MTP-28-A

# A PRELIMINARY COST/EFFECTIVENESS COMPARISON OF MESSAGE AND CIRCUIT SWITCHING IN MILITARY COMMUNICATION NETWORKS SERVICING RECORD TRAFFIC

Stephen A. Sobel

The MITRE Corporation
Washington D.C. Office



JUNE 1966

MTP-28-A

# A PRELIMINARY COST/EFFECTIVENESS COMPARISON OF MESSAGE AND CIRCUIT SWITCHING IN MILITARY COMMUNICATION NETWORKS SERVICING RECORD TRAFFIC

by

Stephen A. Sobel

MITRE(DCA)Division

June 1966

THE MITRE

Washington, D. C. Office

The work reported herein was sponsored by the Defense Communications Agency under Contract AF19(628)-5165.

## ABSTRACT

An analysis is contained herein that compares circuit with message switching in military communication networks that carry record traffic. Also examined are the effects of system parameters on the attainment and measurement of statistical convergence when Monte Carlo simulation is used to study delays caused by traffic congestion in communication networks.

## ACKNOWLEDGMENT

The author is indebted to W. A. McCloskey of the Information Processing Systems Department (D-73) of The MITRE Corporation for designing and implementing the simulation computer program that was developed in support of this investigation.

## FOREWORD

Recent years have witnessed a marked growth in the use of automatic data processing techniques in military command and control systems. This has been accompanied by a significant rise in the volume of digital information that flows between physically dispersed centers of command, surveillance, and control which are jointly involved in the accomplishment of assigned missions.

As communication facilities have been expanded to meet this demand, it has become increasingly more important to ensure that the communication capability provided adequately satisfies all operational needs and does so at a minimum cost. Concurrently, it is essential to develop modern tools to aid in the evaluation of proposed communication system designs.

This report constitutes an initial effort to compare the economic and operational implications of using either fully automated message or circuit switching elements in military communication networks that service digital traffic. The results are presented in two parts.

Part I provides general background material, identifies and briefly examines the relevant issues, proposes a framework for more intensive study, and describes a simulation computer program that was developed to aid in the analysis.

Part II presents the results of exercising the computer model and draws certain conclusions. Requisites for obtaining statistically valid results from Monte Carlo simulations in situations that are similar to those encountered in this investigation are discussed at some length.

## TABLE OF CONTENTS

TABLE OF CONTENTS (Cont'd.)

## TABLE OF CONTENTS (Concl'd.)

## LIST OF ILLUSTRATIONS

# 1.0 INTRODUCTION

## 1.1 General

This document constitutes the first in a series which compares the cost of employing circuit switching with high speed, automated message (store-and-forward) switching in communication networks that carry digital data. The overall study is directed at examining the relative economic, operational and technical feasibility of using these two switching techniques (or a combination of both) to satisfy the communication requirements of military users in a broad spectrum of operational environments.

The analysis contained in Part I, while primarily qualitative in nature, provides a framework for more intensive study in specific areas. In addition, a description is offered of a recently developed simulation computer program (called STARCON) which will assist in determining the relative size of the transmission plant that is required by the two types of networks so that they satisfy equivalent performance criteria.

## 1.2 Background

Digital communication networks with switching centers can be divided into two broad classes. These will be referred to as circuit switched and store-and-forward (message switching) networks. The circuit switched network requires that the entire connection be established from the originating subscriber to the destination subscriber before traffic will flow. If all of the channels in any of tandem trunks needed to complete the connection are in use when a call is placed or if a blocking condition exists in any of the switching centers along the route, a busy signal will be issued and the caller will not be able to release his traffic at that time.[1] In such circumstances, the originating

[1] Of course, this is not strictly true when a preemption capability exists.

subscriber must initiate subsequent connection requests, called "retrials," until the destination subscriber is reached. With message switching, however, traffic may flow through successive links of the route even when the entire end-to-end connection cannot be simultaneously effected. Whenever an impasse is encountered, the traffic is stored at the last switching center reached and subsequently forwarded on as soon as transmission facilities become available.[2]

The store-and-forward technique lends itself to the transmission of digital data and, in general, achieves a higher utilization of transmission facilities than circuit switching when both types of networks satisfy the same delay requirements. This technique thus appears to offer an efficient approach in the design of military data communications systems.

The costs of certain components in store-and-forward systems, however, tend to be significantly higher than their counterparts in circuit switching networks. Furthermore, some components which are common to both systems are used in greater quantity (on a link-by-link rather than end-to-end basis) in the store-and-forward case.

Thus, it does not follow that the store-and-forward technique will necessarily result in the more economical configuration, given that stated performance objectives must be met. It is the purpose of the present investigation to examine those circumstances under which the store-and-forward approach would tend to be more economical and vice versa.

---

[1] Message switching systems, currently in existence, generally require storage of the entire message at each switching center before releasing it to an outgoing trunk.

Some systems engineers in industry contend that the choice between message and circuit switching in military networks can be made simply from a cursory examination of functional requirements, although it is not clear to the author how this should be done. Qualitative statements have been made in the past to the effect that circuit switching is a more economical approach when some of the following conditions are true:

1) transmission facilities are in abundant supply and inexpensive to use,

2) the cost ratio of the switching facilities to the transmission plant is large,

3) messages are long,

4) the number of subscribers per switching center is large,

5) frequent long delays cannot be tolerated,

6) most messages do not have multiple addresses consisting of different combinations of addressees,

7) messages tend to be transmitted at regular time intervals and do not vary excessively in length,

8) messages do not have to traverse too many switching centers from origin to destination,

9) the number of channels per trunk is large,

10) an extensive amount of speed or code conversion is not required, and/or

11) a multiplicity of efficient, alternative paths exists.

A quantitative criterion, however, to select between these switching concepts has apparently never been developed.

It is the author's opinion that as it becomes necessary to supplement existing communication system facilities to cope with an anticipated growth in the volume of record traffic, or in the process of taking appropriate steps to enhance the survivability of the existing physical plant, or simply when it becomes necessary to determine whether a particular set of subscriber communication requirements should be satisfied by circuit switching or message switching elements (or a combination of both), more definitive and defensible guidelines should be available to assist in making this decision.

### 1.3 Limitations of the Current Analysis

Military communication switching networks exist in many forms and configurations and are designed to satisfy many different kinds of performance goals that depend on the operational situation. This initial analysis, however, does not attempt to make an exhaustive comparison of circuit switched to store-and-forward networks in all possible situations. Rather, to reduce the scope of the initial study to manageable proportions, a few basic parameters, susceptible to quantitative analysis, will be isolated and examined. In particular, in operational situations defined by parameters depicted in Section 2.0, the ratio of the transmission capacity required by each switching technique to carry an equal volume of traffic, and still achieve the same value for a given delay statistic, will be determined. It is hoped that this relationship, augmented by a quantitative economic analysis of system elements, can lead to some general guidelines that will aid in choosing between the two switching techniques.

## 1.4 Overview of Relative Component Costs in the Two Types of Switched Networks

Although a comprehensive analysis of the relative costs of fully automated message and circuit switched systems is beyond the scope of the present study, the reader would naturally benefit from a brief qualitative discussion comparing the cost of typical hardware configurations used by the two switching techniques. This is, therefore, offered below.

