663629

DP-1149

AEC RESEARCH AND DEVELOPMENT REPORT

SIMULATION MODEL OF AN IBM 360/65 JOB SHOP - I

F. D. KNIGHT

M. M. LUDEMAN

DU PONT

REG. U.S. PAT. OFF.

Savannah River Laboratory

Aiken, South Carolina

Printed in the United States of America
Available from
Clearinghouse for Federal Scientific and Technical Information
National Bureau of Standards, U. S. Department of Commerce
Springfield, Virginia 22151
Price: Printed Copy $3.00; Microfiche $0.65

663629

DP-1149

Mathematics and Computers
(TID-4500)

SIMULATION MODEL OF AN IBM 360/65 JOB SHOP - I

by

F. Delano Knight
and
Michael M. Ludeman

Approved by

J. E. Suich, Research Manager
Applied Mathematics Division

August 1968

E. I. DU PONT DE NEMOURS & COMPANY
SAVANNAH RIVER LABORATORY
AIKEN, S. C. 29801

CONTRACT AT(07-2)-1 WITH THE
UNITED STATES ATOMIC ENERGY COMMISSION

ABSTRACT

A discussion of a simulation model of an IBM 360/65 computer job shop (written in GPSS/360) is presented, with emphasis on the following points: a description of the system modeled, the GPSS model, the questions considered by the model, the design of the experiments made with the model, and the results obtained.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

If a computer job shop is to be operated efficiently, some quantitative methods are needed to evaluate alternative arrangements and types of equipment, and different sets of operating procedures. Prior to the development of the methods discussed in this report, such evaluations at the Savannah River Laboratory were done with a limited amount of quantitative information. Better methods were desired to substantiate the projected benefits of proposed changes in the operation of the computer system. This report describes one technique that is applicable to this problem.

The approach described in this report is an application of the methods generally known as simulation. The behavior of a complex system in real time is described by logical inter-relationships, which are in turn solved on a high-speed computer. The principal advantage of these methods is that the simulation model can be solved on a computer much more rapidly and economically than the corresponding experiments can be conducted on the real system.

Several general purpose computer languages have been developed for use in simulation problems. Some of these and examples of their application to computer job shops are described in the references. The language chosen for the present work is General Purpose Simulation System (GPSS). This choice was made because of the availability of a version of GPSS operable on the IBM 360/65 and because of the ease with which it is learned and implemented.

Although the job shop at the Savannah River Laboratory has been changed from the form described in this report, the model and the analytical methods serve as references to which subsequently developed models can be compared. The major alterations to the job shop were the installation of the HASP operating system; installation of additional direct access storage; and improvement in the Computer Center peripheral equipment to reduce reliance on the IBM 360/30 for printing.

SUMMARY

An IBM 360/65 computer job-shop model has been written in GPSS/360 with subsequent simulation and comparison of alternative systems. A series of 24 simulation runs were made to investigate the effects of changes in the areas of (1) input rate to the job shop, (2) job-tape parameter definition, i.e., number of jobs/input tape and cumulative run time/input tape, (3) job category definition, (4) environmental factors, and (5) equipment configurations. When the input rate was varied from 360 jobs/day to 180 jobs/day, most measures of performance improved with decreasing input.

The effect of increasing or decreasing use of the job shop by business application users was investigated in conjunction with changing the time for preventive maintenance from 12 noon to 2 p.m. each Tuesday and Thursday to 6 a.m. to 8 a.m. on the same days. When business application use is present, the simulation runs showed that no significant advantage was gained by shifting the preventive maintenance schedule. However, with no business application use, improvement in job turnaround and throughput was significant.

Several different equipment configurations were studied by varying the Central Processing Unit (CPU) from an IBM Model 50 to a Model 75; the system input/output (I/O) device from an IBM Model 2311 disk to a Model 2301 drum; and the printer from an IBM Model 1403-2, 600 lines/minute printer to a Model 1403-N1, 1200 lines/minute printer.

Modeling the IBM 360/65 computer job shop has provided a better understanding of its operations and also of the modeling of real systems for simulation purposes. The model is an illustration of a decision tool usable by management for evaluating various aspects of a computer job shop.

DISCUSSION

DEFINITION OF OBJECTIVES

The major purpose of the comparison of two or more computer job shops is to determine their relative efficiency. However, the efficiency must be defined in terms of some criterion that often depends upon the type of facility that the shop supports. In the present situation, the computer job shop is assumed to serve the scientific users in the Savannah River Laboratory (SRL). Routine accounting and bookkeeping activities are not considered in detail, but are treated as blocks of time during which the computer is not available to the scientific users.

In most computer job-shop operations, the most important function is to process the computer programs as rapidly as possible and to return the output to the user. In this case, "job turnaround" time is the measure of performance. It is probably the most used factor in evaluating the performance of a computer job shop.

One method of minimizing job turnaround is that of job scheduling. The most obvious scheduling policy is first-in, first-out (FIFO). FIFO forces the jobs to be processed by the computer in the same order in which they were submitted to the Computer Center (CC). While this guarantees that no job will receive a shorter turnaround time at the expense of other jobs in the queue, the mean job turnaround time is longer than for a queuing discipline based on priorities.[1] An alternative choice to the FIFO queuing discipline is a priority system based on some combination of job characteristics and the importance of each individual job. This latter idea forms the basis for the evaluations of the systems considered here.

Another factor to be considered in evaluating the performance of a computer job shop is the number of jobs that are completed and returned to the users during the working day shift. The desire is often to maximize the "job throughput" if the job turnaround time is not significantly increased.

Under a priority system, a majority of the users gain in reduced turnaround time while only a small number of users have their turnaround time increased. Under certain circumstances, this penalty is not as harsh as it may seem. An example would be a computer job shop where programs may be submitted only during the day shift, i.e., from 8:00 a.m. to 4:00 p.m. If a user cannot receive his completed program and output before 4:00 p.m., then it is useless to him until the next morning. Programs not completed before 4:00 p.m. will almost always be completed by 8:00 a.m. the following morning, and it is immaterial to the user whether they are completed at 5:00 p.m. or at 7:45 a.m. the next morning. Hence, the objective is to complete the maximum number of jobs during the day shift.

The problems and goals described in the above section form the foundation for the investigations described in the following sections.

SIMULATION

Simulation is described as "a numerical technique for conducting experiments on a digital computer, which involves certain types of mathematical and logical systems that describe the behavior of a ... system over an extended period of real time."[2]

Although simulation is but one of many approaches in modeling a system, its usage has become quite widespread in the past decade as the systems to be modeled become more complex and the ability of the digital computer to handle simulation-oriented problems has been increased.

With this increased interest in simulation, a new family of high-level languages was created to aid the user of this technique. A simulation language is the result of isolating and defining sequences of operations that are common to simulation problems and of defining a single new instruction to perform this sequence. Some simulation languages available and in current use today include:

- DYNAMO by MIT[3]
- General Purpose Simulation System (GPSS/360) by IBM[4]
- SIMSCRIPT by RAND Corp.[5]
- SIMPAC by Systems Development Corp.[6]

GPSS/360 was chosen as the language to use in this study. Foremost in this decision was its present availability for the IBM System/360 and the ease with which it may be learned and implemented.

REVIEW OF THE LITERATURE

The use of simulation models for the study of computer centers has become increasingly important since 1960. A number of studies point out the wide range of problems that can be studied within this small area.

One such simulation was reported by G. K. Hutchinson.[7] This model was formulated to describe the computation center at Lockheed-Sunnydale, as it existed in the spring of 1962. The model was divided into three areas: input generation, simulation, and output analysis. The simulation phase of the model was written in SIMSCRIPT while the input/output (I/O) phases were written in Commercial Translator Language.

The model was evaluated by means of a Final Measure (FM), which was a weighted function of (1) the volume-cost ratio of jobs per dollar, (2) hours of actual run time per dollar, and (3) the reciprocal of the weighted turnaround time. This FM had the advantage of being independent of the run time. The three quantities were normalized by assigning values of unity to average values from preliminary runs. Subsequent runs were compared to this normalized value of 1.00. The FM was defined such that an increase in the FM implied an increase in overall performance.

In the experiment, the effects of six distinct factors were studied at two levels through an analysis of variance technique. These were: (1) presence or absence of a priority system, (2) queuing discipline rule used to determine the next job to run, (3) manpower available, (4) equipment available, (5) volume of incoming jobs, and (6) stacking of jobs by the software system.

The best results, i.e., the highest value of FM, were obtained by using the FIFO within priority class. The model simulated up to 3 days of actual operation time.

J. E. Hawthorne[1] at Union Carbide discussed the same problem on a more limited scale. He described a proposed computer center with two IBM 7090's, four IBM 1401's for I/O, and various other peripheral equipment. Using a FIFO queuing discipline as a base line with which to compare subsequent runs, he selected the following alternative job shops to compare with the original model.

- A job priority monitor system.
- Same with 10% increase in workload.
- Addition of an IBM 360/40 for input and output.
- Implementation of a time-sharing type of operation with 25% of jobs coming from an outside source which gave the programs originating from outside users a lower priority than those from inside users.

Hawthorne used turnaround time and backlog at the end of 17 shifts as his criteria for rating one strategy over another. The FIFO gave the greatest turnaround time of the five, but the backlog at the end of the 17th shift was zero, whereas a backlog existed for each of the other four disciplines. The job monitor priority discipline both with and without the increased loading resulted in the greatest cumulative percent of jobs completed in the least time. However, the backlog at 12:00 midnight, Saturday, was 12-1/2 hours for the regular job load and 22 hours for the increased load. The remaining two experimental configurations decreased this backlog somewhat but also slightly increased their turnaround time. This model simulated 1 week of actual operation.

J. H. Katz at IBM[8] reported on an IBM "feasibility" project that involved a macroscopic SIMSCRIPT model of an IBM System/360 in which all the hardware involved could be changed by merely redefining a parameter for each equipment entity. However, actual configurations were not tested by IBM.

The three systems[7,1,8] were based on a macroscopic model, similar to the model used in this report. Other equally important work has been done using microscopic models.[9]

The three models previously discussed were run for simulated periods of 1 week or less, whereas the SRL model uses a basic time period of 4 weeks. Also, none of the other models considered delays such as preventive maintenance, unscheduled downtime, etc., to the extent as are incorporated into the SRL model.

Since each run of this model simulates 4 weeks of operation, the effects of such interruptions will have a chance to settle from a transient to a steady state.

DESCRIPTION OF GPSS/360

The General Purpose Simulation System (GPSS) is a high-level interpretive language designed expressly for work in simulation studies. It attempts to isolate operations that are common to simulation problems and to replace these operations by a single instruction, so that the correspondence between the model and the actual system is more easily recognized by the programmer.

GPSS is built around a set of simple entities, which are divided into four classes.

	Class	Entities
I.	Basic	TRANSACTIONS
II.	Equipment	FACILITIES STORAGES LOGIC SWITCHES AND GATES

Class		Entities
III.	Statistical	DISTRIBUTION TABLES QUEUES
IV.	Operational and Computational	FUNCTIONS VARIABLES SAVEVALUES

Many real life systems may be represented in terms of a flow chart where the traffic or moving entities proceed through the system as time passes. To capitalize on this feature, GPSS utilizes block diagrams to represent the system to be simulated, with each block capable of representing a single simulated operation. These blocks may represent an action such as: joining (or departing) a QUEUE, a decision to choose one of several alternative paths through the system, or the time required for service by a FACILITY.

