RESIDENT FELLOWSHIP PROGRAM
IN
COMPUTER SCIENCE

FINAL REPORT
NASA GRANT NO. NGR 11-002-158
JUNE 12, 1972
TO
SEPTEMBER 12, 1974

BY
MOIEZ A. TAPIA

PREPARED FOR
NATIONAL AERONAUTICS AND SPACE ADMINISTRATION
LANGLEY RESEARCH CENTER
HAMPTON, VIRGINIA

ATTENTION:
MR. RICHARD MORRIS, NASA TECHNICAL MONITOR AND PRECEPTOR
MR. WILLIAM F. GIANNINI, GRANT ADMINISTRATOR

School of Electrical Engineering
Georgia Institute of Technology
Atlanta, Georgia 30332
TABLE OF CONTENTS

I. General Assignment

II. Specific Assignments
   (a) Study of Disc Performance
   (b) Study of Magnetic Tape Performance
   (c) Study of Application Program Efficiency
   (d) Study of CPU Performance
   (e) Simulation Model for Computer Complex
   (f) Computer Networking
   (g) Ph.D. Thesis Review

III. Other Activities

IV. Conclusion

V. Suggestions for Extension of Work

VI. Acknowledgements

References
RESIDENT FELLOW PROGRAM

IN

COMPUTER SCIENCE

MOIEZ A. TAPIA

I. General Assignment

My principal assignment in the Analysis and Computation Division of the Langley Research Center was the evaluation of the performance of various subsystems in the Computer Complex using Statistical Data Collection System (SDCS). The system uses a large number of high-speed probes, integrated-circuit logic and storage media for capturing and storing desired events.

In collaboration with Richard Staib and Edward Phillips I designed systems for monitoring disc controller, tape controller, program count register and central processor unit for studying the performance of disc storage system, tape storage system, central processor unit and efficiency of application programs respectively. These systems are described in earlier reports (1, 2, 5).

Among other activities, I intensively reviewed and consulted in the preparation of a Ph.D. thesis by Jerry Tucker, who received his Ph.D. degree in June 1974. I reviewed efforts, in the Center, to develop a Simulation Model for the Computer Complex. I was, also, involved in the study of a Computer Network for the Complex. All these activities will be described in the next section.

II. Specific Assignments

(a) Study of Disc Performance

The purposes of studying disc performance as well as the details of the plug-in modules for monitoring the disc controller have been described in earlier reports (1, 2, 5).

The Statistical Data Collection System was connected to discs Z₀, Z₃, and Z₄ and data collected during the week Feb. 25, 1974 through March 2, 1974. Later it was connected to discs B₀ and B₄ and data collected during the week March 25, 1974 through March 30, 1974.
The statistical analysis of various parameters measured is given below:

(1) The total activity time is the time spent in accessing, repositioning, reading and writing. Useful activity time is the time spent in reading and writing records. The average total activity on \( Z_0, Z_3, Z_4, B_0 \) and \( B_4 \) was found to be 35.6, 2.3, 13.0, 40.7, and 28.8% respectively.

(2) The percentage of records written with length less than or equal to seven sectors is around 80% in all the discs. In case of \( Z_3 \) more than 80% of the records written were of one-sector length. This indicates a poor definition of currently used block size. Consideration should be given to redefining block size so that most of the records written would be of 8-sector length. This would reduce time spent in accessing and/or repositioning thus improving overall data transfer rate.

The percentage of records read, with length of seven or less sectors was in the range of 60 to 80.

(3) The maximum number of sectors read and written in a fifteen-minute interval was 191,854 (\( B_0 \)) and 66,198 (\( Z_4 \)) respectively.

(4) The mean repositioning times for \( Z_0, Z_4, B_0 \) and \( B_4 \) were 51.55, 58.37, 35.45, 42.50 and 39.20 milliseconds respectively. The repositioning time had nearly uniform distribution over the range 17-92 millisec. For \( Z_3 \) the range was 62-91 millisec.

(5) 90% of accesses took less than 2 milliseconds. Some accesses took 50 milliseconds.

(6) The number of parity errors per billion bits transmitted was 64, 867 and 81 for \( Z_0, Z_3, \) and \( Z_4 \) respectively.

The percentage of time spent in useful activity, on the average, was found to be 3.42, 0.14, 2.57, 5.06 and 3.93 for \( Z_0, Z_3, Z_4, B_0 \) and \( B_4 \) respectively.

This indicates a very low activity on job disc \( Z_3 \) and hence a very inefficient utilization of the job disc. Consideration should be given to writing job copies on a tape or another disc.

Significant unbalance exists between activities on \( Z_0 \) and \( Z_4 \), and some unbalance between \( B_0 \) and \( B_4 \).
The maximum total activity during a 15-minute interval was found to be 85.9, 9.7, 61.0, 93.6, and 92.2 on discs Z₀, Z₃, Z₄, B₀ and B₄ respectively.