Costs associated with a communication network are derived from three principal sources as represented in Figure 1-1. These are:

1) those related to the transmission plant,[1]

2) those arising from functions performed at the switching centers, and

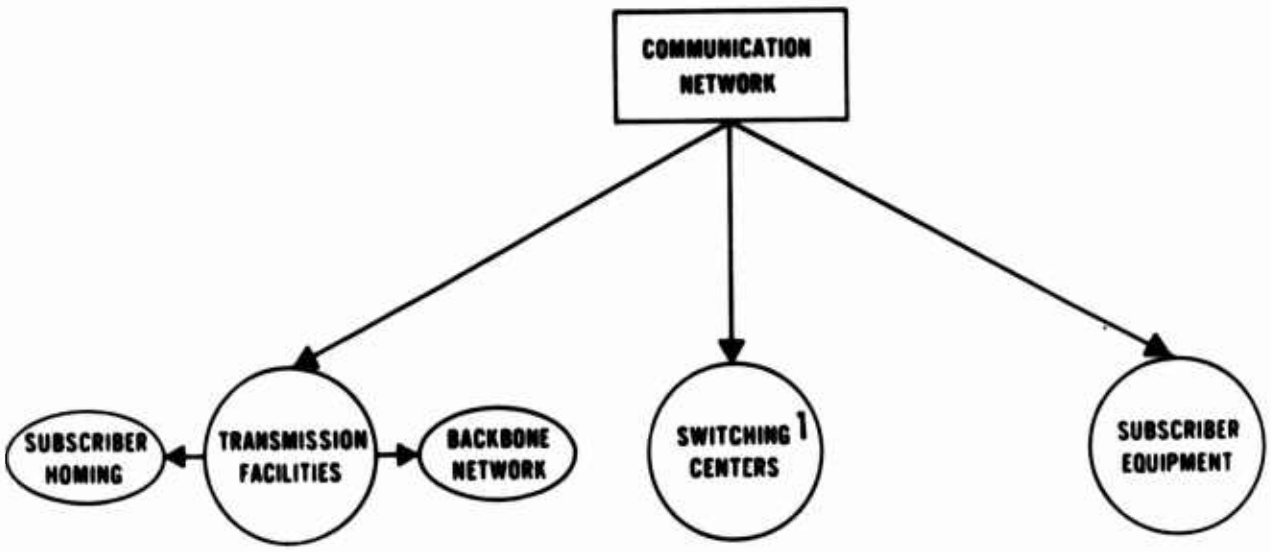
3) those which result from equipment located at the subscriber terminals.

The primary emphasis of the quantitative portion of the investigation to follow is directed at comparing the magnitude of the trunking costs in the two types of switching networks. Nevertheless, an order of magnitude comparison of other system costs may be of value at this time.

Switching centers in circuit switching networks lack provision for message storage, which therefore is necessarily performed at subscriber

---

[1] Transmission facilities can be further divided into two classes: lines joining subscribers with switching centers (commonly referred to as subscriber homing) and trunks linking switching centers themselves and comprising the so-called "backbone" network. It will be assumed that subscriber lines will exist in approximately the same quantity with both message and circuit switching, and hence attention herein is focused on the relative trunking requirements.




[1] A list of some of the parameters affecting the cost of individual switching centers is as follows:

Circuit Switching - the number of trunks and subscriber line terminations (and whether two or four wire), the number of links (i.e., percentage blocking), thru-put bandwidth, switching speed, amount of common control equipment, degree of alternate routing capability, reliability (MTBF), power requirements, physical hardness, and such other capabilities as conferencing, the number of preemption levels, abbreviated dialing, alarms and automatic diagnostics.

Message Switching - the number of trunks and subscriber lines accommodated, the size and nature of the central storage, the planned average utilization of the data processing capability, the maximum permissible message length, constraints on power consumption and physical size, the hardness level, and the capability to perform such ancillary functions as handling multiple precedence traffic, performing format, code and speed conversion, providing security, delivering multiple addressed messages, etc..

Figure 1-1. Simplified Breakdown of System Elements in a Switched Communication Network

terminals. Equipment performing this function at subscriber terminals can have a transfer rate equivalent to the transmission rate and a relatively slow access time. In some situations, the same storage equipment which is part of the data processing installation at the subscriber's facility may be used for this purpose in conjunction with auxiliary buffering devices and control units.

The actual switch connections with circuit switching are generally effected by relatively slow speed electromechanical devices and are controlled by common equipments which are typically electromechanical or more recently solid state. The common control facilities are only in operation while connections are being established. Hence, the amount of this equipment which is required depends on the frequency of message initiations rather than on traffic volume.

With circuit switching, many functions can be performed on an end-to-end basis rather than link-by-link. These functions include error control, message encryption, and modulation. Equipment to perform these operations need only be located at subscriber terminals, and only a single pair of units for each function is involved in the complete transmission of any message from source to ultimate destination.

In general, store-and-forward networks store and retrieve each message many times as it passes from the sending to the receiving subscriber. In a typical configuration, a message, upon entering any switching center along its route, is inserted in a high-speed (orders of magnitude faster than the transmission rate) random-access input register, commonly magnetic core. If channels in the required outgoing trunk are available, the message will be transferred to an appropriate output register and then offered to the channel. If, on the other hand, no outgoing channels are free, the message is transferred from the input register to a central storage facility which may take the form of

a magnetic disc or drum.[1] Logical and administrative operations are performed in practice by one or more medium size general-purpose electronic computers, and the switching funotion is accomplished by means of transferring messages through the internal computer memory.

Key stream generators are commonly used in communication systems as they provide for a high level of message security.[2] Synchronization requirements, however, preclude operating these devices end-to-end in a store-and-forward network.[3]

Error control is frequently performed by providing for the retransmission of message blocks in which errors are detected. This technique can only be conveniently executed, in the case of store-and-forward systems, on a link-by-link basis.

In order that the data be processed at each switching center traversed, data modems are required to perform the waveform conversion. This equipment,

---

[1] If these become saturated, the data is generally then transferred to tape. In some systems all messages pass through secondary storage.

[2] Also, when used on a link-by-link basis, they can provide for call flow security, i.e., whether or not traffic is flowing cannot be monitored by the enemy. On the other hand, when used on a link-by-link basis, cryptographic security within the switching centers is reduced.

[3] Conceivably, a message switch could be designed and operated with a cut-through capability which would enable a key stream generator to be used on an end-to-end basis. Nevertheless, doing this is tantamount to simulating a lower cost circuit switch with a message switch since the advantages of store and forward, in this mode of operation, are completely nullified.

therefore, must be provided at each trunk channel and subscriber line termination in all switching centers. [1]

It is evident from the foregoing that there are many sources which tend to cause the costs of switching center equipment associated with store-and-forward systems to be greater than with circuit switching systems. Memory devices, in the former case, must operate at much higher speeds and be more versatile.[2] More sophisticated equipment is used to perform the switching function. Equipment to perform the error control function, achieve security, and provide for waveform conversion is typically present at all switching centers in addition to subscriber terminals.

While the cost of switching facilities for store-and-forward systems tends to be greater than for circuit switching systems, the higher trunk occupancy, $\rho$, permitted by message switching realizes savings in transmission facilities costs. Since total network costs are the sum of switching, terminal, and transmission facility costs, of concern is whether the savings in transmission (and perhaps subscriber equipment) costs afforded by the store-and-forward technique more than offsets the increased cost of switching facilities.

---

[1] As is also true with error control and security equipment; naturally with wideband transmission facilities and multiplexing, the statement has to be appropriately modified.

[2] Of course, as previously noted with circuit switching, delayed messages are stored at subscriber terminals.

## 2.0 SCOPE OF CURRENT INVESTIGATION

As stated earlier, the initial quantitative investigation is primarily concerned with determining the relative transmission capacity that is required by the two switching concepts to carry the same traffic volume and satisfy the same delay criterion.[1] This comparison is performed under different sets of conditions which are defined by the following parameters:

1) the number of tandem trunks traversed by each message,
2) steady-state channel loading,
3) the number of channels per trunk group (i.e., total trunk capacity),
4) holding time (i.e., message length) distribution,
5) transient loading (size and duration),
6) delay criterion (i.e., average or specified percentile delay),
7) permissible delay (stated in average number of holding times),
8) forced hold[2] and speed-up[3] with message switching, and
9) intra-switching center delay and retrial interval with circuit switching.

---

[1] With circuit switching, it is assumed that blocked calls are placed again at fixed time intervals until served. A constant retrial interval was adopted as it is felt that terminal equipment could be used to perform the function in this fashion automatically.

[2] Forced hold refers to the operational doctrine that often is employed in store-and-forward systems wherein messages arriving at a switching center are completely stored before an initial attempt is made to transmit them out of the center over an appropriate trunk. Some reasons for this practice are speed-up, error correction (although this merely requires the complete receipt of entire blocks), and general processing convenience.

[3] Speed-up refers to the operational doctrine with forced hold store-and-forward networks whereby messages are transmitted through backbone trunks at rates greater than which they enter and leave the system due to subscriber terminal equipment limitations.

Section 3.0 describes the computer program that was used to investigate the above parameters. The manner in which these parameters are specified in the program input and the format of the resulting output are provided therein. Other performance considerations affecting the choice between circuit and message switching, but not treated by the current computer model, are briefly discussed in Section 4.0.