The basic entity in GPSS is the TRANSACTION. The TRANSACTIONS are generated, or created, at a predetermined rate and are moved through the system on a first-in, first-out basis within priority classes. Associated with each TRANSACTION is a set of attributes called PARAMETERS, which are used to define characteristics of the individual TRANSACTIONS.

The FACILITY entities are used to represent servicing units that can service only one TRANSACTION at a time. STORAGE entities, on the other hand, can service multiple entries at any given time. Associated with each of the entities is a subset of attributes designated as Standard Numerical Attributes (SNA's). These are primarily referencing aids, e.g., Q1 is the current length of QUEUE 1, S4 is the current contents of STORAGE 4, P46 is the value in transaction PARAMETER 46, etc. They are often used in control or decision-making situations, such as testing whether a QUEUE has a certain length, or if a FACILITY has serviced a certain number of units. These give the GPSS language the ability to dynamically modify the parameters of a system.

LOGIC SWITCHES are used to control the flow of TRANSACTIONS through the system. They are tested by using a GATE block to determine if certain conditions prevail that will allow the TRANSACTION to continue through the system.

The statistics gathered by the GPSS model are accumulated in DISTRIBUTION TABLES defined by the modeler. The quantities that may be tabulated are the values of any SNA's, the difference between two consecutive SNA's, interarrival rates of TRANSACTIONS, and the time that a TRANSACTION remained in the system.

The operational entities include ADVANCE blocks, which simulate the passing of time by a TRANSACTION, and TRANSFER blocks, which allow a TRANSACTION to choose alternative routes through the system in a pseudo-undetermined manner. SAVEVALUES and MATRIX SAVEVALUES are used to store constants used within the system.

VARIABLES are arithmetic combinations of the values of the various SNA's, including other variables. FUNCTIONS express the relationship between two variables, one of which is often the value of an SNA. GPSS/360 also has eight random number generators, which may be used as the independent variable in a FUNCTION definition.

The GPSS program begins operation by the generation of TRANSACTIONS that are moved through the system by the GPSS scan. The system maintains an absolute clock and records the time at which an event is due to occur. Those TRANSACTIONS that cannot move until a later clock time are merged into a FUTURE EVENTS CHAIN ranked in ascending order according to the clock time at which they may next move. All remaining TRANSACTIONS are placed in a CURRENT EVENTS CHAIN ranked on a PRIORITY basis.

The GPSS scan cycles through the CURRENT EVENTS CHAIN until no more TRANSACTIONS may be moved at this particular clock time. At this point, the absolute clock advances to the time of the most imminent departure of TRANSACTIONS from the FUTURE EVENTS CHAIN. The GPSS scan places all the appropriate TRANSACTIONS on the CURRENT EVENTS CHAIN, and begins cycling through the CURRENT EVENTS CHAIN, starting with the TRANSACTION with the highest PRIORITY.

The GPSS scan now attempts to move each TRANSACTION through the system as far as it will go. This sequence is repeated until there are no TRANSACTIONS within the system, or the program is terminated by outside means.

The overall length of the run may be specified either in terms of clock time or by the number of TRANSACTIONS that are to pass through the system. At the end of a run, certain standard statistics may be printed out, including facility statistics and storage statistics, which consist of the number of entries, average length of time spent by transactions within FACILITIES or STORAGES, average utilization, and average contents of the storage entity.

Other automatically gathered statistics include a tabular printout of QUEUE statistics. Also, at the user's option, a printout of the contents of the CURRENT and FUTURE EVENTS CHAINS is available.

THE SYSTEM MODELED

Equipment and Facilities

The system under consideration has equipment at two locations. The terminal equipment in SRL consists of an IBM 2540 card reader for input and an IBM 1403-2 printer for output. (Since all of the equipment considered in the model was manufactured by International Business Machines, subsequent references will be made by model number only, e.g., 2540, 360/65.) All jobs from SRL are submitted through the terminal, and with the exception of the few jobs which must be run on the 7090 Emulator, all are entered into the system from this location.

The central processing unit (CPU) is located in an adjacent building approximately 1000 feet from SRL. The equipment at this Computer Center (CC) includes a 360/65, a 360/30, a 2540 card reader, and a 1403-2 high-speed printer used by the 360/30. Also available are two 2403 tape control units each with a 9-track drive, four 2402 tape units with 9-track drives, and four 2402 tape units with 7-track drives. Of the six 9-track tape units, two are reserved for input and output operations through SRL, two are used by the 360/65 for input and output of the current job being processed, and two are available for mounting special volumes needed by the job being processed. Should a job require three or four private volumes, the I/O operations between SRL and the CC are temporarily closed down. Also with the CPU, there are five 2311 disk units, but only one may be accessed by a user; the remaining four serve as direct access storage for the system.

The 360/30 plays a minor role in the model. It is used by SRL as an auxiliary printer during the evening and night shift if a backlog exists for the 1403-2 printer in SRL. Should the printer in SRL break down, however, the 360/30 and printer are preempted and used in conjunction with the 360/65 to process jobs from SRL.

Personnel

The work week consists of 15 shifts, with shifts running 12-8 a.m., 8-4 p.m., and 4-12 p.m. At the terminal location, there are two operators and one clerk on duty from 8 to 4 p.m., and one operator on duty each of the other two shifts. Two operators and one clerk are on duty at the 360/65 during the day shift. On each of the other two shifts, only one operator is on duty with the 360/65. There is also one operator on duty at the 360/30 on each shift.

The distinction between operators and clerks lies in their respective functions. Operators are responsible for the actual operation of the equipment; they feed cards into the reader, mount tapes, and perform similar tasks. Clerks, however, have no contact with the machines, but perform paperwork associated with each job and provide security measures for incoming jobs requiring it.

The final person considered in the system is the messenger. He makes one round trip per day from the CC to SRL. His duties consist of bringing output from the previous night's printing by the 360/30 to SRL and returns to the CC with any new emulator programs submitted since his last trip.

Job Characteristics

Jobs to be processed originate at the terminal location between 8 a.m. and 4 p.m. They are divided into separate categories, each with clearly defined characteristics. They are defined as follows:

Category	Characteristics
1	- Express runs. Run time less than 6 minutes, printed output less than 75 pages. The assumption is made that 75 pages is equivalent to 1750 lines of output. No special tapes or disks required. Priority 100 during the day shift, but low priority at night.
2	- Regular runs. Run time less than 20 minutes. No more than two special tapes and/or one disk. No special handling required. Will run during the day when Category 1 is empty. Will run at night before any other job. Priority 90.
3	- Long runs. More than 20 minutes run time or more than two special tapes and/or one disk required - less than 2 hours. Priority 80.
4	- Classified jobs. Special control clerk handling. Priority 70.
5	- Special handling required by clerk but not classified. Priority 60.
6	- Emulator jobs. Sent by messenger to the CC to be put on tape and run there. Priority 50.
7	- Very long jobs - 2 hours or longer. Priority 40.

Procedure to be Simulated

Jobs are submitted at the terminal and are placed in queues according to their category. During the day shift, the jobs in Category 1 have top priority and are placed onto tape at the CC via the card reader. To make this input tape, an operator mounts a scratch or blank tape on a unit, and signals the operator at the terminal to feed cards into the card reader. When the tape is completed, the operator dismounts the tape and releases the tape unit from service. A tape is considered complete when the number of jobs or cumulative run times exceeds predetermined limits for each category. As the queue of Category 1 jobs decreases, tapes of jobs from the remaining categories are made and placed in queues waiting for the CPU.

When a tape is completed, the operator determines whether it is the correct time of day to run a tape of that category. If not, then the successive tapes in the queue are examined to find one that is appropriate to run. If there are no such tapes, then the first tape is run.

If the operator now finds that the CPU and two tape units are available, the input tape is mounted on one of the tape units and a scratch tape is mounted on the second to accumulate output. The operator next determines whether any private volumes are required for the first job on the tape. If there are, he obtains them from the tape library. At this point, the job starts; when it is completed, the procedure described above is repeated. When all jobs on the tape have been run, the operator dismounts the input tape, returns it to a pool of scratch tapes, dismounts the output tape, and places it in a queue waiting for the printer. If more than two tape units are required to mount private volumes, the I/O link with the terminal is closed before the job is processed. If private volumes are required, the operator searches for these volumes, finds an available tape unit, mounts the volumes, and then signals the computer to proceed. The time required to process each job is given as a function of computer steps and I/O steps and varies for each job. If the printer at the terminal operates simultaneously with the CPU, the run time is increased 5%. As each job is completed, the operator dismounts any private volumes that were used and returns them to the library.

If the time is between 8:00 a.m. and 4:00 p.m., the output tape is placed in a queue to be printed on the printer at the terminal. If the output tape is completed during the evening shift, it may be printed on the terminal printer or on the printer connected to the 360/30.

As the jobs are printed, the output is stacked next to the printer and when approximately five jobs have been printed, an operator removes the stack of jobs from the printer, separates them, and places them in the boxes belonging to the originator. A variation from this routine occurs at night when the operator is likely to let more than five jobs accumulate before he separates then.

The only deviation from the above routine is found in jobs from Category 6. These jobs are carried to the CC by the messenger each day and placed on a tape. When no other jobs remain in the system, the operators close down the Operating System (OS) and bring up the 7090 Emulator. When all of these jobs are completed, the Emulator is closed down, and the OS is reinitialized.

THE SIMULATION MODEL

Equipment and Facilities

The mnemonics for the equipment entities considered by the model are as follows:

Location	Equipment	Mnemonic
SRL	1403-2 Printer	PRNT1, PTN1
	2540 Card Reader	READR, CPU1
CC	360/65	CMPTR, CPU2
	360/30	PRNT2, MOD30
	2402, 2403	TAPE1 - TAPE4
		TAPE9, TAP10
	2311	DISK
	360/65 used as a 7090 Emulator	EMLTR

The remaining equipment in the system played only a minor role in the model and was not considered in any detail. Where a piece of equipment has two mnemonics, i.e., CMPTR and CPU2, one was used to control program logic and the other was used to gather statistics. A system flow chart is given in Figure 1.