(b) **Study of Magnetic Tape Performance**

The study of performance of the magnetic tape storage system was undertaken in order to obtain information needed for determining the requirements and specifications for a Mass Storage System for the Complex. The plug-in modules designed for the purpose have been described in an earlier report (1). Ed Phillips and I de-bugged and checked out the system with the help of John Gambill. The SDCS was connected to the tape controller for nearly 6 consecutive days during July 1974. Partial analysis of the data gathered has been recently obtained. The following statistics are based on analysis of data gathered from 8:20 a.m. to 8:20 p.m. on Monday August 5, 1974.

Mean Record Length = 236 CM words

# Records Written (during the twelve hour period) = 90,923

# Records Read = 286,580

# Records Read or Written = 370,449

# Back Spaces = 4729

# Parity Errors = 8621

# Files = 347

# Records/File = 1068

Mean File Size = 251,906 CM words

# Tape Mounts = 346

# Rewind/Unloads = 356

# Maximum Transports Active During 30-second Interval = 9

# Rewinds = 545
These results were compared with the information on dayfile. It was observed that as far as the monitored tapes, other than the calcomp ones are concerned, there is close agreement between the data gathered by the SDCS and those on the dayfile. Ed Phillips has looked into this problem recently. The calcomp tapes at the end of a file backspaces. This is not recorded on the dayfile, but is observed by the SDCS and recorded.

Ms. Barbara Polak has been processing the data gathered on tapes DEP 10 through DEP 14. Detailed analysis will, then, be obtained.

(c) **Study of Application Program Efficiency**

The superiority of a hardware monitor over a software monitor has been discussed earlier (2). A hardware system monitoring an application program can be used for debugging the program. Since it gives the distribution of frequency of execution of various instructions, it can be used to compare a pair of high-level languages, in terms of various arithmetic and logical operations they involve for a given computational job.

The details of the plug-in modules I designed are given in earlier reports (1, 2, 5). Ed Phillips and I debugged and checked out the modules wired by John Gambill. The SDCS with the plug-in modules in it, was connected to the Computer A in August 1974, and a test program written by Joseph Drozdowski was monitored. The analysis of data gathered has been obtained recently (September 1974). The system did record the frequency of various instructions executed. However, the system erroneously recorded a very frequent reference to the address 0. This fault, probably, is due to the inappropriate delay in the system between occurrence of the program clear pulse and the loading of the data from P register into the shift register in the system. Ed Phillips is considering this possibility and attempting to remove the fault.

(d) **Study of CPU Performance**

In order to study the nature of computing done at the Center in terms of frequency of use of various arithmetic units, conflicts in their accesses, number of instructions executed per second, etc. I designed plug-in modules for monitoring a CDC-6000 central processor unit.
The plug-in modules designed for the purpose can be used for determining efficiency of different higher level languages in terms of frequency of distinct arithmetic operations that they involve for a given computing job. For example, a comparison between the efficiencies of Fortran-Run (currently in use at the Center) and Fortran-Extended (to be used in the near future) languages for a given job in terms of distribution of various arithmetic operations, can be made.

The modules will be built in the near future and the SDCS with these plug-in modules in it will be connected to a CDC-6000 central processor unit.

(e) Simulation Model for the Computer Complex

The need for and the philosophy of developing a simulation model for the Computer Complex have been described in an earlier report (2). George Canovos, Joseph Drozdowski and I have been reviewing the efforts underway for developing simulation model for the Computer Complex.

Robert Deacle has been debugging the model left to us by Sherman at the end of Summer 1973. The model at present takes nearly 0.93 unit time to run jobs which took 1 unit of time under real ICOPS system. Also, to run jobs that took 24 hours under real ICOPS system, the model takes nearly 16 minutes of CPU time and 160K words of CM.

The inaccuracies in the model stem from lack of certain information available from dayfile as well as discrepancies of the model itself. The dayfile does not tell us when a job goes into pre-mount and when user subroutine EVICT releases a tape. The model does not use information regarding data cell down time and WAITGO control card. Also, the model is not handling big jobs running at midnight, the way they should be handled.

The accuracy of the model will be improved in the next couple of weeks by refining the model as well as modifying the dayfile subroutines.

After "reasonable" accuracy has been obtained, the effect of changes in algorithms for assigning printers on the throughput will be determined using the model. The effect of changes in other algorithms and/or resources will be determined later.

Simulation models and hardware monitors are complementary and indispensable for understanding behavior and improving the efficiency of the computer complex.
(f) **Computer Networking**

Networking of the computers at the Center has been underway for quite some time. The advantages of a computer network, "local" or nationwide, have been outlined in an earlier report (5) which, also, discusses different configurations and/or protocols for such a network.

A special study group comprising Donald Booth, Richard Hofler, et.al. examined several configurations for computer networking during the summer of 1974.

Due to more intensive involvement, on my part, in tape monitoring and development of an application program efficiency monitor, I could not exchange views on the various approaches the group was considering.

(g) **Ph.D. Thesis Review**

I intensively reviewed and guided Ph.D. dissertation research as well as writing of the dissertation of Jerry Tucker, in the area of Transitional Calculus using Boolean Differences.

He established conditions for integrating Boolean differentials and obtained algorithms for integrating compatible differentials and exact differentials. He gave an algorithm for piecewise-integrating any differential.