## 3.0 DESCRIPTION OF STARCON[1] MODEL

### 3.1 General

The STARCON computer program has been developed to provide a vehicle for comparing the effects of certain variables (specified in Section 2.0) on the relative transmission capacity required by circuit and message switched communication networks. This comparison is performed by carrying out a Monte Carlo simulation in each situation. The program is written in FORTRAN IV,[2] contains in the order of 600 statements, and is presently operating on a CDC 1604 machine with a 32K word core memory. The operating speed varies considerably with the choice of input parameter values; typical values for the processing rate are presented in Section 3.0 of Part II.

Appendix A-I discusses the computer program logical design, procedures for specifying and inserting the input data. It also describes the content and format of the computer printout.

### 3.2 Configuration

The program treats the single node (switching center) as well as the multiple node case. In the latter situation, a "ring" configuration is utilized. This is illustrated in Figure 3-1. All messages flow counterclockwise and traverse the same number of tandem trunks. The number of nodes in the ring and the number of tandem trunks traversed by each message are specified by input parameters.

---

[1] Switching Technique Analysis for Record Communication Networks.

[2] With the exception of a short subroutine (written in CODAP) which reads the 1604 clock. The FORTRAN IV language used on the CDC 1604 is actually designated as FORTRAN-63.

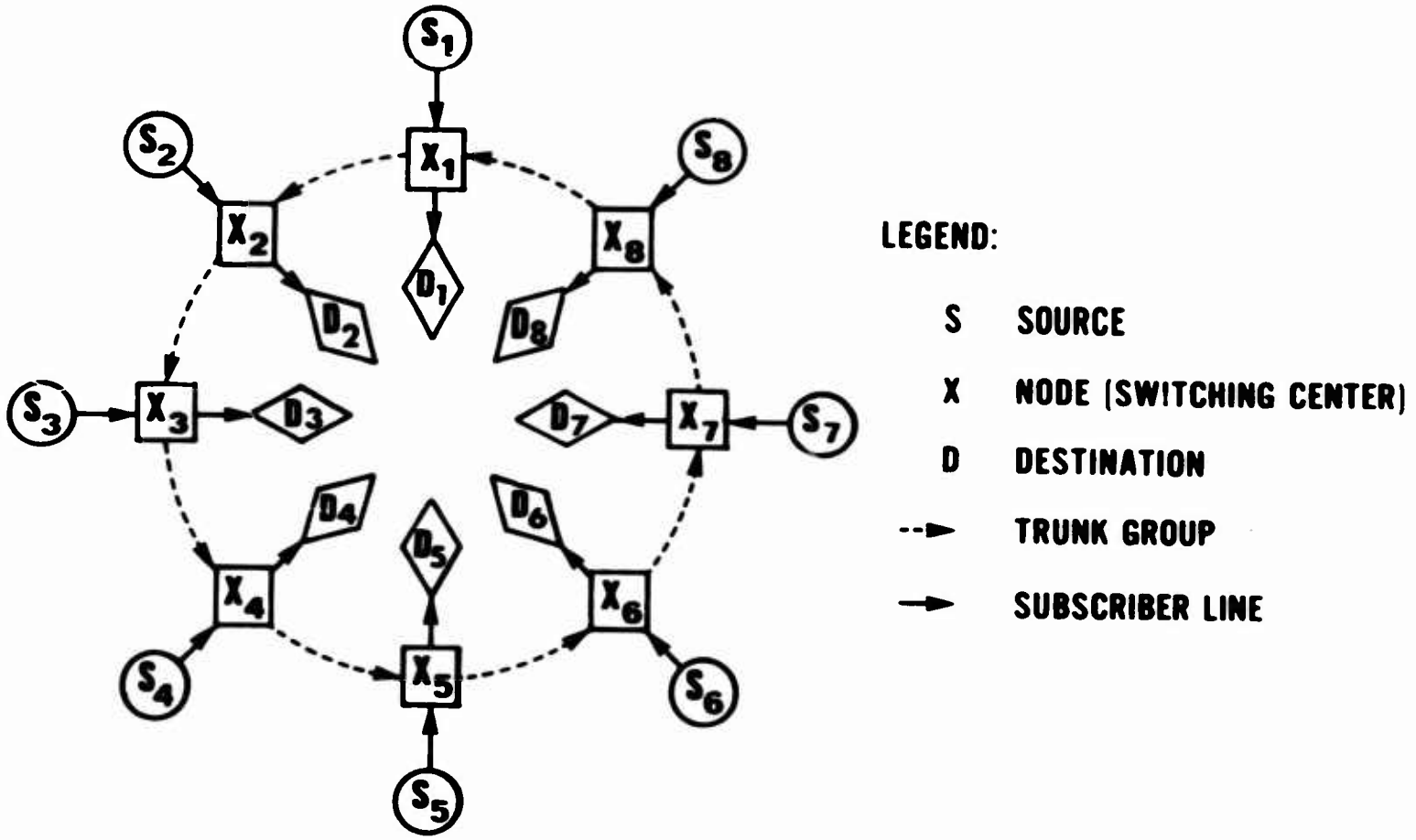



Figure 3-1. Ring Configuration: Multiple Node Cases

The ring configuration is employed here as a convenient means for monitoring the effects of tandem switching centers. The symmetry offered by this network not only reduces the programming burden, but allows samples of message delays experienced between different pairs of originating and receiving subscribers to be analyzed collectively since all pairs exhibit the same steady-state frequency distribution.

Subjecting a single type of configuration to analysis will naturally limit the generality of the result. In the ring configuration, messages arriving at a switching center come from only two sources: local subscribers (treated as a single Poisson source) and one other switching center. In the store-and-forward case, this creates only one of many possible situations with regard to statistical dependence between message interarrival times and message lengths. Configurations, (1) which have messages entering a switching center from many other switching centers (each of which contributes only a small portion of the total load), (2) which provide for the two-way flow of traffic, (3) in which different messages traverse different numbers of switching centers, and (4) in which messages originating from a given source have different destinations, may exhibit different characteristics.

### 3.3 Input/Output Variables

So that the reader can become more familiar with the STARCON program's capabilities, a description of some of the input and output variables is presented below. A more detailed account of the design of the program is offered in Appendix A-I.

### 3.3.1 Input Variables

#### 3.3.1.1 Model Number

This parameter takes on four values, differentiating among the classes of switching networks that may be simulated. The reader is referred to Figure 3-2.

Model One: This model refers to a store-and-forward network in which messages can flow out of a switching center as soon as an outgoing channel becomes available. With this model, speed-up (see Paragraph 3.3.1.11) is not possible. This model is referenced in Appendix A-I as the Queueing-Forced Hold case.

Model Two: This refers to a store-and-forward network in which a message must be completely received at a switching center before it can be released to an outgoing trunk channel. This is referenced in Appendix A-I as the Queueing-Forced Hold case.

Model Three: This refers to a circuit switching network in which no delays are encountered at switching centers and retrials are executed at fixed time increments. This model is referenced in Appendix A as the Retrial-No Delay case.

Model Four: This refers to a circuit switching network in which connection delays are experienced at switching centers. During these delays, all trunks between the originator and the switching center involved are held idle. This model is referenced in Appendix A-I as the Retrial-Delay case.

#### 3.3.1.2 Number of Nodes

This parameter refers to the number of switching centers in the ring configuration. A maximum of thirty-two switching centers can be handled.