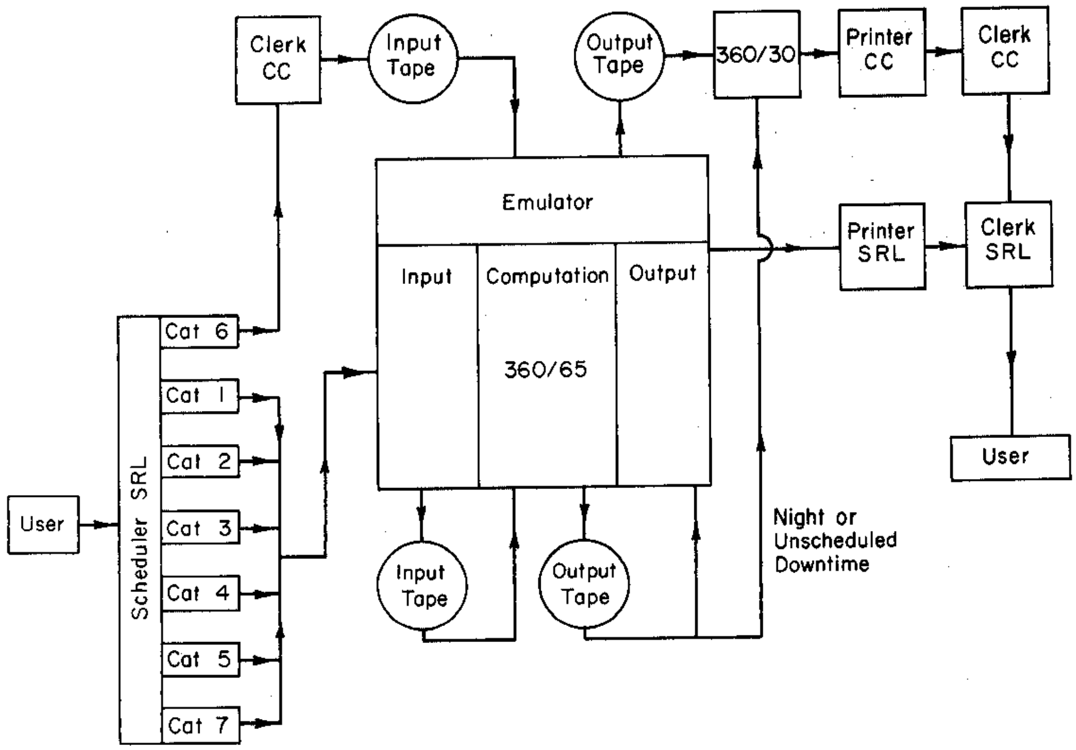

FIG. 1 SYSTEM FLOW CHART

Operating Procedure in the Model

Jobs are generated between 8:00 a.m. and 4:00 p.m. at a rate comparable to a typical input stream at SRL. Each job is assigned certain characteristics, which are retained throughout the simulation. These characteristics are assigned to the jobs in a random manner from a distribution that represents the actual characteristics of jobs at SRL. These characteristics include:

Number of cards in the program deck.
Processing time on the computer.
Expected lines of output.
Number of private tape volumes required.
Number of disks required.
Security classification.
Building of origination.
Amount of clerk-handling required.

If special clerk-handling is required, the job is taken by the clerk in SRL and processed before it enters the input queue. Each job is sorted into one of the seven categories. During the day, jobs from Category 1 (Express), are processed immediately, while the remaining jobs are placed in queues at the terminal. The first job in the Express queue seizes the input partition of the computer, CPU1, and requests that an input tape be mounted. When this is done, the job is read into READR and placed on tape. The remaining jobs on the Express queue are read into READR until the number of jobs on the tape reaches a predetermined maximum, or until the accumulated run time of the jobs on the tape reaches or exceeds a predetermined maximum. When this happens, the tape is closed out and the operator dismounts the tape and places it in the run queue. The next job on the Express queue now starts the process again, and the cycle continues until there are no further jobs in the Express queue. Then the procedure is repeated for each of the other categories until all of the jobs are placed on tape.

As soon as a tape is placed in the run queue, a determination is made whether the first tape in the run queue can be processed by the CMPTR. If it can, then two tape units are seized for input and output to CPU2 and the tape begins to run. The jobs are fed into CMPTR one at a time. If any private volumes are required for any job, the required volumes are obtained from the library and mounted on available equipment. If three or four private tapes are used, then all I/O operations with the terminal are temporarily halted. Each job on the tape is processed in turn, and the results are put on the output tape. As each job is completed all private volumes that were required are dismounted, and the tape units are freed for the next job. The completed output tapes also form a queue. If the tape joins the queue after 4:00 p.m., it goes to the night print routine; if it joins before 4:00 p.m., it goes to the day print routine.

In the day print routine, the tape waits until an operator is available to seize the printer PTN1 and mount the tape. A switch is then set to prevent a second tape from being run on the printer until the first one is finished. Each job seizes the printer, advances the time required to print the output, and then releases the printer. As the jobs stack up at the printer, the operator at the terminal (if not busy) removes the jobs in batches of five or more, separates them, and distributes them into the proper boxes. When a tape ends, the operator at the terminal removes the remaining jobs and distributes them, while the operator at the CC dismounts the output tape.

The night print routine is similar to the day print routine in most respects. If PTN1 is in use, the operator attempts to seize MOD30. If both are in use, he waits until either is free and then seizes the free machine and mounts the tape. Since the MOD30 is printing at the CC, the jobs are broken down there and then brought by messenger to the terminal. The operator at the terminal now allows ten jobs to stack up before separating them and distributing them to the boxes.

Emulator

Category 6 jobs form a queue at the terminal until the messenger arrives and carries them to the CC. The assumption is made that these jobs are placed on tape sometime during the day so that when the 7090 Emulator is brought up sometime late in the evening or early morning, they are on tape ready to run. This assumption will have very little effect on the model. After the Emulator jobs have been run, the Emulator is closed down and the 360/65 Operating System is reinitialized. The output from the Emulator jobs is printed by the MOD30 (PRNT2) and is returned to the terminal by the messenger the following morning.

INTERRUPTIONS IN THE OPERATION OF THE MODELED SYSTEM

Preventive Maintenance

Each Tuesday and Thursday at approximately 12:00 noon, all equipment is stopped after completing the job currently being processed. The system stays down for 2.5 ±.5 hours and is returned to operation. This represents the time that the IBM engineers use to perform routine maintenance.

Initial Program Load (360/65)

The initial program load (IPL) is a program which is used to load the Operating System into the hardware of the computer system. This is always done at the start of the work week. It is also necessary to re-IPL whenever the Operating System collapses due to a system malfunction or operator error.

Unscheduled Downtime

Unscheduled downtime is an unexpected failure of one or more pieces of equipment and is handled in different manners depending on which piece of equipment has failed. These possibilities are:

- CPU2 Failed - The entire system shuts down for 3 ±2 hours. Accounts for 70% of the unscheduled downtime.
- CPU1 Failed - Accounts for 10% of the unscheduled downtime. Two possibilities are: (1) less than 2 hours - 6% of the time, and (2) more than 2 hours - 4% of the time. Both are handled the same way. In reality, if (2) occurs, the jobs are sent to the CC by messenger to be put on tape, but this procedure has not been included in the model. Instead, each job is delayed an appropriate amount of time before being allowed to proceed through the system.
- PTN1 Down - Accounts for 10% of the unscheduled downtime. The MOD30 is immediately seized and used as a printer, and the jobs are sent by messenger to the terminal. The MOD30 prints 1.5 times as fast as PTN1.
- Three or More Disks Down - The system can operate with two 2311 disks down. If three or more go down, the entire system shuts down as though CPU2 went down. This accounts for 10% of the unscheduled downtime.

OTHER CONSIDERATIONS IN THE MODEL

Scheduled Business Applications

Each day, the 360/65 is seized at 10:00 a.m. and at 2:00 p.m. to run commercial data processing jobs. These jobs account for a total of 10-13% of the run time on the 360/65, and each run lasts for 1.25 ±.25 hours, after which the machines are again available to SRL users.

Abort Ratio

Since the assumption that all jobs will run as expected is unrealistic, only 64% are assumed to run with the expected computing time and lines of output. The remaining 36% are considered to have aborted, and their run time is less than or equal to the original run time with their output somewhere between 0 and 1.5 times the expected amount.

External Usage of the 360/30

A routine is included which treats the MOD30 as though it is occupied 100% from 8:00 a.m. to 4:00 p.m. and occupied 50% of the time between 4:00 p.m. and 8:00 a.m. No real statistics were available. Since the 360/30 is used in support of work outside SRL, it is not used during the day shift unless the 360/65 experiences unscheduled downtime during the day. During the two night shifts, it is in use about 50% of the time and functions as an auxiliary printer the remainder of the time.

Ratio of Computing Time to I/O Time

In order to study the effects of faster CPU and I/O devices, the run time must be separated into components representing computing and I/O operations. The total run time is assigned to each job from a distribution representing the actual run times obtained from job-shop records. This parameter is redefined internally as follows:

$$(\text{Run Time})_{ADJ} = (\text{Run Time})_{ORIG} \left[P(K_0 - K_1) + K_1\right]$$

where

K_0 - ratio of the 360/65 compute speed to the speed of the CPU under consideration

K_1 - ratio of the 360/65, 2311 disk I/O speed to the I/O speed of the devices under consideration

P - percent of the total time spent in the computing step of each job (Values of P are given by the distribution of Figure 2)

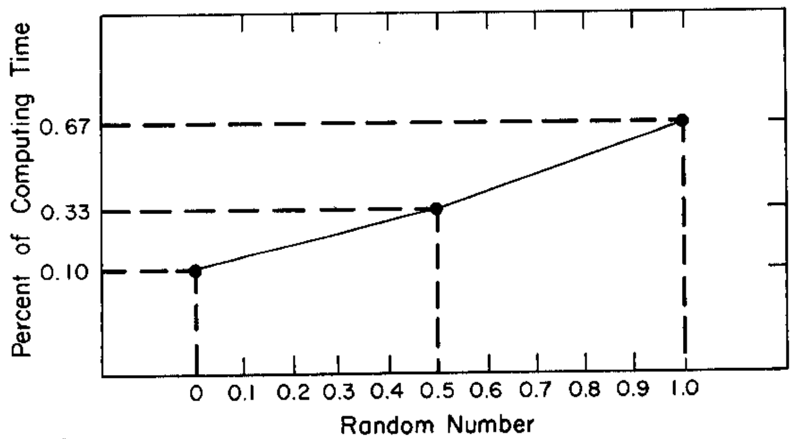

FIG. 2 RATIO OF COMPUTING TIME TO I/O TIME

For the purpose of this study, the computing speed of the 360/75 was considered to be 1.40 times faster than that of the 360/65; and the computing speed of the 360/65 was considered to be 3 times faster than that of the 360/50. Also, the 2301 drum was assumed to be 3 times as fast as the 2311 disk when used for user I/O operations.

The use of this procedure is illustrated in the following numerical example. Consider an Express job with a run time of 24 units and compare the 360/65 CPU equipped with a 2311 disk I/O to a 360/75 CPU equipped with a 2301 drum I/O. Then,

$$K_0 = \text{ratio of } \frac{360/65}{360/75} \text{ speeds} = \frac{1.0}{1.4} = 0.70$$

$$K_1 = \text{ratio of } \frac{2311 \text{ disk}}{2301 \text{ drum}} \text{ I/O speeds} = \frac{1.0}{3.0} = 0.33$$

Consider the value of P obtained from Figure 2 to be 0.50. Then,

$$(\text{Run Time})_{ADJ} = 24 \left[(0.5)(0.37) + 0.33\right] = 12$$

Job Input Control

If jobs do not get through the system within a certain period of time, the rate of input is assumed to decrease considerably because there are only a limited number of jobs available from the users. Much of the job throughput for a given day consists of jobs in the debugging stage, which are submitted several times during the day. If one of these jobs does not return, the user does not have an opportunity to resubmit it that day. A built-in function automatically decreases the rate of input when the number of jobs within the system exceeds 200. Should as many as 400 jobs be within the system, the input is reduced to zero. When the number of jobs within the system is decreased to 150, the job input rate is returned to normal.

Relation Between Input and Throughput Rates

The job input rate for each run was controlled by permitting only a determined percent of the jobs generated to enter the system, with the remainder terminating. Because this was done by the program on a probabilistic basis, a job input rate of 350 jobs/day might result in an actual input of 345 one day and 360 the next, with an average of 350 jobs/day for the entire run. Under ideal conditions, the job input rate and the job throughput rate should be equal. However, if the system becomes saturated, the job input control portion of the model would decrease the input rate, and the job throughput for the model configuration would be decreased correspondingly. In general, when the nominal job input rate is higher than the job throughput rate, it may be assumed that the system was saturated for portions of the run.

