CALCULATION

| STEP | VARIABLE | CALCULATED OR ASSIGNED |
|---|---|---|
| 1 | Salary Mean | $60,000 |
| 2 | Salary Std.Dev | $10,000 |
| 3 | User Mean | 50 |
| 4 | User Std.Dev | 20 |
| 5 | NumberC | 0.40 |
| 6 | Effic_Fact 'C' | 0.80 |
| 7 | Users | User Mean + User Standard Deviation * NORMAL |
| 8 | Salary | Salary mean + Salary Standard deviation * NORMAL |
| 9 | Batch Job Time Mean | 2.7 min |
| 10 | Batch Job Time Std.Dev | 1 min |
| 11 | Number Batch Jobs | (Job Mean (3) + Job S.D. (0.1) * NORMAL) / Users |
| 12 | Batch Job Time | Batch Job Time Mean + Batch Job Time S.D. * NORMAL |
| 13 | Batch Minutes | Number of Batch Jobs * Batch Job Time |
| 14 | Batch Costs 'C' | Batch Minutes*Salary*Users*MinMonth*NumberC*Effic_FactC |
| 15 | Batch Costs | Batch Costs "C" + Batch Costs "B" |
| 16 | User Cost | Avail Cost + Response Time Costs + Print Costs + Batch Costs |
| 17 | Total-Cost | User Cost + Computer Center Expense |

Although not shown in this example all variables are recalculated to a user minute basis.

The equations that MONTE uses to calculate Costs and Expenses are simple, but typically involve multiple steps. The final calculation for determining the costs of batch job turn-around time for Group 'C' users for example, is shown in STEP 14. Working backward we can see that the first variable (Batch Minutes) is obtained in STEP 13 in which the Number of Batch Jobs is multiplied by the Batch Job Time.

The number of Batch Jobs (11) is specified to be a number selected from a Normal Distribution having a mean of 3 and a standard deviation of 0.1. With such a "tight" standard deviation, the numbers selected will cluster very tightly around the value 3. Three jobs per minute will yield approximately 1400 jobs during an eight hour day.

The number selected from the distribution in STEP 11 represents all batch jobs run during the day. This variable must be divided by the number of users on the system (STEP 7) so that costs can be properly assigned.

The final calculations for Batch Costs 'C' in STEP 14 multiply the Batch Turn-Around Time in Minutes by the Salary per user minute, by the Number of Users on the system at that particular time. This number is then multiplied by the number of minutes in the month so that a monthly cost can be obtained. The last calculations weight the result by the percent of users in User Group C. That number is finally multiplied by the efficiency factor that the user has specified.

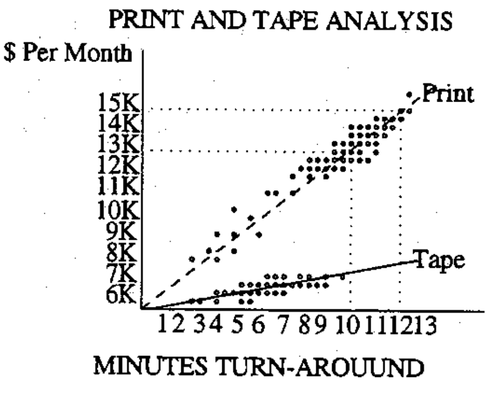
## 8. OUTPUT

The output from MONTE consists of a series of tables and graphs. A Summary TABLE provides information on both the dependent and independent variables. For each variable summary statistics such as mean, standard deviation, maximum and minimum values are displayed. This information should be used for initial model verification as well as for final analysis. The table also provides information on various percentages and ratios which are useful in establishing the relationship between variables.

As part of the analysis MONTE will also produce a series of plots. Some of these are used during the initial verification of the model to assist in determining what type of sampling distribution should be used. Others show the relationship between a particular service level indicator (such as response time, or network availability), and the cost of providing that level of service.

## 9. EXAMPLE

The following example compares the effect of spending money to add printers and decrease print turn-around time, or add additional tape drives to decrease tape mount times.