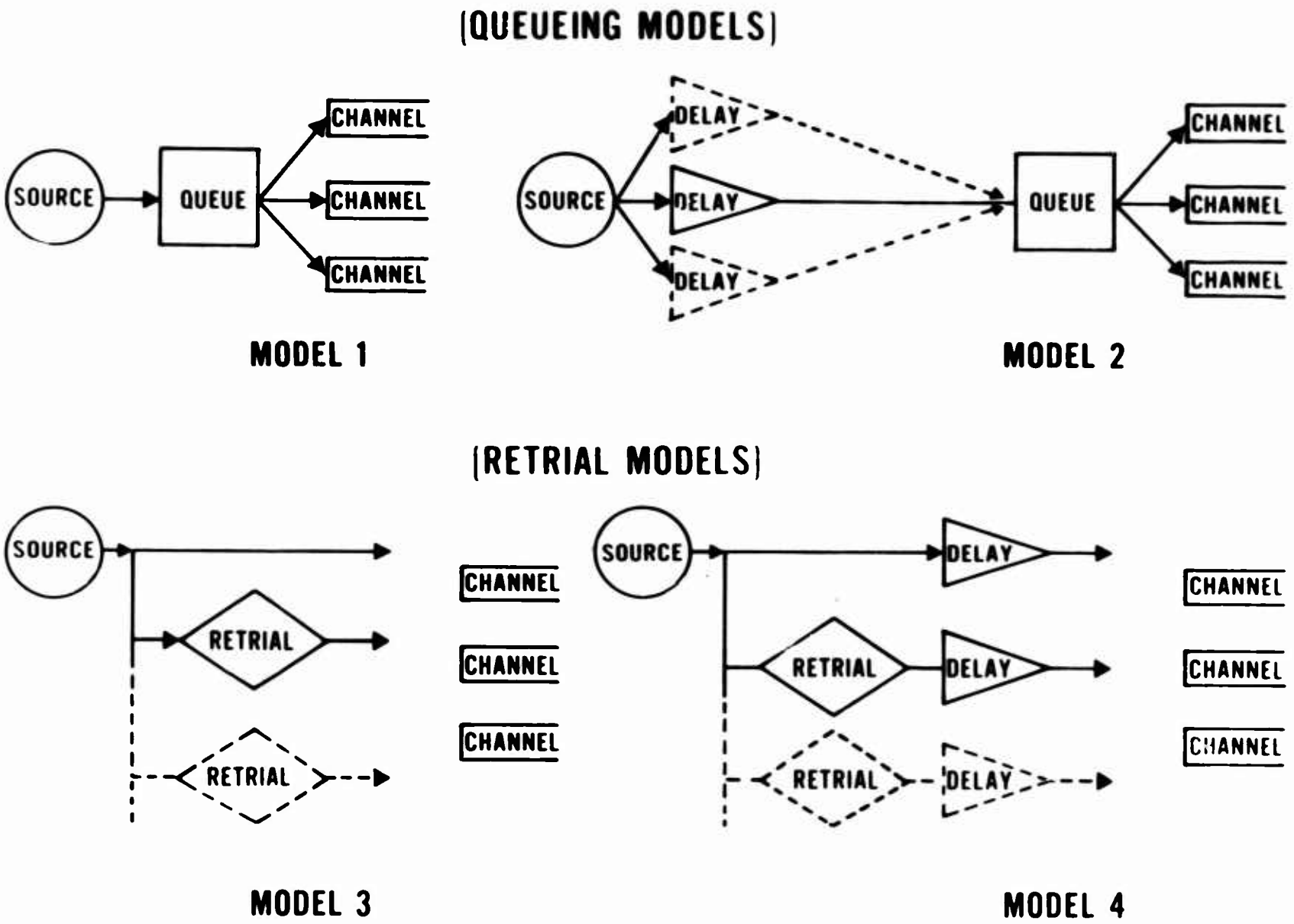



Figure 3-2. Models Simulated by the Computer Program: Single Node Cases

In the special case when one node is specified, the configuration investigated is composed of a single source feeding into a single switching center that in turn releases messages to a single out-going trunk group.

#### 3.3.1.3 Number of Channels

This variable specifies the number of channels in each trunk group.

#### 3.3.1.4 Number of Erlang Phases

This parameter specifies the number of phases in the Erlang statistical distribution characterizing the message length variability. When the number one is specified, the exponential distribution is used. Zero causes fixed length messages to be generated.

#### 3.3.1.5 Initial Random Number

This variable selects the starting value for the pseudo random number generator.[1] It permits the results of similar cases to be compared for different sequences of message lengths and message origination interarrival times.

#### 3.3.1.6 Fractile

This parameter permits one to determine that message delay such that the number of shorter delays comprises a stated fraction of the total number of all observed delays. This fraction, varying

---

[1] The pseudo random number generator used by STARCON is the RANF FORTRAN-63 library routine as designed for the CDC 1604-A computer. It uses a generation method similar to that discussed by Rotenberg [17], takes the form of

$$x_{i+1} = (2^{10} + 1)x_i + 101 (\text{modulo } 2^{36}),$$

and has a serial correlation that has been investigated by Coveyou [5].

between 0 and 1, is referred to as the fractile. For example, if the input fractile of 0.9 is specified and 10,000 delays are observed, the delay will be found (and its length indicated in the computer output) such that there are exactly 9,000 shorter delays in the total sample. The fractile, thus, is a sample statistic and as such has no direct influence on the simulation.

#### 3.3.1.7 Transient Coefficient and Related Variables

In order to produce an environment in which the traffic load is periodically increased for short durations, this series of parameters is provided. The number of messages generated at the higher rate per cycle, the ratio of the transient load to the steady-state load, and the total number of messages generated per cycle can be specified by input variables.

#### 3.3.1.8 Number of Tandem Trunks Spanned by Each Message

Self-explanatory.

#### 3.3.1.9 Number of Initial and Observed Delays

The simulation can be run for a specified period before data is recorded in order to minimize the effects of a poorly chosen initial state. The number of recorded delays is also specified in the input.

#### 3.3.1.10 Channel Load Factor

This variable specifies the long-term average utilization of each channel.

#### 3.3.1.11 Transmission Speed-up Factor

In the case of the Queueing-Forced Hold case (Model Two) trunks can support transmissions at rates greater than which messages enter and leave the backbone network because of subscriber terminal equipment limitations with low volume subscribers. The amount of this transmission speed-up is designated by this input variable.

#### 3.3.1.12 Retrial Time Interval

This parameter is applicable to the Retrial Models (Three and Four). It designates the length of the time interval between successive attempts to place calls that have been blocked.

#### 3.3.1.13 Switching Delay Constant

This variable specifies the duration of the connection delay in Model Four.

### 3.3.2 Output Variables

#### 3.3.2.1 Average and Maximum Queue

For Models One and Two this represents the average and maximum number of messages simultaneously stored at a switching center. For Models Three and Four, this reflects the average and maximum number of blocked messages simultaneously awaiting transmission at a source.

#### 3.3.2.2 Average and Maximum Hold Time Queue

This is similar to the preceding output variables, but instead of considering the number of stored messages, the sum of their lengths is given. These variables, thus, can shed some light on storage requirements.

#### 3.3.2.3 Mean Message Delay

This represents the average time interval between when an originating subscriber first desires to release a message and when the ultimate recipient begins to receive this traffic.

#### 3.3.2.4 Ninety Percent Confidence Bound for the Mean Delay

This variable aids the program user in determining whether a sufficient number of delays have been observed. The estimate of this bound attempts to take into account any correlation which may exist among observed values.

#### 3.3.2.5 Fractile Message Delay

See input variable, Paragraph 3.3.1.6

#### 3.3.2.6 Ninety Percent Confidence Bound for Fractile

An estimate of the ninety percent confidence bound for the actual fractile associated with the computed fractile message delay is offered here.

### 3.4 Program Verification

Extensive steps were taken to ensure the proper operation of the program. Nevertheless, since many of the cases treated by the program have not yet been solved by analytical methods nor have been investigated by other simulation programs known to the author, no fully conclusive test existed to protect against the possibility of unobtrusive logical or truncation errors.

An approximate idea of the number of input parameters included in the program was provided in Paragraph 3.3.1. Although the logic associated

with each of these parameters was individually checked for particular values, proper operation for all possible combinations of these parameters over a wide range of values was not feasible to carry out. Furthermore, the complexity of the logic in certain instances approached sizable proportions. In a limited number of situations, exact[1] or approximate[2] analytical results existed which could be utilized.

### 3.5 Run Length Criterion

One fundamental problem associated with Monte Carlo digital simulation is in determining the number of observations (i.e., the sample size) which should be executed during any given run. If the number of observed delays is too small, errors in the result may render it unsuitable for its intended application. If too many messages are processed, the cost incurred in obtaining the result will be excessive.

The first concern, in arriving at an appropriate number, is determining the maximum size error that can be tolerated. This generally can be related directly to the purpose for conducting the simulation, to the questions which it hopes to answer, and to costs associated with using data with errors of different magnitudes. In the present situation, we are trying to ascertain the ratio of transmission capacity required by the two switching concepts. For the computer runs executed to support the conclusions offered

---

[1] Examples of exact results are exponential, fixed, and two-phased Erlangian hold times for the single node, Model One case.

[2] Examples of approximate solutions are for the short retrial, exponential hold time, single node, Model Three case and the exponential hold time, ring configuration, Model One case.

in Part II, a tentative ten percent error criterion[1] was arbitrarily adopted in lieu, of a formal analysis of the situation requirements.

After an acceptable error standard has been determined, it is necessary to develop an efficient and accurate measure of errors actually incurred. Classical procedures concerned with independent observations from a universe are not directly applicable since in the present situation adjacent message delays may be highly correlated.