VALIDATION OF THE MODEL

Before a model can be effectively used to determine or predict performance, every effort must be made to ensure that the model is a true representation of the system being modeled. To accomplish this, it is first necessary to ascertain that the postulates and assumptions underlying the model represent the system. This done, further effort must be made to guarantee that the model behaves in agreement with the real system. Often the degree of verification of a model such as this is limited by the complexity of the model and the original assumptions on which the model was based. Although accurate statistics with which to compare the results of the model were unavailable, the general trend of the results establishes that this model represents the Computer Center job shop adequately.

In the case of a simulation model such as described in the previous section, where a pseudo-random number sequence plays an important role in the behavior of the model, it is advisable to determine if the behavior of the model is dependent upon the sequence of random numbers provided. GPSS provides for eight distinct random number sequences so that the modeler has no difficulty in providing different random number sequences for the overall model. To dispel any doubts that the behavior of the model was a result of a favorable random number sequence, a series of three stability runs were made before embarking on the runs described in the section on Experimental Design. Because the function that assigns the CPU processing time is a sensitive point within the model, the three stability runs were made using a different random number sequence to assign CPU processing times to the jobs as they were created. The resulting statistics were compared using chi square and analysis of variance techniques, and no significant difference in the runs was observed at the 99% level.

Another type of stability run was made to determine if the proposed 4-week run length would be sufficient to allow the effects of preventive maintenance and unscheduled downtime to settle into a steady-state condition. Runs of 10 and 12 weeks were made and the results were compared with runs of 4 and 1 week. Runs of 1 week showed a very marked difference in results because the unscheduled downtime could occur either 0, 1, or 2 times while runs of 10 and 12 weeks duration varied little from those of 4 weeks, except that unscheduled downtime occurring at random intervals did not occur frequently enough in a 4-week run for its effects to be smoothed out. Also, a slight degradation in mean turnaround time for the various categories was apparent in the longer runs.

VALIDATION OF THE MODEL PARAMETERS

For the model to represent the real system, the parameters within the model must correspond closely with those in the real system, both in their position within the model and in its value. For this reason, many of the input and constant parameters were determined by a statistical analysis of data from the Computer Center. Within the model, the areas where real statistics are used included the following items:

Job Characteristics

A continuous cumulative frequency distribution of the numerous job characteristics was obtained using data from the Computer Center from a one-month period. These cumulative distributions were obtained for the following quantities:

- Number of cards per job.
- Number of lines of output per job.
- Computing time per job.

Further, discrete distributions were obtained for each of the following characteristics:

- Number of private tape volumes required by each job.
- Number of jobs which required a 2311 disk unit.
- Classified or unclassified.
- Number of jobs that required special clerk-handling for reasons other than classification.
- Frequency of job aborts.

Preventive Maintenance

The length of time required by the IBM engineers to perform preventive maintenance on the system was obtained from the System/360 log. The time was averaged over a 20-week period and was found to be 4.95 hours per week. Because preventive maintenance occurs twice a week, an average time of 2.5 hours per occurrence with a uniform spread of .5 hour on either side (2 to 3 hours) was believed to be a valid approximation to the data.

Unscheduled Downtime

The lengths of the periods of unscheduled downtime were also obtained from the System/360 log. Over the same 20-week period, this amounted to an average of 7.25 hours per week. Although the frequency of the occurrences of these downtimes was not known, an average of 2 per week was considered reasonable. Because the length of the runs of the model allowed for only a small number of occurrences of the unscheduled downtime, these occurrences were chosen at predetermined intervals within each model.

Initial Program Load

A study of an 8-week period showed it was necessary to re-IPL the System/360 an average of 60 times per week. From this study, it was decided to force the model to re-IPL every 2.5 $\pm$0.5 hours, each with a duration chosen from a uniform distribution from 2 to 6 minutes.

Further Considerations

Certain data were not readily obtainable or were not believed to be significant enough to warrant further study. In either case, the data were estimated by persons familiar with the area and the estimate was used by the model. These statistics included:

- The time required by an operator to mount and dismount scratch tapes was defined in terms of an exponential function ranging from 1 to 5 minutes.
- The time required by an operator to mount and dismount private volumes was defined in terms of an exponential function ranging from 1 to 20 minutes. The extreme value represents the situation where a specific tape is misplaced by an operator.

- The messenger in the model was scheduled to make one round trip at approximately 7:30 a.m., requiring 15 minutes in each direction. Since the messenger serves merely as a vehicle by which jobs are transported from one location to another, there was no effort to validate this time.
- The distribution of job interarrival times was estimated by discussion with personnel in the Computer Center. The distribution used for most of the runs is assumed to be given in Figure 3a.
- For the instances where the job input stream is increased to 360 jobs/day (Runs 3 and 24), the job interarrival times are assumed to be given by Figure 3b.

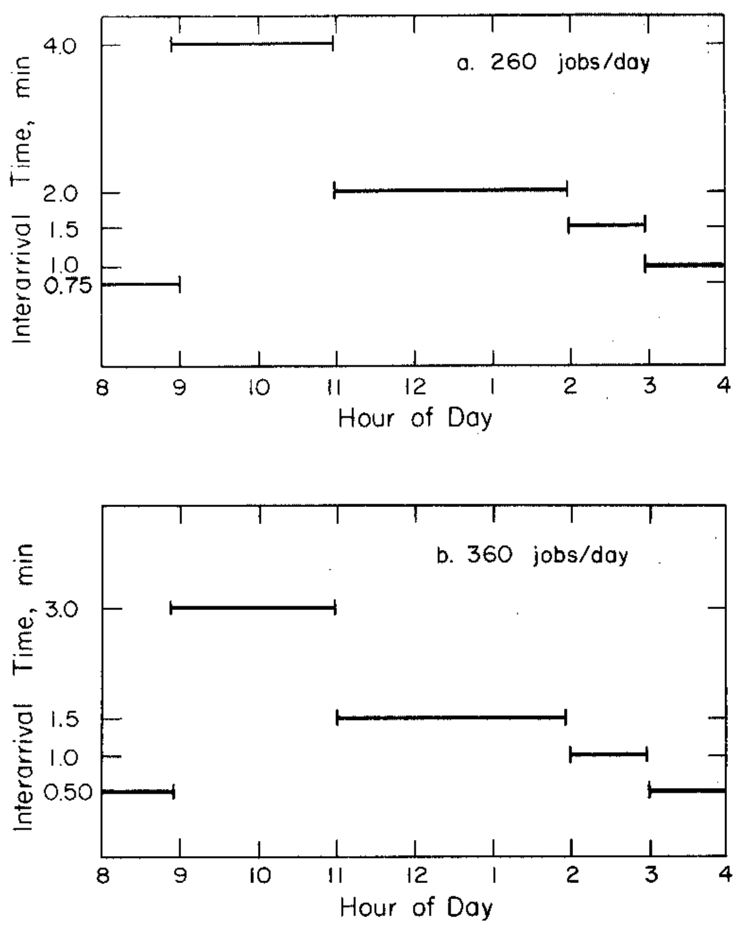

FIG. 3 JOB INTERARRIVAL TIMES

EXPERIMENTAL DESIGN AND RESULTS

To make effective use of this model, a series of experiments were devised to obtain data for comparisons between various operational strategies and various equipment configurations. In Table 1, the following notation will be used.

INPUT RATE = 260	- defines the job input rate to be 260 jobs/day
JOB TAPE = 8-16/12-120	- defines the limits on the job tape made in the Computer Center to be between 8 and 16 jobs/tape, with a cumulative run time between 12 and 120 minutes
EXPRESS = 6/1750	- defines the parameters for the Express category to be less than 6 minutes of run time and less than 1750 lines of output
CPU = 360/65	- defines the central processing unit to be an IBM 360 Model 65
I/O DEVICE = 2311	- defines the input/output device to be an IBM 2311 disk drive
PRINTER = 1403-2	- defines the printer to be an IBM 1403 Model 2, 600 lines/minute printer

Description of Output Statistics

The question of which statistics should be gathered to evaluate a simulation model is generally answered by those who must make decisions based on the results of the simulation. As a representative example of the statistics obtained by the model, the following statistics were tabulated for each of the 24 simulation runs.

- Job input rate.
- Job throughput rate - This number remains relatively constant unless the system performance is such that a large backlog is allowed to build up.

- Average number of jobs completed during the day shift (8:00 a.m. to 4:00 p.m.) - This number is expected to remain relatively constant unless some factor greatly influences the processing speed of one of the main pieces of equipment.
- Jobs which are completed during the day shift of the day submitted, in both Category 1 and Category 2 - This statistic is a good measure of the performance of the system because a large value indicates a short job turnaround time.

TABLE 1

DESCRIPTION OF THE 24 RUNS STUDIED BY THE MODEL

Series	Run	Input Rate	Job Tape	Express	CPU	I/O	Printer	Comments
I	1	260	8-16/12-120	6/1750	360/65	2311	1403-2	
	2	180	8-16/12-120	6/1750	360/65	2311	1403-2	
	3	360	8-16/12-120	6/1750	360/65	2311	1403-2	
II	4	260	4-8/6-60	6/1750	360/65	2311	1403-2	
	5	260	4-8/3-30	6/1750	360/65	2311	1403-2	
	6	260	2-4/3-30	6/1750	360/65	2311	1403-2	
	7	260	12-24/18-120	6/1750	360/65	2311	1403-2	
	8	260	16-32/40-120	6/1750	360/65	2311	1403-2	
III	9	260	8-16/12-120	2/1750	360/65	2311	1403-2	
	10	260	8-16/12-120	15/1750	360/65	2311	1403-2	
	11	260	8-16/12-120	2/4000	360/65	2311	1403-2	
	12	260	8-16/12-120	6/4000	360/65	2311	1403-2	
	13	260	8-16/12-120	15/4000	360/65	2311	1403-2	
	14	260	8-16/12-120	FIFO	360/65	2311	1403-2	
IV	15	260	8-16/12-120	6/4000	360/65	2311	1403-2	No scheduled business applications
	16	260	8-16/12-120	6/4000	360/65	2311	1403-2	Increased scheduled business applications
	17	260	8-16/12-120	6/4000	360/65	2311	1403-2	Preventive maintenance at 6-8 a.m. on Tues and Thurs
	18	260	8-16/12-120	6/4000	360/65	2311	1403-2	Preventive maintenance 6-8 a.m. Tues and Thurs and no business applications
V	19	260	8-16/12-120	6/4000	360/65	2301	1403-2	
	20	260	8-16/12-120	6/4000	360/50	2311	1403-2	
	21	260	8-16/12-120	6/4000	360/75	2311	1403-2	
	22	260	8-16/12-120	6/4000	360/75	2301	1403-2	
	23	260	8-16/12-120	6/4000	360/75	2301	1403-N1	
	24	360	8-16/12-120	6/4000	360/75	2301	1403-N1	

- Job turnaround time - An adjusted job turnaround time was defined to reflect the actual environment of the Computer Center. Because the users from SRL work only the day shift, only those jobs completed before 4:00 p.m. each day are of immediate value. For those jobs, the adjusted job turnaround time was defined as the actual time in the system. For jobs which are not completed before 4:00 p.m., it is immaterial to the user when they finish as long as the completed job and output are available at the start of the following day. For these jobs, the adjusted job turnaround time is defined as the time the job was in the system from the time of input until 4:00 p.m. For jobs which are input on one day and not returned by 8:00 a.m. the following morning, the adjusted job turnaround time is given by the time spent in the system on the first day until 4:00 p.m. plus a penalty function of twice the time spent in the system on the second day before the job returns. This penalty function was chosen to reflect the importance attached to returning a job to the user by the morning of the day after submittal. For jobs which have not been completed by 4:00 p.m. the second day, the adjusted job turnaround time is defined as the total time the job was in the system.