The results are a significant contribution to the design of sequential systems using edge-sensitive flip-flops and to fault analysis.

He successfully defended his thesis and received the Ph.D. degree in June 1974.

**III. Other Activities**

Numerous professional activities such as presenting papers, publishing papers, attending seminars and conferences have been described in earlier reports (1, 2, 5). The activities not described earlier will be described here.

The research pertaining to necessary and sufficient conditions for the solution of a system of Boolean equations having unknown functions of a specified number of variables is proposed to be reported in a NASA Technical Note to be written by Jerry Tucker and I. The proposed Note will be submitted for consideration in the near future.
I attended a five-day Seminar on "Computer Performance Evaluation" held at the College of William and Mary in July 1974.

I spent several days looking into the various parameters of operating systems of a CDC-6000 computer, that could be monitored and analysed for performance evaluation and improvement. This is, indeed, a very useful and significant activity which could not be pursued to completion due to limitations on time.

IV. Conclusion

The design of plug-in modules for monitoring disc storage system and those for monitoring tape storage system has been successfully completed. Its application has given useful data whose analysis has given us insight into the status of performance of these storage systems and has indicated ways of improving the performance of the systems.

The design of modules for monitoring application program efficiency is almost completed, with some minor faults in the systems to be taken care of.

The drawings for the design of modules for monitoring central processor unit of a CDC-6000 computer have been completed.

Important aspects of a local computer network have been outlined.

The Ph.D. research was successfully completed which led to graduation of Jerry Tucker in June 1974.

The necessary and sufficient conditions for the existence of solution of a system of Boolean equations have been established, and a simple algorithm for obtaining the solutions has been obtained.

Publications accomplished include a paper in IEEE Transactions on Circuit and System Theory, a review of Wai-Kai Cheu's paper on characterization of complete directed trees and two-trees, another review of text-book "Applied Graph Theory" by Marshall, three papers presented at different conferences and a NASA Techbrief (5).

Research on solutions of Boolean equations will be submitted in a Technical Note for consideration in the near future.
The opportunity to study the hardware as well as the software systems in the Computer Complex, the experience in sophisticated design of systems to monitor some of the computer subsystems and their interfaces using modern digital design techniques, interaction with a number of experienced engineers and computer scientists, involvement in theoretical research related to digital system design - all this has tremendously contributed towards making my Residency a very worthwhile and rewarding experience for me. The extent to which this will affect the relevance and practicality of my teaching and research is immeasurable.

V. Suggestions for Extension of Work

In a dynamic environment such as exists at the Center computer performance evaluation would have to be an ever-continuing activity. It is conceivable that additions to the complex and/or changes in the equipment or configurations will take place off and on. The effect of these changes on the performance of the computer systems will have to be determined by using hardware and/or software monitors and/or simulation models.

The plug-in modules designed for disc and tape storage systems may be used to observe the effect on performance, of any changes in different resource-allocation routines and/or various algorithms controlling processes in systems. In fact, a "best" algorithm (e.g. track assignment algorithm) for a given operation in these systems can be arrived at by varying these algorithms and observing their effects on the performance.

The plug-in modules for tape storage can be modified so that the SDCS will not record "EOF" when the tape is backspacing or rewinding.

The performance of an operating system has a tremendous effect on the overall performance of the computer using it. Consideration should be given to designing modules which can monitor various parameters of the operating system of the complex, and to determine their effect on the performance of the overall system.

Transitional Calculus as proposed by Jerry Tucker has a very good potential to make significant contribution to analysis as well as design of digital systems. It is, however, in its infantile stage. Work in this area should be continued. Necessary and sufficient conditions for a differential expression to possess an exact integral need to be obtained.

The Transition Calculus as reported in (6) assumes one variable change at a time. It should be generalized to allow multiple simultaneous changes in variables. Techniques for handling incompletely specified circuits should be developed.
Possibility of extending the applicability of Transition Calculus to circuits with input constraints should be investigated.

VI. Acknowledgements

The author wishes to express his sincere gratitude toward the NASA Langley Research Center for the award as well as the extensions of the grant which made it possible for him to pursue Resident Fellowship. Sincere gratitude is also expressed to the ASEE-Ford Foundation Resident Fellow Program which provided this invaluable experience for the author.

Special mention must be of his sincere gratitude towards Messrs. Andrews, Gribble, Eskew, Morris and Young who provided encouragement, support and guidance for the activities pursued while the author was at the Center.

The author is, also, very grateful to Messrs. Staib, Tucker, Phillips, Cuddihy, Booth Drozdowski, Hofler, Canovos and Gambill for their willingness to help and cooperate, delightful discussions and participation in the author's various activities.

Thanks are due to Ms. Judy R. Neil who willingly typed many manuscripts.

The author is, also, very grateful to Dr. D. T. Paris, Director of School of Electrical Engineering, Georgia Institute of Technology who recommended the author for the Resident Fellowship and gave it the highest priority among his academic activities.

The author would also like to thank Dean Walter Carlson who nominated him for the Fellowship.
References