Toward this end, it was decided to construct a criterion to reflect errors experienced (both for average and fractile statistics) which takes into account existing correlation. The criterion adopted does not provide estimates of actual errors encountered but only provides indirect indications of the "maximum likely" error; i.e., ninety-five percent confidence bounds. Error indices used by STARCON are described below.

### 3.5.1 Confidence Interval for the Average Message Delay

To obtain an estimate for the mean of the message delay distribution, a simple arithmetic average of the observed values is used by the program. Of interest, therefore, is the stochastic error which is associated with this statistic. A procedure for estimating this error is offered herein.

---

[1] The ten percent error criterion refers to a policy in which runs are made sufficiently long so that the computed 95% confidence interval does not exceed ten percent of the sample performance statistic. This policy, however, was not adhered to in cases where (1) it led to excessive computer running times and (2) greater error still permitted sufficient discrimination in arriving at reliable conclusions.

Observations of message delays are accumulated into blocks in the order they are produced. The size of the blocks is determined so that twenty blocks contain the entire sample. An average value, $z_k$, is then computed for each block. The adjacent block averages are generally much less correlated than individual observations. It is assumed that correlation which does persist in block averages can be approximately described as decreasing exponentially as a function of distance (i.e., number of block averages away).

Provided the serial correlation decays sufficiently rapidly, the variance of the sample average, $z_{av}$, i.e.,

$$z_{av} = \frac{1}{20} \sum_{k=1}^{20} z_k$$

is asymptotically equal to the expression

$$\sigma^2_{z_{av}} = \left(\frac{\sigma^2_z}{20}\right) \left[1 + 2 \sum_{k=1}^{\infty} \rho_k\right]$$

where $\sigma^2_z$, the variance of the block average, can be estimated by

$$\sigma^2_z \approx \frac{1}{19} \sum_{k=1}^{20} \left(z_k - z_{av}\right)^2$$

and where $\rho_k$ is the serial correlation coefficient of order k.

If we now make the geometrical decay assumption, i.e., $\rho_k = \rho_1^k$, we have

$$\sigma^2_{z_{av}} = \frac{\sigma^2_z}{20} \left[ \frac{1 + \rho_1}{1 - \rho_1} \right]$$

If in addition we assume the distribution of $z_{av}$ is normal by invoking the Central Limit Theorem, the ninety-five percent confidence interval simply becomes:

$$z_{av} \pm 1.96\sigma_{z_{av}}$$

Standard techniques are employed to estimate the population serial correlation coefficient of order one.

The technique described above is certainly superior in the present application to any method which ignores serial correlation. Nevertheless, it suffers from certain shortcomings. First, it does not produce a true ninety-five percent bound since the serial correlation estimate may be biased, it is not linearly introduced into the expression for $\sigma_{z_{av}}$, and it may not decay exponentially. Second, the use of block averages instead of individual observations reduces the number of degrees of freedom in the estimated statistics and thus increases the volatility of the confidence interval.

Serious consideration has recently been given to employing more advanced techniques that have resulted as an outgrowth of time series analysis. These techniques, however, significantly increase the computational burden which in turn adds correspondingly to the computer processing time.

### 3.5.2 Confidence Interval Associated with the Fractile Message Delay

Straightforward procedures are used to determine the message delay for the sample that corresponds to the fractile stipulated

in the input. Once this message delay has been computed the question arises as to what range of population fractiles it is likely to represent. The following technique was employed to answer this question.

Observed values of message delays are transformed into a binary series. This is accomplished by converting all values equal to or less than the computed fractile message delay to one, while other values are set to zero. This new series is then processed in the identical manner as the original message delays in Paragraph 3.5.1, or, in general, as if it constituted the sample generated by the simulation.

The average of the new series is, of course, equal to the fractile stipulated in the input. The variance of this statistic is used as a criterion to reflect its stochastic accuracy as an estimator. The resulting variance can then be used to approximate a fractile confidence interval.

If one desires some insight regarding the range of delays implied by the computed fractile confidence interval, the original sample can be inspected to determine which sample delay lengths correspond to the upper and lower bounds of this interval.

## 4.0 OTHER CONSIDERATIONS

### 4.1 General

The STARCON model does not permit investigation of all parameters affecting the choice between message and circuit switching. A brief qualitative discussion of some of the considerations not treated by the present model is offered below. This discussion should serve to broaden the base for comparing the two switching concepts as well as to illustrate certain shortcomings that could result from using the model's findings without proper discretion.

### 4.2 Configuration and Traffic Distribution

To study the effect of tandem trunks, the ring configuration was employed in the STARCON Model. Other configurations with different network topologies and message distribution patterns may exhibit dissimilar characteristics and should also be examined. Furthermore, the effects of adaptive routing should be studied, since this capability might not enhance the performance of the two switching concepts equally and at the same time may not be as easy to implement in each case.

### 4.3 Dialing and Header Processing

Although the STARCON model examines message delays with both switching concepts, it does not consider the time associated with dialing[1] when circuit switching is used.[2] In the case of message switching, the time involved

[1] Of course this is a function of the mechanism used, e.g., dial pulses, touch tone, etc., and whether the operation is manual or automated.

[2] Assumptions made with regard to delays other than from dialing, in establishing end-to-end connections, are fairly rudimentary with the existing STARCON Model. A fixed delay is assumed to be encountered at each tandem switching center for linking the input and output channel. During this period all channels involved in the connection between the originator and the switching center are considered in a busy state.

in composing the header and in transmitting it through each link along the message's route was also not taken into account. In subsequent studies, these factors may well deserve special attention.

### 4.4 Multiple Addressed Message

The effectiveness of both switching concepts to distribute multiple address messages and to offer a general broadcasting capability was not examined by the model. It would appear that message switching offers greater flexibility in performing these functions, especially when many different combinations of addressees from messages originating at the same source are required. In the case of circuit switching, it is necessary to decide whether a simultaneous transmission policy should be adopted (with or without camp-on), whether messages should be forwarded on from recipient to recipient, whether messages should be transmitted sequentially in time and all emanate from the original source (as if they were different messages), or whether a combination of the above should be used. Furthermore, end-to-end error correction using a feedback scheme becomes difficult to perform with circuit switching when multiple addressed messages are involved.

### 4.5 Message Precedence

Methods by which messages of different precedence levels are serviced with the different switching techniques should be compared. Store-and-forward offers greater facility in processing a large number of precedence levels. On the other hand, circuit switching provides for the immediate interruption (preemption[1]) of messages of lower precedence but then is not very efficient (requiring complete retransmission) in handling the interrupted traffic.

---

[1] Of course the store-and-forward approach does not preclude provision of a preemption feature.

## 4.6 Performance in Post-Attack Environment

In the post-attack environment, it is necessary to (1) service generated traffic with a reduced facility capacity and (2) discover and utilize existing paths (after suffering losses in transmission plant and switching centers), which may have been inefficient while all facilities were intact. The first capability is examined by the model but the second is not. Store-and-forward networks often do not have provisions for automatically rerouting traffic, while this capability generally does exist with circuit switching systems. Nevertheless, the entire question of the relative performance of the two switching concepts in the post-attack environment requires more intensive investigation. Circuit switching systems may tend to offer (for the same total cost) more possible paths between given pairs of subscribers and in this sense are more survivable.

## 4.7 Security

When key stream generators are used with message switching to provide message security, it is typically performed on a link-by-link basis. Hence, traffic is in the clear (or at most is protected by off-line super-encryption) within each switching center. Circuit switching, on the other hand, can use these crypto generators on an end-to-end basis. Nevertheless, when this is done, call flow security may not be realized and signaling information will be in the clear. Before key stream generators can be used on an end-to-end basis in large common user networks, furthermore, certain operational problems require solution with regard to the distribution, selection, and matching of codes.

## 4.8 Acknowledgment

With circuit switching, the originator of a message is naturally aware of whether and when the addressee has received his traffic. In the case of message switching, such timely acknowledgment, in general, is not

possible. Some store-and-forward systems, moreover, offer no provision for an automatic acknowledgment of receipt; thus, protection against "lost" messages is lacking.

### 4.9 Equipment Flexibility

Store-and-forward switching offers greater flexibility through its ability to perform such operations as format and code conversion and speed translation. This permits the use of tandem circuits with different bandwidths as well as the pairing of dissimilar types of subscriber terminal equipment. This latter equipment, thus, can be selected on the basis of individual user requirements.