  The mean adjusted job turnaround time was collected for each of the seven categories. Two other tables tabulated the cumulative percent of jobs that were returned to the user within a certain length of time. One table used the adjusted job turnaround time, and the second used the absolute job turnaround time. A final set of tables tabulated the same statistics for the adjusted turnaround time for each of the seven categories.

Because the results of a simulation study of a computer job shop are generally used in making decisions regarding proposed changes in hardware, software, and operating strategies, some method of ranking the various alternatives in a definite pattern is necessary. One such technique consisting of the use of a performance index is described in Appendix B.

Series I

This series of runs was designed to illustrate the effects of varying the job input rate for the original model configuration: CPU = 360/65, I/O = 2311, Printer = 1403-2. The job input rate was varied from 180 to 360 jobs/day with the results shown in Table 2. The results obtained were as expected with increased throughput resulting in increased turnaround time. In the case where job input was 180 jobs/day, there were not enough jobs in Category 1 and Category 2 to keep these queues from becoming empty during the day. If this occurs during the day, jobs from Category 3 are permitted to run during the day shift. Since by definition these jobs run from 20 minutes to 2 hours each, a tape with several Category 3 jobs could tie up the computer for several hours during the day, thus decreasing the job throughput for that day by a significant amount.

TABLE 2

OUTPUT STATISTICS FOR SERIES I: INPUT RATES

CPU = 360/65 I/O = 2311 Printer = 1403-2

Express = 6/1750 Job Tape = 8-16/12-120

Run:	1	2	3
Job input/day	260	180	360
Job throughput/day	257	191	269
Average number of jobs completed day shift	85	57	72
Jobs completed day submitted			
Category 1	57	52	33
2	2	3	0
Mean adjusted turnaround, minutes			
Category 1	525	243	1112
2	400	332	450
3	427	375	474
4	504	381	615
5	824	417	662
6	3564	1093	4612
7	867	417	1184
Average idle time, minutes/week	2183	3212	1919

Figure 4 shows the cumulative percent of jobs turned around as a function of the job turnaround time, with decreasing turnaround time occurring as the input level drops. Although the job input rate for Run 3 was set at 360 jobs/day, the system was unable to process these jobs rapidly enough to prevent a large backlog from building up. When this happened, the input rate was reduced by the Job Input Control section of the model, which only allowed a maximum of 400 jobs within the system at any given time. Because of this, the average job throughput was reduced to only 270 jobs/day, which was but a small improvement over Run 1. However, due to the manner in which jobs were input to the model in Run 3, the job turnaround statistics were greatly increased.

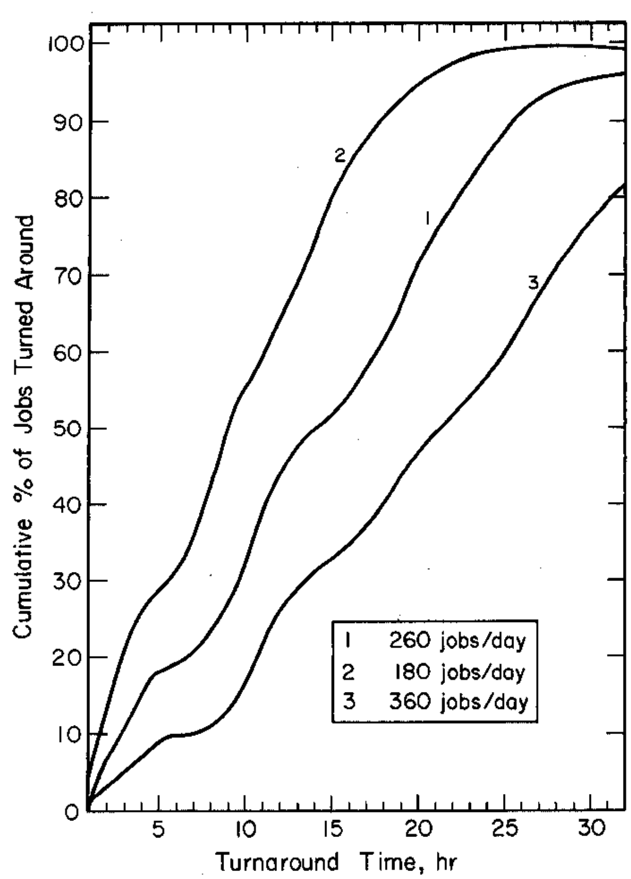

FIG. 4 EFFECT OF INPUT RATE - SERIES I

The level area which occurs in each curve at approximately 5 to 8 hours from submittal is the result of changing the scheduling algorithm at 4:00 p.m. from that of running short jobs first to that of running longer jobs first. Jobs that are completed within 8 hours of the time of input decreases from 43% at an input rate of 180 jobs/day to only 11% at an input rate of 360 jobs/day.

Series II

When the operators in the Computer Center made up the job input tape using jobs from SRL, no specific rules were in effect as to the number of jobs placed on a given tape, or as to the cumulative lengths of job run time of jobs placed on the tape. The second series of runs compares various alternatives to the originally chosen parameters of 8 to 16 jobs and 12 to 120 minutes of run time (8-16/12-120).

Table 3 compares the output statistics for the six runs involved, and Figure 5 presents the cumulative distribution curves for the original parameters and the two extreme alternatives. Because of its overall better performance characteristics (shorter turnaround time for major categories, more jobs completed on day of submittal), the job-tape parameters from the original run (8-16/12-120) were chosen for the remainder of the runs made.

Run 6, which allowed very short job tapes to be made, could not match the performance of the "base case." This can be explained by realizing that a job tape with only 2 to 4 jobs on it would usually have a cumulative run time of only 5 to 10 minutes, and could require an additional 10 minutes of time for operator intervention. With an overhead equal to the run time, the central processing unit can be utilized only 50% of the time, causing an increase in job turnaround times for nearly all jobs. Run 8, however, considers very long job tapes and, after the initial delay each day caused by waiting for the first tape to be made, performs in a manner quite similar to the base case.

TABLE 3

OUTPUT STATISTICS FOR SERIES II: JOB-TAPE PARAMETERS

CPU = 360/65 I/0 = 2311 Printer = 1403-2

Express = 6/1750

Run:	1	4	5	6	7	8
Job-tape parameters						
Number of jobs	8-16	4-8	4-8	2-4	12-24	16-32
Run time, minutes	12-120	6-60	3-30	3-30	18-120	40-120
Job input/day	260	260	260	260	260	260
Job throughput/day	257	239	243	228	258	259
Average number of jobs completed day shift	85	84	75	79	82	79
Jobs completed day submitted						
Category 1	57	44	46	51	39	46
2	2	1	1	0	2	0
Mean adjusted turnaround, minutes						
Category 1	525	885	911	905	689	605
2	400	432	405	617	422	493
3	427	373	384	349	419	416
4	504	423	440	490	452	428
5	824	749	431	536	839	573
6	3564	3768	4071	4339	4040	3667
7	867	1188	754	992	1043	732
Average idle time, minutes/week	2183	2350	2377	2222	2114	2257

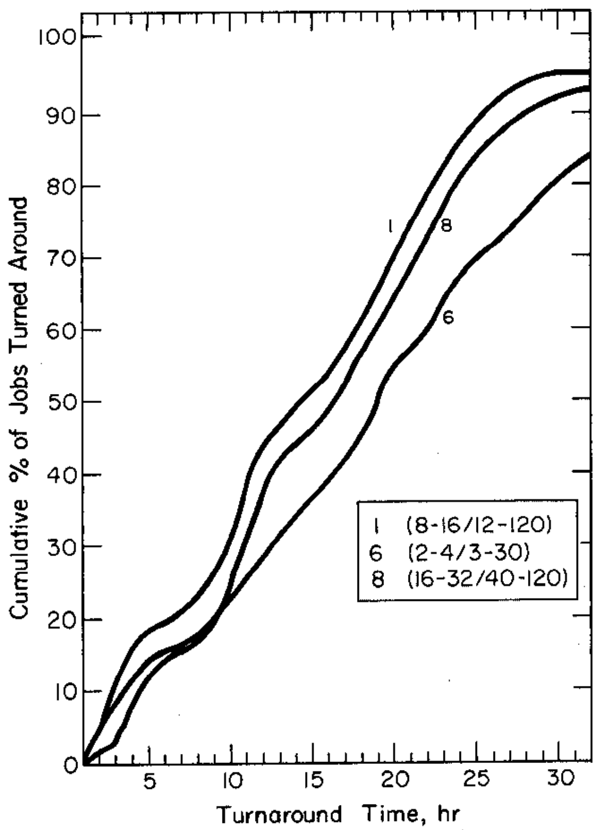

FIG. 5 EFFECT OF JOB-TAPE DEFINITIONS - SERIES II

Series III

The third series was designed to determine the effects of varying the run time and output parameters on the Express category. Runs 1, 9, and 10 compare results obtained by varying the run time parameter and keeping the output limit constant at 1750 lines. Runs 11-13 vary the run time parameter while holding the output parameter constant at 4000 lines. Also included in this series of runs is the classical first-in, first-out (FIFO) queuing situation. In this last model, all jobs are defined as either Category 1 or Category 6 (Emulator).

Table 4 compares the output statistics for these seven runs, while Figure 6 gives the cumulative job turnaround for the second set of runs. When examining the output statistics in Table 4, for Runs 9 and 11, it can be seen that the mean adjusted turnaround times for Category 1 are significantly shorter than those for the other runs. This is because the Express category as

TABLE 4

OUTPUT STATISTICS FOR SERIES III: CATEGORY DEFINITIONS

CPU = 360/65 I/O = 2311 Printer = 1403-2
Job Tape 8-16/12-120

Run:	7	9	10	11	12	13	14
Express parameters	6/1750	2/1750	15/1750	2/4000	6/4000	15/4000	FIFO
Job input/day	260	260	260	260	260	260	260
Job throughput/day	257	259	260	258	258	259	259
Average number of jobs completed day shift	85	75	70	81	88	72	50
Jobs completed day submitted							
Category 1	57	53	44	57	61	43	42
2	2	6	0	6	0	0	0
Mean adjusted turnaround, minutes							
Category 1	525	325	602	433	567	672	445
2	400	742	365	555	377	349	
3	427	396	397	493	414	399	
4	504	409	367	431	408	378	
5	824	580	504	534	502	472	
6	3564	3672	3889	7367	4064	3908	3876
7	867	1347	421	671	453	410	
Average idle time, minutes/week	2183	2094	2191	1474	2075	2215	2289

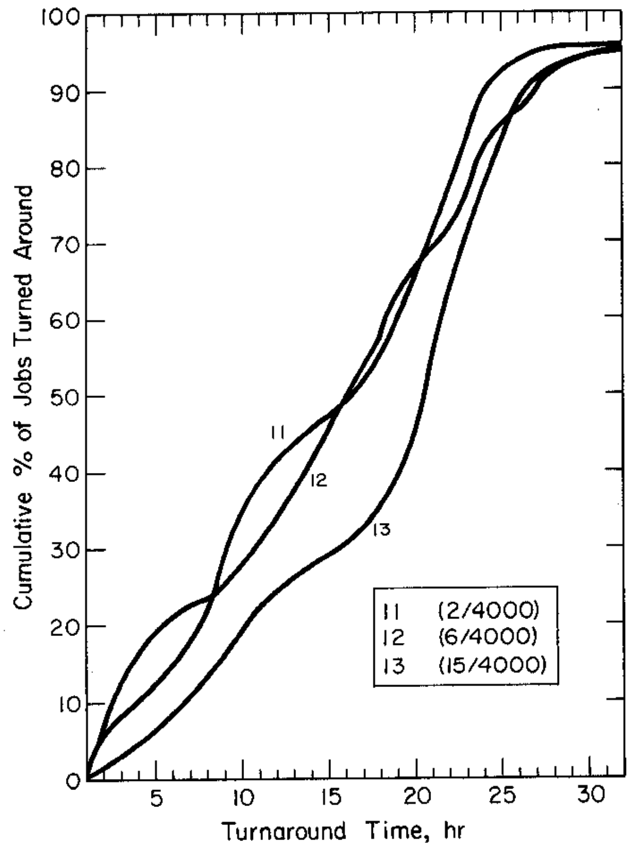

FIG. 6 EFFECT OF CATEGORY DEFINITIONS - SERIES III

defined in these two runs allowed only 66% as many jobs to be classified as Express jobs (because of the 2-minute run time restriction) as compared to the "base case." This made it easier for the model to complete all Express category jobs during the day, and even allowed the Regular (Category 2) jobs to be processed during the day shift. In an analogous manner, Runs 10 and 13 show a very low adjusted turnaround time for Regular jobs because the definition of the Express category leaves very few jobs to be classified as non-Express. Because the job shop under consideration is most interested in jobs which return within 8 hours of input (since these are the only jobs which could return during the day shift), the Express parameters for Run 12 were chosen for the remainder of the runs made in the experiment.