As part of the output MONTE will produce a graph showing the relationship between user costs and both tape and print turn-around time. It is apparent from the slope of the lines that slow print turn-around time is more "expensive" for the user than slow tape mount time. It can also be seen from the plot that decreasing the turn around time by approximately two minutes for each print job will decrease the user cost by approximately two thousand dollars per month. The graph shows the spread of data points so that the mean, minimum and maximum values can be evaluated.

Using this information the computer center manager could evaluate the benefits associated with service level changes. If the manager knew for example, that it would cost approximately $100K per year to purchase departmental printers, but that this would improve service by reducing the print turn around time by $150K per year, this would seem to be a good argument for purchasing the equipment.

## 10. DISCUSSION

The MONTE model can be used to establish a cost based service level metric if all the variables except those in the **performance** Class are kept stable. (See TABLE 1) If this is done the user cost will change as the service levels variables shown in TABLE 3 (availability, response time, etc.) fluctuate. Decreasing user costs reflect better service; increasing costs are due to a degradation in service. More detailed information showing why service levels are increasing or decreasing can be obtained from the summary table or the graphs.

The data analyst can configure MONTE to model the effects of changes in the user population by modifying the **LOAD** variables. If the FINANCIAL and PERFORMANCE class variables are kept constant the model will show what effect changing the number of users or their work habits will have on service. This information could be extremely useful in times of consolidation or movement of user groups between machines.

Because the MONTE model is distribution based the user will be able to examine the maximum and minimum points as well as the most "likely" numbers. Computer center managers may want to take a conservative approach in their use of the model and plan for worst case scenarios. MONTE is well suited for running these type of risk analysis investigations.

Since computer center performance is so tightly coupled with hardware performance, increasing this performance usually (but not always) results in an increase in computer center expenses. Increased performance will result in greater productivity and less costs to the user community. Decreased performance on the other hand, usually results in increasing user costs and a reduction in computer center expenses.

Viewed from a strict cost/expense perspective User Costs will be minimized when perfect service is provided. From a computer perspective reducing service will probably result in lower system expenses. Eliminating the systems entirely will result in the greatest cost saving to the Comp Center. Obviously neither of these approaches is satisfactory. From a corporate mangement standpoint perhaps the philosophy should be to minimize the sum of User Costs plus System Costs. This would result in the greatest cost/expense savings. Another approach would be to minimize costs at a service level which would be acceptable to the user community.

## 11. CONCLUSION

This paper described the MONTE model and showed how it could be used. The information provided by the model can be used by management when business decisions that affect service levels are being made.

## 12. REFERENCES


- Boies,S.J.,(1974),"User Behavior on an Interactive Computer System," *IBM Systems Journal* 13, No.1,2-19
- Doherty,W.J.,Thadhani,A.J."The Economic Value of Rapid Response Time", IBM Publication
- Ghanem,S.B.(1975),"Computing Center Optimization by a Pricing-Priority Policy," *IBM Systems Journal*, No.3,
- Hertz,D.B.,(1979),"Risk Analysis in Capital Investment," *Harvard Business Review*, Sept-Oct
- Lewis,L.J.,(1976),"Service Levels: A Concept for the User and the Computer Center", *IBM Systems Journal* 15,No.4
- Streeter,D.N,(1975),"Productivity of Computer Dependent Workers," *IBM Systems Journal* 13, No.3,292-305
- SAS User's Guide: Statistics,Version 5 Edition,(1985),SAS Institute Inc.
- Merges,M.J.,Mutlu,H.B.,(1989)" An Availability Model for Distributed Computing Environments", *IEEE*, June
- Lee,K.Y.,(1989),"Cost Analysis on Lost CPU Time Attributable to Power Outages", AT&T Bell Labs Technical Memorandum, 45261-890330.01TM, March 30
- Monte Carlo Method,(1951),National Bureau of Standards Applied Mathematics Series - 12, U.S. Government Printing Office, Washington D.C., June 11


## 13. ACKNOWLEDGEMENTS


The author would like to thank the management of AT&T Bell Laboratories for the freedom to pursue this work. I would also like to thank my fellow employees at Bell Labs; Duncan Clarke, Henry Tate, Betul Balasaygun and particularly Al Terry for their comments and suggestions.