### 4.10 Message Quality

With store-and-forward networks, errors are corrected and the signal is reshaped at each tandem switching center along the message's route. Circuit switching, on the other hand, may perform these functions only at subscriber terminals; in certain instances, the signal quality may suffer on this account. Nevertheless, this problem is not intrinsic with the circuit switching concept and can be avoided by the installation of appropriate equipment when needed at the switching centers.

### 4.11 Message Length

Certain store-and-forward networks are designed such that specific sections of input and output core registers are associated exclusively with individual trunk channels and subscriber lines. This places a limitation on message length which does not exist with circuit switching. Other designs for store-and-forward, however, can avoid this restriction.

## 4.12 Subscriber Terminal Equipment Requirements

Message switching allows a subscriber to dispatch traffic immediately, thus eliminating the need for temporary storage, and further avoiding the necessity of repeated attempts to release a message when earlier attempts have failed. Thus, message switching places a smaller burden on subscriber equipment,[1] although such functions as error control, security, header composition, etc., still have to be performed at the terminal stations. On the other hand, message switching places a significantly higher burden on the switching centers which may reprocess a single message several times as it passes from origin to destination.

Certain functions which are performed automatically with message switching have typically been performed manually with circuit switching. An example is the dialing operation and the execution of retrials when a busy signal is encountered. Another is the performance of message logging operations. Thus, in certain instances, it would be necessary to develop new terminal equipment in order to offer a subscriber the same service with circuit switching as with store-and-forward.

---

[1] Certain types of transmitting devices use punch cards, paper tape, or magnetic tape as a data source. In such cases, a subscriber may not have to make any special provisions to store data that must be temporarily retained with circuit switching.

## 5.0 CURRENT PROGRAM AND AREAS WORTHY OF MORE INTENSIVE STUDY

The current effort is channeled into three areas. These ar[illegible]s follows: first, an examination is being made of various criteria to mea[illegible]e the statistical accuracy of simulation results (this investigation is warranted since certain conventional techniques have proved unreliable in cases recently simulated); second, the STARCON Model is being exercised to determine the extent to which message switching actually reduces required transmission plant capacity in specific operational situations; and finally, a more intensive investigation is being conducted to determine the manner in which functional requirements in military communication systems dictate which switching technique should be used.

Possible extensions to the current program include the following: an investigation of more generalized configurations than STARCON can currently simulate could be conducted to study how changes in r[illegible]twork topology and message distribution patterns affect the efficiency of the two switching concepts; and an effort to construct meaningful parametric cost models of circuit and message switching centers could be undertaken to assist in the overall cost/effectiveness analysis.

(This page intentionally left blank)

# APPENDIX A-I

## COMPUTER PROGRAM LOGICAL DESIGN AND OPERATING INSTRUCTIONS

### 1.0 INTRODUCTION

In Section 6.0 (Part I), certain aspects of the STARCON Model are described. These include a summary of the parameters investigated by the Model, an abbreviated discussion of the computer program's input and output variables, an indication of the steps taken toward program verification, and finally an explanation of the criterion which is used to determine what constitutes a sufficient computer run length.

This appendix discusses the logical design of the STARCON Model and presents certain operating details. In particular, the deck structure, the format and content of the card input, and the computer printout are described.

## 2.0 DESCRIPTION OF THE PROGRAM LOGIC

The program consists of five sections (see Figure A1):

1) The Main Program

2) The Initialization Routine

3) The Simulation Routine

4) The Histogram Routine

5) The Statistic and Output Routine

The Main Program reads all of the input into the program and structures the program. It is designed so that different cases can be simulated during one computer run.

The initialization routine transforms the input to computer units and sets up initial dummy messages at each source (i.e., originating subscriber) in the system. These dummy messages start the simulation process. They are not included in the statistical results.

The Simulation Routine performs the simulation of the models described in this paper. The simulation consists of continually maintaining and updating five lists of describing the messages in the system. These five lists describe

1) the time at which the message originated;

2) the time at which the message will change (or attempt to change) its state;

3) the length (time duration) of the message;

4) the present position of the message in the system; and

5) the position in the system at which the message will exit.

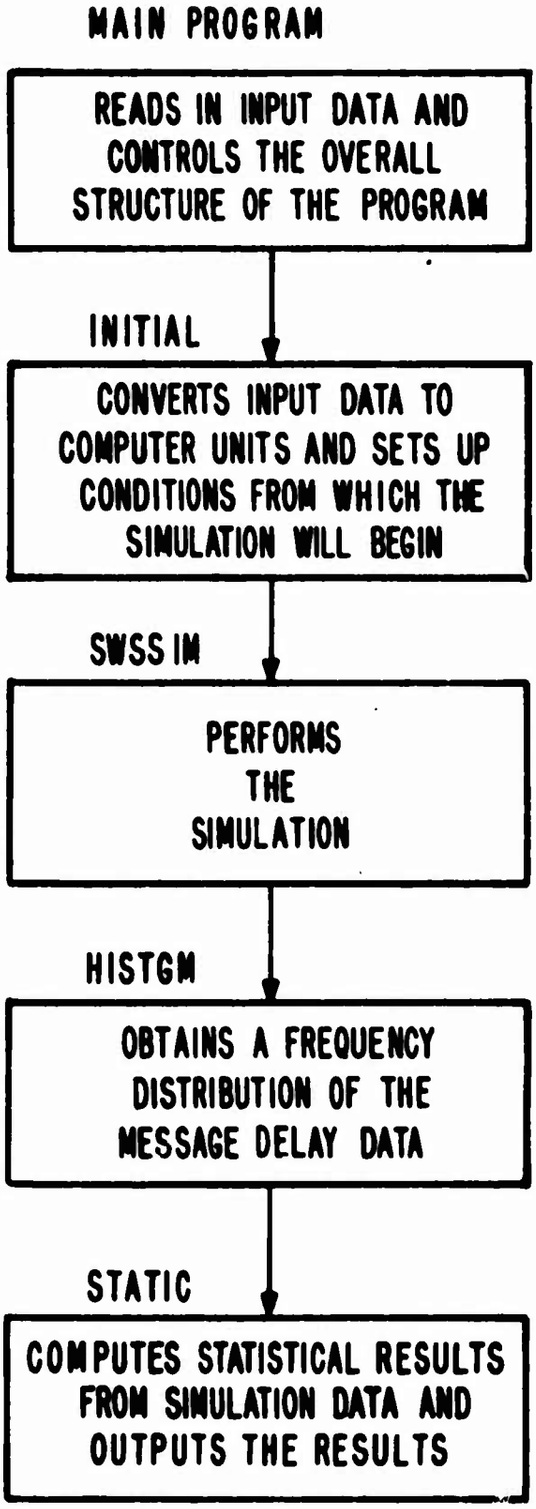



Figure A1. Overall Program Flow Chart

These five lists completely describe the status of a message and are ordered according to the time at which the message will change state. There are three possible states into which a message can be categorized:

1) a source message,

2) a channel message, and

3) a queue message.

A source message is defined as a message that has not as yet entered the system but will be the next message to originate from a given source. There is always one source message for each source in the message roster. At the time a source message enters the system, a new source message is generated for that origination point.

The channel message can be further divided into two substates. The first is the actual channel message. This is a message which is utilized to simulate the occupancy of a channel by a message. When a message tries to use a channel and finds one free, not only is an actual channel message created, but a temporizing channel message is inserted at the next node pending a determination of the message's next state at this new location. All actual channel messages are eliminated from the message roster when they arrive at the top of the list. Temporizing channel messages more generally consist of all types of messages which do not naturally fall under the categories of source, actual channel, or queue messages. Examples are a message in the forced hold state in Model 2, a message experiencing a switching center connection delay in Model 4, as well as a message which has entered a new stage but whose fate has not yet been determined.

A queue message is, as its name implies, a message which is waiting in a queue either for a channel to become available in the queueing models or for the retrial interval to elapse in the circuit switching models.

The Simulation routine always examines the first message in the list, determines what is to happen with that message, records information for the statistical analysis, reschedules the message, and examines the new first message in the list. This process is continued until the desired number of message delays are obtained.

The Simulation routine can be run in two modes of operation which are determined by the input parameters. The first mode is to simulate the system's operation without monitoring and statistically analyzing the delays incurred. This mode is used to let the system build up to a likely state, i.e., a state such that certain statistical quantities will converge more quickly than they would if monitoring began when the system was in the null state. After this state has been attained, the mode of operation is changed to include a statistical analysis of the system's experience.