Figure 7 compares the cumulative job turnaround for Run 12 and Run 14 (FIFO). While FIFO does yield a low mean turnaround time, it is not until 45% of the jobs are returned to the user that the cumulative performance curve demonstrates an improvement over the new "base case."

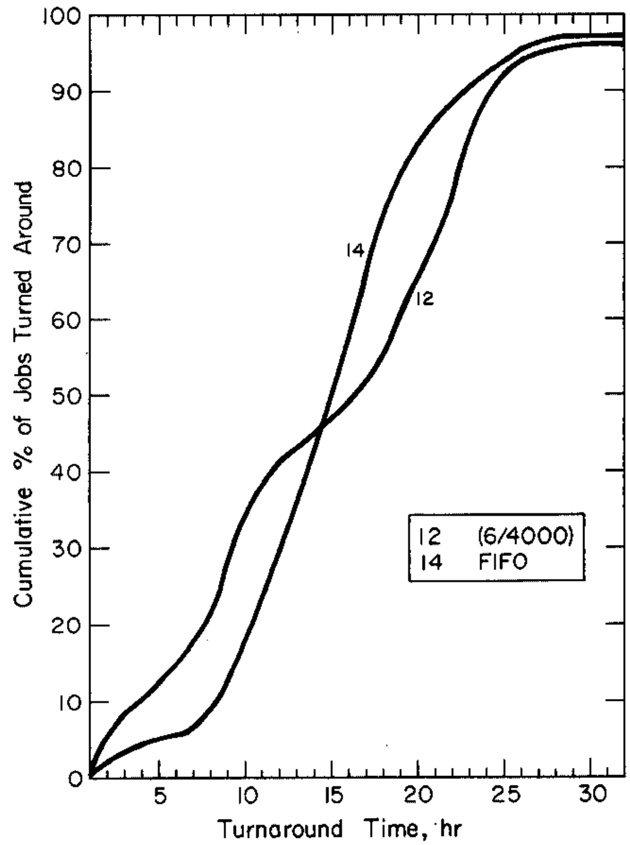

FIG. 7 EFFECT OF FIFO - SERIES III

Series IV

This series of runs considers several major changes in the environment surrounding the Computer Center and evaluates their effects on overall performance. There are three environmental factors which, by their frequency or magnitude, create a discontinuity in the smooth flow of SRL jobs through the computer systems, i.e., preventive maintenance, the use of the 360/65 for business applications, and unscheduled downtime. Because of the unpredictable nature of the latter item, little can be done about it. Because of the requirements of the business applications, it is not feasible to move these programs off the day shift. A natural course of events would lead to the point where the business applications would increase considerably and require additional computer time, and finally culminate by being taken off of the current 360/65 and placed on an equivalent machine dedicated to the processing of business oriented programs.

Runs 15 and 16 show what would occur if the business applications were removed from the 360/65 or increased by a factor of 50%. The results given in Figure 8 show that when the business applications are increased by 50% the jobs returned within 8 hours of input drop from 22 to 12%. Similarly, when

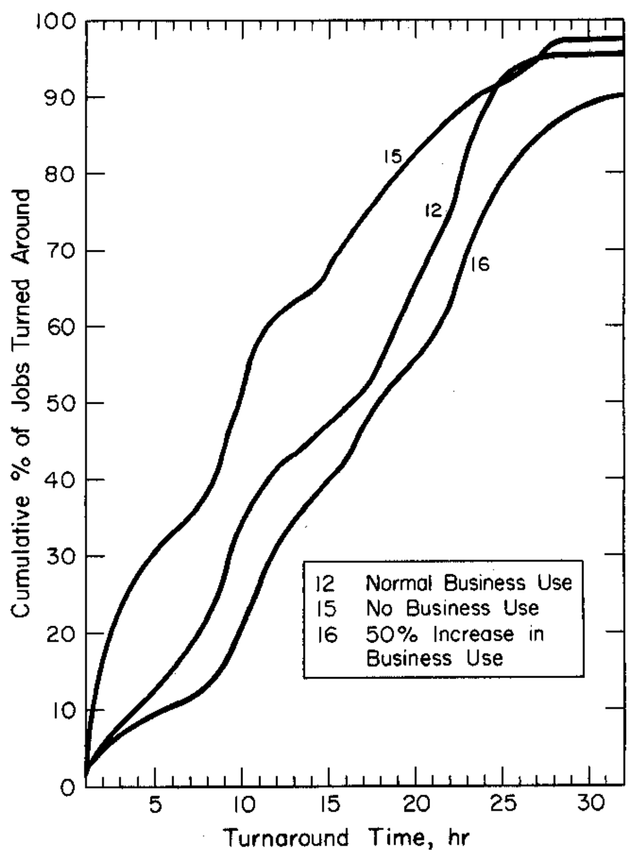

FIG. 8 EFFECT OF VARYING BUSINESS APPLICATIONS - SERIES IV

the business applications are withdrawn from the 360/65, the jobs returned within 8 hours increase to 38%. Run 17 considers the scheduling of preventive maintenance at 6:00 a.m. on Tuesday and Thursday mornings as opposed to the standard time of 12:00 noon on those days. Finally, Run 18 considers preventive maintenance at 6:00 a.m. in the instance where no business applications are done on the 360/65. Table 5 compares the output statistics, while Figures 8, 9, and 10 compare the cumulative distribution curves for various cases.

It should be noted that a rescheduling of the preventive maintenance has a negligible effect on SRL turnaround in the situation where the business applications require 3 hours per day (Figure 9). However, when the business applications are transferred off of the 360/65 (Figure 10), the improvement in the cumulative turnaround curve is apparent. Although there

TABLE 5

OUTPUT STATISTICS FOR SERIES IV: ENVIRONMENTAL FACTORS

CPU = 360/65 I/O = 2311 Printer = 1403-2
Express = 6/4000 Job Tape = 8-16/12-120

Run:	15	16	17	18
Job input/day	260	260	260	260
Job throughput/day	263	248	260	265
Average number of jobs completed day shift	101	64	85	117
Jobs completed day submitted				
Category 1	86	42	48	107
2	3	0	1	6
Mean adjusted turnaround, minutes				
Category 1	380	823	664	238
2	362	435	434	367
3	416	398	403	402
4	444	442	519	425
5	540	421	695	564
6	2848	4053	4055	2497
7	837	845	1273	501
Average idle time, minutes/week	2449	2146	2097	2617

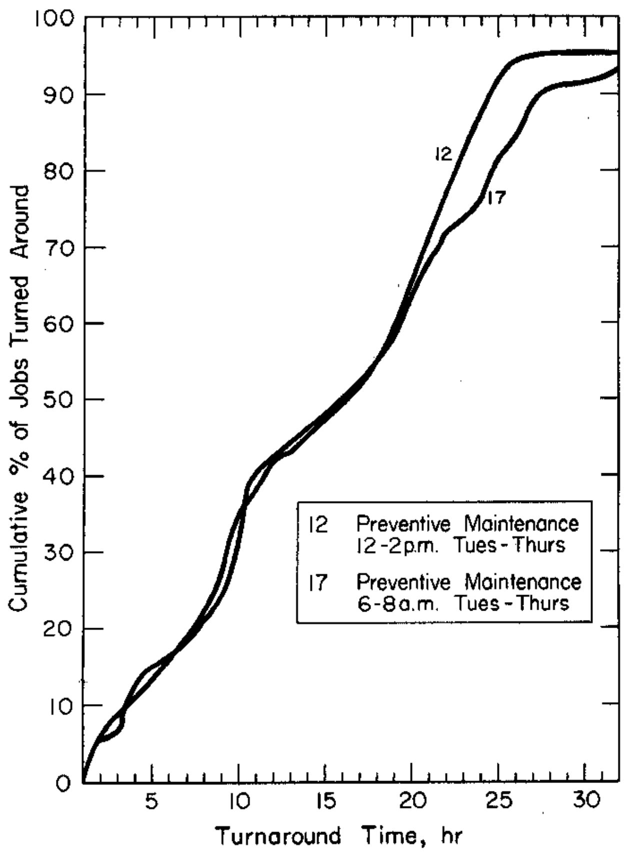

FIG. 9 EFFECT OF RESCHEDULING PREVENTIVE MAINTENANCE - SERIES IV

probably are several factors involved, one possible explanation lies in the fact that the smoothness of the system's operation was already disrupted by the business applications in Run 17, hence, the rescheduling of preventive maintenance at 6:00 a.m. would not allow the system to smooth out its operations. Rather, the system would still have a 3-hour interruption each day rather than a 3-hour interruption for 3 days and a 5-hour interruption for Tuesday and Thursday. In the second instance, where there is no business application usage, the time spent by preventive maintenance is the only outside interruption to the system, and its rescheduling to 6:00 a.m. allows for a virtually uninterrupted 8-hour work day in the Computer Center.

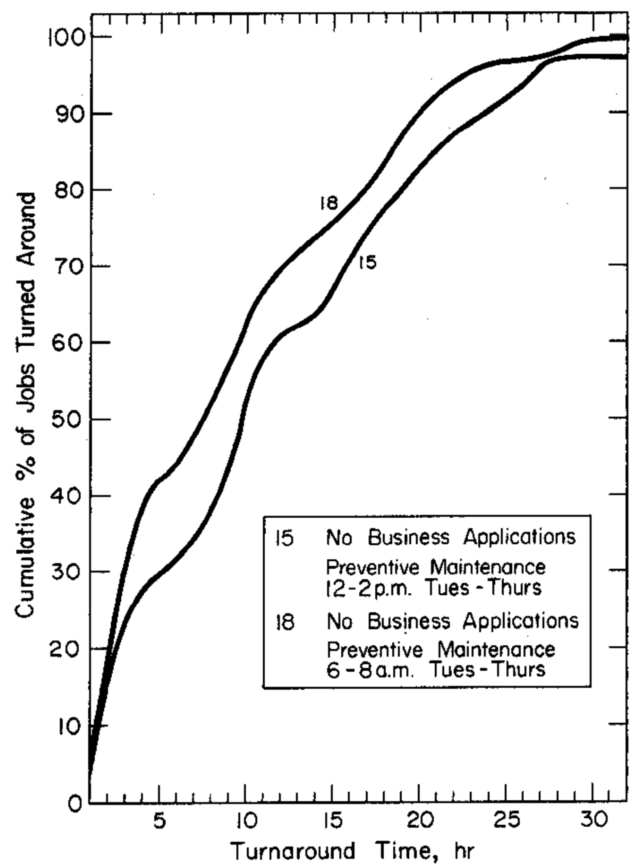

FIG. 10 EFFECT OF RESCHEDULING PREVENTIVE MAINTENANCE WHEN NO BUSINESS APPLICATIONS ARE SCHEDULED - SERIES IV

Although not completely verified by the model, it would appear that these observations hold only for a system that does not consistently have a large backlog at 8:00 a.m. each morning. If this should occur, then the throughput of jobs that return the same day they were input should decrease drastically, and job turnaround should also increase.