The Histogram routine processes the waiting time data to obtain a frequency distribution of waiting times. Message waiting time is defined as the duration between the time that a message enters (i.e., is first offered to) the system and the time it has begun to be received by its destination; in other words, waiting time is the total time that a message is in the system minus the time duration (i.e., length) of the message. This information is recorded on tape by the Simulation routine and read in and processed by the Histogram routine. The Histogram routine constructs a frequency distribution of waiting times and determines a waiting time value corresponding to the specified fractile.

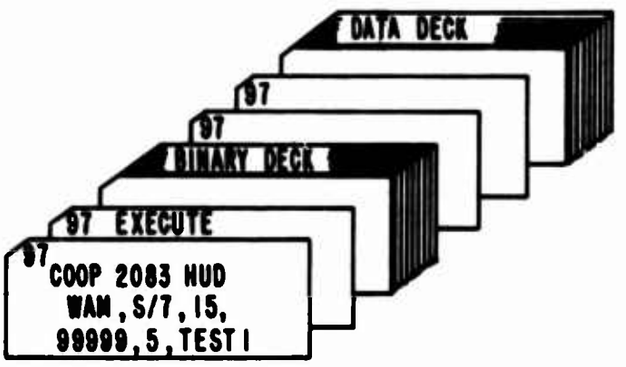
The Statistic and Output Routine computes, from the simulation output, the sample statistics described in paragraph 3.3.2. The routine divides the

simulation output into block averages consisting of both twenty and also fifty consecutive delay observations. Correlation coefficients and confidence bounds are then computed from these block averages. A more complete description of the output is given in Section 3.3 of this Appendix.

## 3.0 COMPUTER PROGRAM OPERATING PROCEDURE

### 3.1 Deck Structure

The structure of the deck for this job, to be run on a CDC 1604, under the CO-OP Monitor System in the GO mode of operation is as described below.




The first card, the job card, describes the following information:

a) 7-9 punch in column one defines the card as a monitor card.

b) COOP - tells the computer that the COOP Monitor System is to be used.

c) 2083 HUD - is the project identification number for the program.

d) WAM - the user's initials.

e) S/7 - specifies that a scratch tape designated as symbolic unit 7 is to be used.

f) 15 - designates in minutes the maximum running time requirements of the program.

g) 99999 - specifies the maximum number of lines of output which the program could produce (99999 is an artificially large limit since the program produces only a few lines of output per case).

h) An indicator that designates which dump and diagnostic procedures will occur should an error occur in the program.

i) A user identification name for the program. (Note - the written information on the card must be ended with a period.)

The second card in the deck is an execute card. This specifies to the computer that the deck is a GO job (i.e., the program is to execute and there are no routines to be compiled). The execute card must have a 7-9 punch in column one and must end with a period. The second card is followed by the binary deck representation of the program. The binary deck must be terminated by two cards with a 7-9 punch in column one. The input data deck immediately follows.

## 3.2 Input Data Cards

The reader is referred to Figure A2 which depicts the format of the input cards.[1] Three cards are required to specify each case. The following explanation is offered for the nomenclature used herein.

[1] Blank fields denote decimal integers (Iw in FORTRAN-63 notation) which are right justified; fields containing a decimal point denote floating point decimals without exponent (Fw.d in FORTRAN-63 notation).

1) Case Number

This number is used to identify the output from a run and has no influence on the program's operation.

2) Model Number

This number determines which of the four models is to be simulated. The reader is referred to Paragraph 3.3.1.1 of Part I.

3) Number of Nodes

There are only two possible circuit configurations which can be simulated with this program, namely:

a) A system with one source, one unlimited queue, and one trunk group which can have one or more channels.

b) A ring type system having from two to thirty-two sources, unlimited queues, and trunk groups with one or more channels per trunk.[1]

4) Number of Channels

This variable specifies the number of channels that make up a trunk group. (Note: each trunk in a ring configuration will have the same number of channels.)

[1] There is no limit to the number of channels that make up a trunk group. Nevertheless, since the number of channels and nodes influence the number of messages in the message roster, certain combinations of values of each could cause computer core to be exceeded. Since the number of messages in the system at one time is determined by a random process, the only means of determining whether a system with a large number of trunks and channels will not exceed the core limitations is by experimentation.

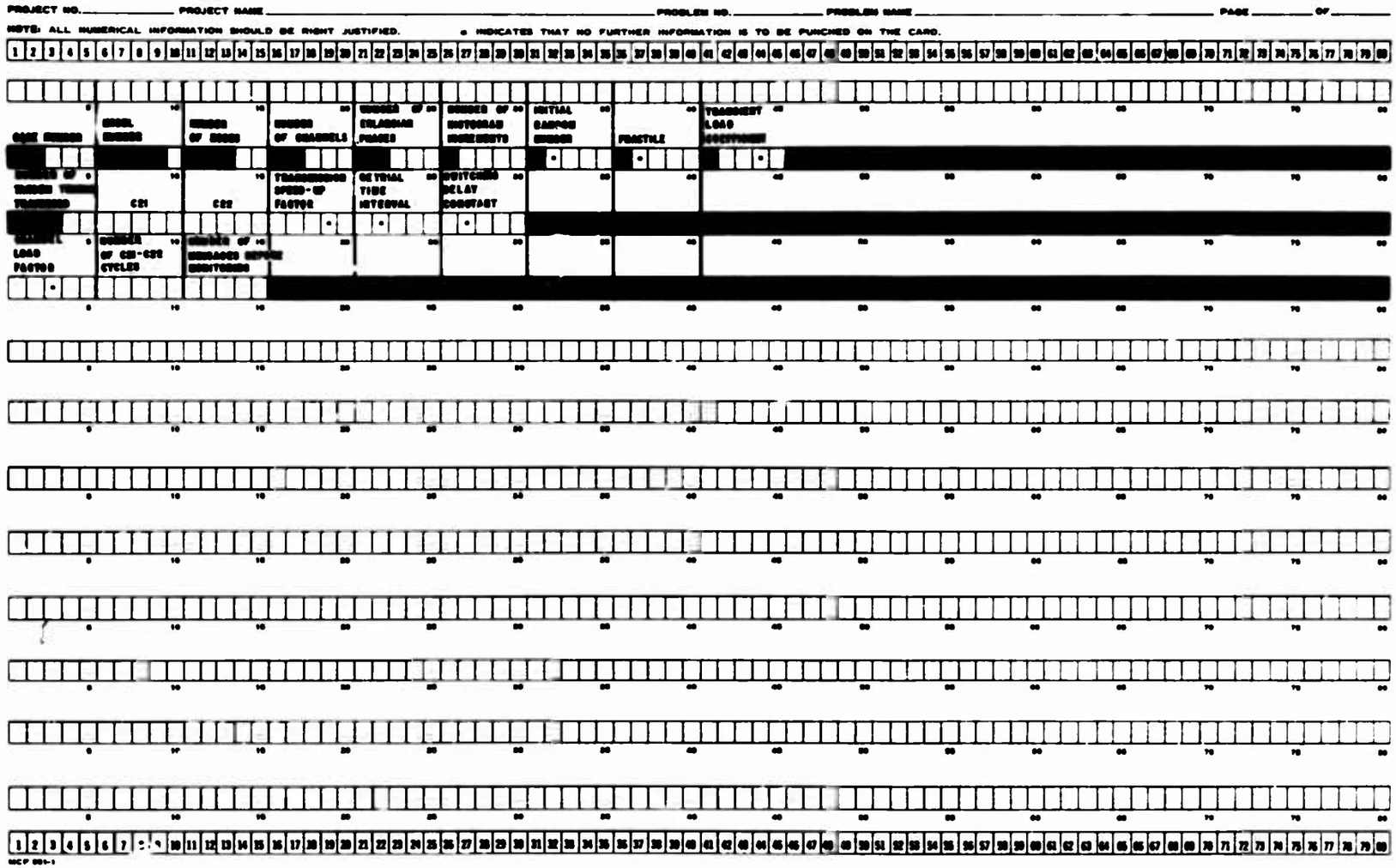



Figure A2. Input Format

5) Number of Erlang Phases

The Erlang Distribution can be described as the distribution which is the result of the sum of K independently exponentially distributed random variables, each one having a mean of $K^{-1}$. This distribution is used to determine the length of the messages (i.e., their holding time). In the special case where zero is specified, constant holding times are produced.