Series V

The final series of runs were concerned with the overall performance of the system as various pieces of equipment were replaced by faster or more efficient hardware. Table 6 presents the output statistics for Runs 19-24. Figures 11 to 13 compare the various hardware changes and demonstrate their effects on the cumulative job turnaround times. This series of runs confirmed the expected improvements due to the upgrading of hardware. For example, as shown in Figure 11, the system with the 360/75 outperformed the system with the 360/65, although not by as much of a margin as might be expected since the 360/75 is considered to be a 40% faster machine. This result is attributed to the lower average percent of computing time to I/O time in our job mix for all cases. Figure 12 compares the 360/65 with 2311 disk I/O to a 360/75 with a 2301 drum I/O. The jobs turned around within 8 hours of input increased from 22 to 50%, and 95% of the jobs were returned to the user within 16 hours of input.

TABLE 6

OUTPUT STATISTICS FOR SERIES V: EQUIPMENT CONFIGURATION

Express = 6/4000 Job Tape = 8-16/12-120

Run:	19	20	21	22	23	24
CPU	360/65	360/50	360/75	360/75	360/75	360/75
I/O	2301	2311	2301	2301	2301	2301
Printer	1403-2	1403-2	1403-2	1403-2	1403-N1	1403-N1
Job input/day	260	260	260	260	260	360
Job throughput/day	263	203	262	264	262	362
Average number of jobs completed day shift	80	65	81	76	87	100
Jobs completed day submitted						
Category 1	79	21	68	75	86	93
2	1	0	0	1	1	0
Mean adjusted turnaround, minutes						
Category 1	226	1523	499	245	225	341
2	317	374	360	314	299	323
3	347	782	365	380	361	358
4	371	3003	398	371	379	367
5	477	5672	624	492	371	316
6	995	5759	2955	2399	887	2781
7	395	5240	409	349	352	414
Average idle time, minutes/week	2897	1428	2467	3046	4525	2754

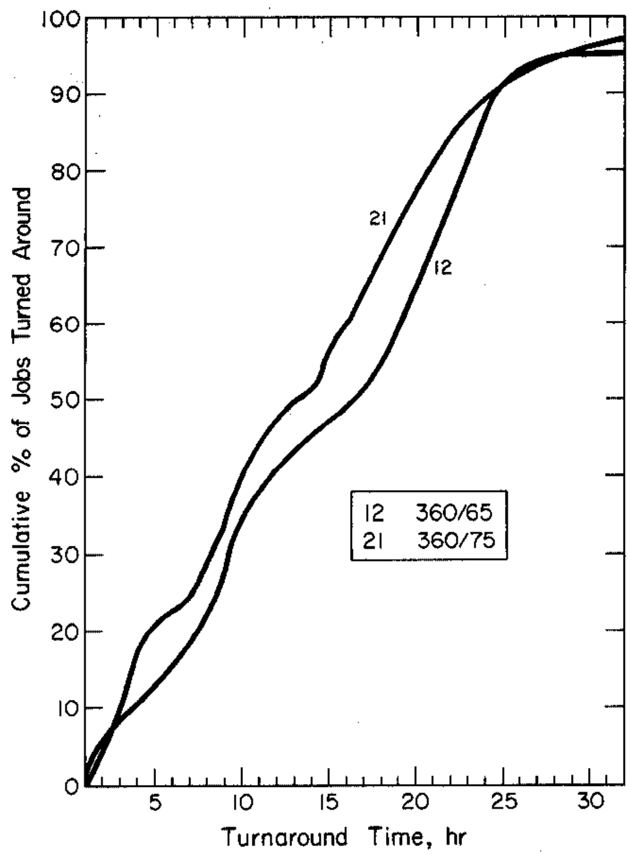

FIG. 11 COMPARISON BETWEEN 360/65 AND 360/75 WITH 2311 DISK - SERIES V

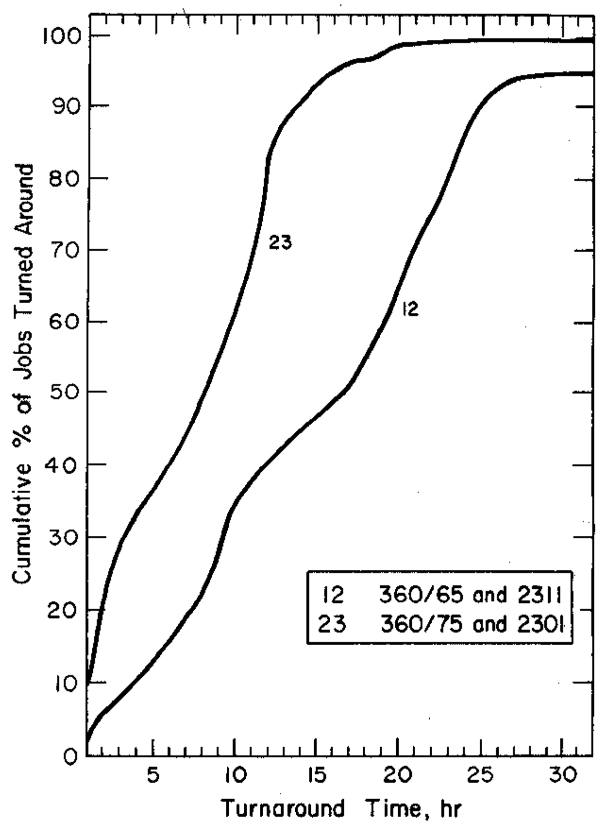

FIG. 12 COMPARISON BETWEEN 360/65 WITH 2311 DISK AND 360/75 WITH 2301 DRUM - SERIES V

Figure 13 shows the resulting curve when a 1403-N1 printer with a print speed of 1200 lines/minute replaces the older 1403-2 printer. For the first 8 hours, an apparent improvement is seen as the Express category jobs are run and the faster printer keeps the length of the print queue small. However, the degree of improvement indicates that increasing the printer speed has little effect on the turnaround time.

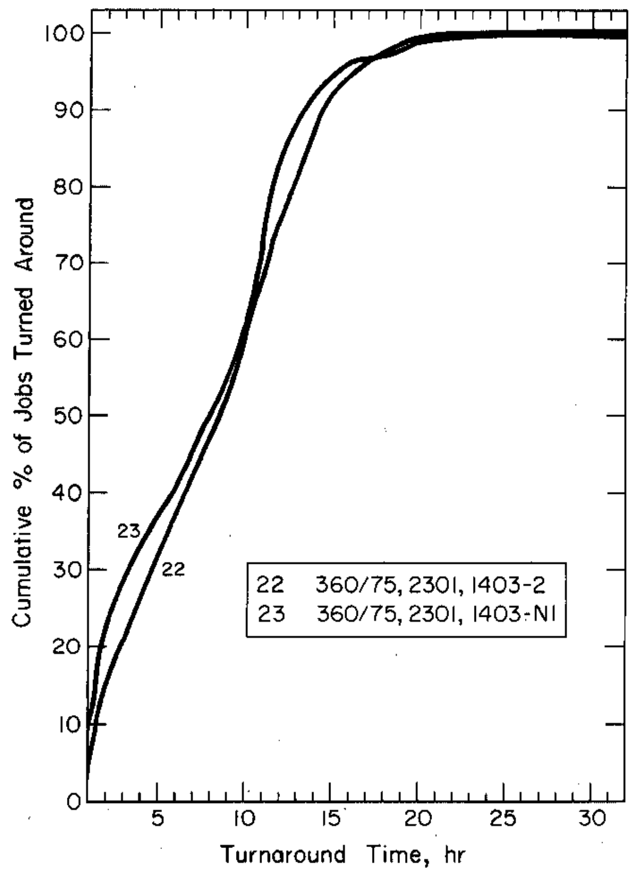

FIG. 13 COMPARISON BETWEEN 1403-2 PRINTER AND 1403-N1 PRINTER - SERIES V

Figure 14 compares the resulting curves for the 360/75 with a job input rate of 260 and 360 jobs/day with the curve for the base case. This comparison shows that an increased input produces a system in which a lower percent of jobs return within a given length of time.

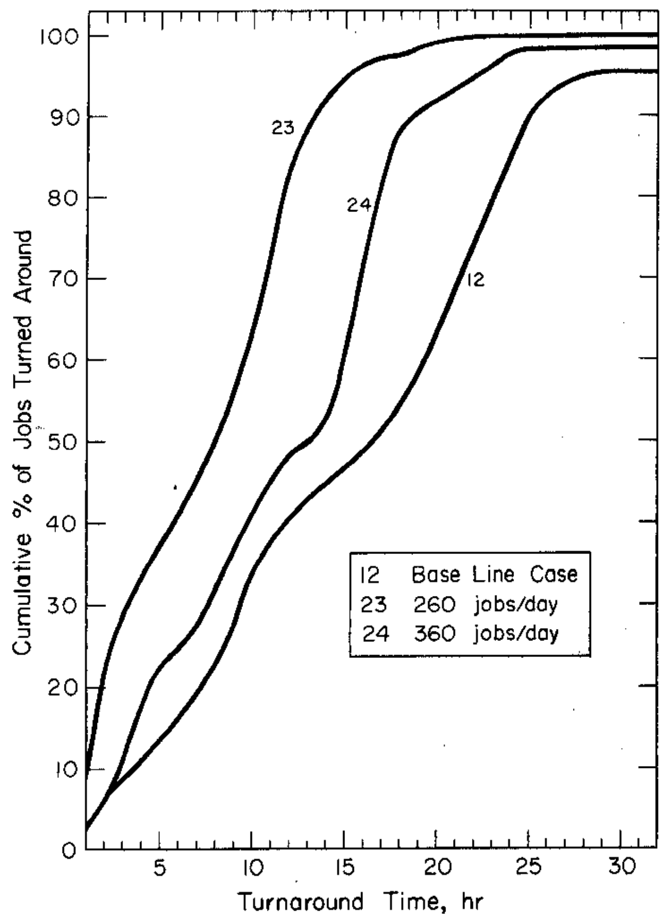

FIG. 14 COMPARISON BETWEEN RESULTS FOR TWO JOB INPUT RATES USING THE 360/75 - SERIES V

APPENDIX A

SYSTEM OVERVIEW

In order to provide a continuous overview of the flow of jobs through each computer job-shop model, certain statistics were collected and arranged in a tabular form to present a 28-day snapshot of the system behavior. In Table A-1, each row represents one day's activity (a day is considered to be from 8:00 a.m. to 8:00 a.m.) and each column contains the following statistics:

Column 1	Number of jobs in system at 8:00 a.m.
Column 2	Number of jobs input during day.
Column 3	Number of jobs which are completed from 8:00 a.m. to 4:00 p.m.
Column 4	Number of jobs which are completed the same day they were input.
Column 5	Number of jobs in system at 4:00 p.m.
Column 6	Number of jobs in system at 12:00 p.m.
Column 7	Number of jobs which are completed from 8:00 a.m. to 8:00 a.m.

TABLE A-1

SYSTEM OVERVIEW FOR RUN 1

Day	1	2	3	4	5	6	7
1	0	251	92	92	159	92	251
2	0	251	29	29	222	168	166
3	85	259	131	130	213	145	308
4	36	276	49	49	263	170	204
5	107	233	29	4	311	257	189
6	152	0	152	0	0	0	152
7	0	0	0	0	0	0	0
8	0	281	77	77	204	125	221
9	60	245	63	36	242	165	183
10	122	241	133	22	230	154	351
11	12	276	24	24	264	195	239
12	48	244	80	72	212	153	215
13	78	0	78	0	0	0	78
14	0	0	0	0	0	0	0

TABLE A-1 (cont'd)

Day	1	2	3	4	5	6	7
15	0	262	79	79	183	108	262
16	0	280	92	92	188	155	242
17	38	246	125	125	159	79	248
18	36	245	89	82	192	121	227
19	54	256	146	146	164	105	233
20	77	0	77	0	0	0	77
21	0	0	0	0	0	0	0
22	0	254	77	77	177	121	174
23	80	250	86	29	244	169	206
24	124	260	89	0	295	234	270
25	114	286	99	32	301	232	216
26	184	251	142	7	293	213	250
27	185	0	95	0	90	49	185
28	0	0	0	0	0	0	0

APPENDIX B

PERFORMANCE INDEX

A performance index consisting of weighted functions of the various output statistics can be a useful technique for comparing and ranking various alternatives studied using a simulation model. A major objection to this method of evaluation is that some one or some group of persons must place an explicit value or weight on each of the results from the simulation, and a set of weights that will satisfy everyone concerned is often difficult to obtain. This difficulty may be overcome in part by a careful analysis of the actual objectives of the computer center rather than the biased viewpoint of each of the users. If the sole purpose of the computer system is to provide rapid turnaround for all users, then the performance index must be defined to reflect this goal. If the computer center exists to process a certain category of jobs regardless of the cost or delay to other users, then this goal must be reflected in the choice of the performance index.

The following section will present an illustrative example of a performance index (PI) of the type under discussion. It should be emphasized that this PI is an example only and is presented here to demonstrate the type of analysis required to formulate a PI. It will be used to rank the 24 runs described in this report as an example of its application. Although not illustrated in this example, the cost of each alternative should also be considered in some manner in the computation of the performance index.

The following system performance criteria were assumed for this example. The prime objective of the computer center is to provide a rapid job turnaround time for all Category 1 (and possibly Category 2) jobs and obtain a 24-hour job turnaround for the other jobs. A parallel objective is to process as many jobs as possible during the day shift and return the maximum number of jobs from Category 1 (and Category 2) to the users prior to 4:00 p.m. each day.

Because a decrease in the job turnaround for one category would generally result in an increased turnaround for some other category, a method must be devised to permit acceptable combinations of increases and decreases. If a 30-minute decrease in job turnaround for Category 1 is acceptable when accompanied by a simultaneous increase of 1 hour in the job turnaround of Category 2, then the weights should be chosen to reflect this balance of values.

Since we wish to use the PI to obtain a ranking of all of the systems, we will compare each alternative system with the base case obtained in Run 1 and define the various terms in the PI in such a manner that an improvement in a statistic for the alternative system is reflected by an increase in the PI. To effect the combination of the performance characteristics previously mentioned, a set of "trade-off" criteria which define equivalent improvements were established. Effectively, these "trade-off" values define the weights given to the various terms in the model. For our purposes, the following outcomes will be considered equivalent when comparing an alternative model with the base line case.

1. Each decrease in job turnaround time for Category 1 by 1 hour.
2. Each decrease in job turnaround time for Category 2 by 2 hours.
3. Each decrease in job turnaround time for Categories 3 to 7 of 10 hours.
4. Each increase of total day shift throughput of 10 jobs/day.
5. Each increase of day shift throughput for Category 1 of 10 jobs/day.
6. Each increase of day shift throughput for Category 2 of 10 jobs/day.
7. Each increase in throughput capability of 100 jobs/week. Throughput capability is a function of both throughput and idle time.

Also, each of the outcomes could also be described in their negative context. This would allow the PI to show that outcome (1) accompanied by the negative of outcome (2) would result in a net change of zero to the PI. Since the PI is to be used for ranking purposes, the PI for Run 1 was normalized to a value of 0.0 and the decision was made to let each of the equivalent outcomes result in a change in the PI by a value of 0.05. For this example, the following form was chosen for the performance index.

$$PI = W_1T_1 + W_2T_2 + W_3T + W_4S + W_5S_1 + W_6S_2 + W_7I$$

where each term is defined below.

The term T_1 measures variation in the adjusted turnaround time for Category 1. Since a 1-hour reduction in Category 1 job turnaround was defined to vary the performance index by 0.05, the term T_1 was defined as

$$T_1 = [525 - T_1(J)]$$

and

$$W_1 = 0.00083$$

where 525 is the turnaround time for Category 1 from Run 1 in minutes and $T_1(J)$ is the turnaround time for Category 1 from Run J.

Similarly

$$T_2 = [400 - T_2(J)]$$

$$W_2 = 0.00042$$

where $T_2(J)$ is the Category 2 turnaround time for Run J.

The variation in adjusted turnaround time for Categories 3 to 7 is measured by T, given by:

$$T = \sum_{i=3}^{7} \left[T_i(1) - T_i(J)\right]$$

where $T_i(1)$ is the adjusted job turnaround times for Category "i" in Run 1 and $T_i(J)$ is the same for Run J.

$W_3 = 0.000083$

The fourth term, S, measures variation of the day shift throughput rate for all categories and has the form:

$S = [B(J) - 85]$

$W_4 = 0.005$

where 85 is the day shift throughput for Run 1 and B(J) is the day shift throughput for Run J.

Again in a similar manner, the terms S_1 and S_2 measure the variation in the number of jobs in Categories 1 and 2 completed on the day submitted and are defined:

$S_1 = [C(J) - 57]$

$W_5 = 0.005$

$S_2 = [D(J) - 2]$

$W_6 = 0.005$

where 57 and 2 are the jobs in Categories 1 and 2 completed on the day submitted for Run 1, and C(J) and D(J) are the same statistic for Run J.

The final term in our performance index evaluates the variation in idle time. Because idle time represents time during which the computer could have been processing additional jobs had they been available, it appears reasonable to convert idle time into a term which would indicate the system's capability to have processed additional jobs, assuming the job characteristics were to all come from the same distributions. The throughput capacity TC(J) for Run J is computed as follows:

$$TC(J) = (\text{Jobs Completed}) \left(\frac{\text{Total Time}}{\text{Time Used}} \right)$$

$$TC(1) = (257)(5) \left(\frac{168}{168-36} \right)$$

$$TC(1) = 1640$$

Throughput capacity could be computed on a daily rather than weekly basis with no loss of generality. The final term is given by

$$I = TC(J) - 1640$$

and

$$W_7 = 0.00050$$

Table B-1 ranks the 24 runs according to the defined criterion.

TABLE B-1

JOB STATISTICS AND PERFORMANCE INDEX
FOR 24 SIMULATION RUNS

Run	No. Completed Day Shift	Completed Day Submitted Cat 1	Completed Day Submitted Cat 2	Adjusted Job Turnaround Times Cat 1	Cat 2	Cat 3	Cat 4	Cat 5	Cat 6	Cat 7	Idle Time, min.	Through-put Rate	PI[(a)]	Rank
1	85	57	2	525	400	427	504	824	3655	867	2183	257	-0.000	11
2	57	52	3	243	332	375	381	417	1093	417	3212	191	0.282	7
3	72	33	0	1112	450	474	615	662	4612	1184	1919	269	-0.798	23
4	84	44	1	885	432	373	423	749	3768	1188	2350	239	-0.447	20
5	75	46	1	911	405	384	440	431	4071	754	2377	243	-0.441	19
6	79	51	0	905	611	349	490	536	4339	992	2222	228	-0.599	22
7	82	39	2	689	422	419	452	839	4040	1043	2114	258	-0.297	17
8	79	46	0	606	493	416	428	573	3667	732	2257	259	-0.149	13
9	85	53	6	325	742	396	409	580	3672	1347	2094	259	0.009	10
10	70	44	0	602	365	397	376	504	3889	421	2191	260	-0.132	12
11	81	57	6	433	555	493	431	534	7367	671	1474	258	-0.321	18
12	88	61	0	567	377	414	418	502	4064	453	2075	258	0.027	9
13	72	43	0	672	349	399	378	472	3908	410	2215	259	-0.177	14
14	50	42	0	445	0	0	0	0	3876	0	2289	259	-0.194	15
15	101	86	3	380	362	416	444	540	2848	837	2449	263	0.514	6
16	64	42	0	823	435	398	442	421	4053	845	2146	248	-0.475	21
17	85	48	1	664	434	403	519	695	4055	1273	2097	260	-0.234	16
18	117	107	6	238	367	402	425	564	2497	501	2617	265	0.913	2
19	80	79	1	226	317	347	372	475	995	395	2897	263	0.772	4
20	65	21	0	1523	374	782	3003	5672	5759	5240	1428	203	-2.513	24
21	81	68	0	499	360	365	398	624	2955	409	2467	266	0.250	8
22	76	75	1	245	314	380	371	492	2399	349	3046	264	0.624	5
23	87	86	1	225	299	361	379	371	887	352	3243	262	0.913	2
24	100	93	0	341	323	358	367	316	2781	414	2754	362	1.025	1

(a) The weights used to calculate this set of performance indices (PI) were:
W(1) = 0.00083
W(2) = 0.00042
W(3) = 0.00008
W(4) = 0.00500
W(5) = 0.00500
W(6) = 0.00500
W(7) = 0.00050

REFERENCES

1. J. E. Hawthorne. *A Simulation Study of a Proposed Computer Installation.* M. S. Thesis, University of Missouri, Oak Ridge National Laboratory, Oak Ridge, Tennessee (1966) (K-DP-1599).

2. T. H. Naylor, J. L. Balintfy, D. S. Burdick, and K. Chu. *Computer Simulation Techniques.* John Wiley and Sons, New York (1966) p 3.

3. A. L. Pugh. *DYNAMO User's Manual.* The MIT Press, Cambridge, Mass. (1963).

4. *General Purpose Simulation System/360 User's Manual.* International Business Machines Corporation, White Plains, N. Y. (1967) (H20-0326-0).

5. H. M. Markowitz, B. Hausner, and H. W. Karr. *SIM-SCRIPT, A Simulation Programming Language.* Prentice Hall, Inc., Englewood Cliffs, New Jersey (1963).

6. *SIMPAC User's Manual.* Systems Development Corporation, Santa Monica, California (1962) (TM 602/000/00).

7. G. K. Hutchinson. "A Computer Center Simulation Project." *Communications of the ACM 8,* No. 9, 559-568 (1965).

8. J. H. Katz. "An Experimental Model of System/360." *Communications of the ACM 10,* No. 9, 694-702 (1965).

9. N. R. Nielsen. "The Simulation of Time Sharing Systems." *Communications of the ACM 10,* No. 7, 397-412 (1967).