6) Number of Histogram Increments

This parameter is used to determine the number of increments in the histogram which is used to determine the waiting time value corresponding to the input fractile (see below). The maximum value which may be used for this parameter is one thousand.

7) Initial Random Number

This parameter affects the sequence of pseudo-independent rectangularly distributed numbers which are used by the simulation.

8) Fractile

This parameter permits one to determine that message delay such that shorter messages comprise a stated fraction of all delays observed. This fraction is referred to as the fractile.

9) Transient Load Coefficient

This parameter, whose value is greater than one, is used to increase the message generation rate during the C21 part of a monitored cycle as described below in the description of C21 and C22. This parameter should be set to zero if C22 is zero.

10) Number of Tandem Trunks Traversed

This parameter specifies the number of tandem trunks through which each message will be routed. This number must be greater than zero and less than or equal to the number of nodes in the system.

11) C21

In normal operation, this parameter specifies the number of monitored messages which must pass through the system. The program also provides the possibility of alternating between two message generation rates. In this case, C21 specifies the number of monitored messages per cycle (see variable 17) that are generated at a higher than normal rate. This higher rate is equal to the normal generation rate (variable 16) times the Transient Load Coefficient (variable 9).

12) C22

This parameter is specified as zero when C21 specifies the total number of monitored messages. When the system is to be simulated with alternating generation rates, C22 specifies the number of messages in each cycle which will be generated at the normal rate.

13) Transmission Speed-up Factor

This parameter is only applicable to Model Two. It specifies that wideband trunk channels are being used and defines how many times faster these channels are than the subscriber lines.

Note: Since it is assumed that cost per trunk is proportional to trunk bandwidth, the program internally divides the number of channels per trunk (variable 4) by the

speed-up factor. Thus, caution must be exercised in insuring that the number of channels per trunk is evenly divisible by the speed-up factor.

14) Retrial Time Interval

This parameter is only applicable to Models Three and Four. It determines the time interval which an originator subscriber will wait between successive attempts to acquire an end-to-end connection.

15) Switching Delay Constant

This parameter defines the message delay time at each switching center to complete a connection. This constant applies only to Model Four.

16) Channel Load Factor

This parameter is used by the program to determine the message generation rate, $\lambda$. The formula used is:

$$\text{Message Generation Rate}^{1} = \frac{(\text{channel load factor}) \times (\#\text{ of channels})}{\#\text{ of trunks through which each message must pass}}$$

If the Channel Load Factor is preceded by a negative sign, messages originate from each source periodically at $1/\lambda$ intervals. If the Channel Load Factor is positive, message initiations from each source are Poisson distributed with a mean rate of $\lambda$. The same expected load is offered to all trunk channels in the network.

---

[1] The denominator contains the mean message length, but since it is always assumed one, it is omitted from the formula.

17) Number of C21-C22 Cycles

This specifies the number of cycles (each consisting of C21 plus C22 messages) that are to be used in the simulation. If C22 is equal to zero, the number of C21-C22 cycles should be set equal to one.

18) Number of Messages Before Monitoring

This variable defines the number of messages that must pass through the system before message delays are monitored. This variable may be specified as zero.

## 3.3 Computer Printout

The reader is referenced to Figure A3 where a sample printout appears. The first five lines of output represent the input data in the order in which they appear on the input coding form. The next six lines comprise the output and describe the following statistical parameters in the order listed below.

a. TRAFFIC CARRIED - channel load factor times the number of channels per trunk.

b. QAV - average number of messages in queue per switching center.

c. Q MAX - maximum number of messages in queue at a single switching center.

d. Q HOLD TIME AV. - average total length of messages in queue per switching center.

e. Q MAX HOLD TIME - maximum total length of messages in queue at a single switching center.

f. XM - mean waiting time of a message in the system.

```
                                        CASE NUMBER     1
MODEL NO.  1  NO. OF NODES   1  NO. OF CHANNELS    2  NO. OF ERLANGIAN PHASES    1
NO. OF MIST. INCR. 1000  RANDOM NO.  .700  FC  .900  F      0
L =    1   C21 =   2000   C22 =      0   Z =      0   R =      0   B =      0
LOAD FACTOR    .80  NUMBER OF MONITORED GROUP PAIRS    1  NUMBER OF SIGNALS BEFORE MONITORING    200
                    TIME BEFORE SIMULATION =     .0197
 TRAFFIC CARRIED   1.0000
                     TIME AFTER SIMULATION =     .3625
   QAV =   2.8981  Q MAX =    17  Q HOLD TIME AV. =   2.8364  Q MAX. HOLD TIME =  21.5275
   XP  =   1.7883   E20  =    .4517   E50  =    .3865   R20  =    .3143   R50  =    .3827
BL  =       0   BU  =  10.4255
   XP# =    .9000   E20# =    .0973   E50# =    .0847   R20# =    .2122   R50# =    .3016
         WF =   5.1502     KWF =   494
                    TIME AFTER LAST CASE =     .4697
```

Figure A3. Sample Printout

g. E20 - 95% confidence interval for the average waiting time (calculated from the averages of twenty blocks that collectively comprise the entire sample); the bound is expressed as a fraction of the computed mean waiting time.

h. E50 - 95% confidence interval for the average waiting time (calculated from the averages of fifty blocks that collectively comprise the entire sample); the bound is expressed as a fraction of the computed mean waiting time.

i. R20 - first sample serial correlation coefficient computed from the twenty block averages.

j. R50 - first sample serial correlation coefficient computed from the fifty block averages.

k. BL - the maximum of (1) zero or (2) the mean waiting time minus four times the sample standard deviation.

1. BU - the mean waiting time plus four times the sample standard deviation.

m. The next five outputs XM≠, E2Ø≠, E5Ø≠, R2Ø≠, R5Ø≠, are computed in the same manner as their counterparts; i.e., XM, E2Ø, E5Ø, R2Ø, R5Ø, respectively. Nevertheless, the former set of statistics are computed from data which results after the original sample of message delays are transformed into a binary series. The reader is referred to Paragraph 3.5.2 of Part I for a more detailed description of this process.

n. WF - the waiting time value corresponding to the fractile specified in the input.

o. KWF - the interval in the Histogram in which the value WF occurs.

## 1.0 INTRODUCTION AND DELINEATION OF CONTENTS

A cost/effectiveness study has been undertaken that compares circuit and message switching in military communication networks that carry record traffic. The first phase of this effort was directed at identifying the types of functional requirements that are typically levied on such communication systems and at studying the ability of both circuit and message switched networks to provide a suitable operational capability to support these functions. As part of this inquiry, an investigation was initiated to compare the cost of corresponding system components in both types of communication networks when equivalent performance capabilities are offered by each.

The results of this first phase are summarized in Part I, and it is presumed the reader is familiar with its contents. Part I also contains a description of a digital simulation computer program, referred to as the STARCON[1] Model, that was developed to measure quantitatively the actual savings in trunking capacity that message switching permits in a variety of operational situations.

Since Part I was issued, a series of demonstration computer runs has been conducted, exercising various features of the STARCON Model. The results from these runs, in addition to providing the information sought above, afforded, as a by-product, insight into certain other areas. First, it was demonstrated that a commonly employed technique to measure the statistical accuracy of simulation results will in certain instances grossly understate the size of the error and in general not utilize all the relevant information in the sample. Second, it was found that in many cases of interest the number of

---

[1] Switching Technique Analysis for Record Communication Networks

messages that required processing in order to obtain a suitably reliable estimate of the average message delay, exceeded by orders of magnitude what had been commonly thought by some simulation practitioners to be a sufficient quantity. Third, an analytical approximation which has been proposed to estimate delays in certain store-and-forward situations is shown to be invalid. Finally the processing speed of the program (expressed in minutes per thousand simulated message delays) experienced during the demonstration runs has been tabulated for specific values of input parameters with each switching model. Except when high channel loading was used with multiple node configurations and when the effects of a rapid retrial rate were examined with circuit switching, these processing times fell within design goals.

## 2.0 COMPARISON OF TRANSMISSION FACILITY REQUIREMENTS

Part I indicated that a valid cost comparison between message and circuit switched systems should consider all major system components which include the switching centers, the backbone network transmission facilities, the subscriber homing lines, and the subscriber terminal equipment with its support. In order for the comparison to be meaningful, the systems involved should offer equivalent performance capabilities and the impact of providing any special operational features demanded by specific military users should be reflected in each case.

Execution of the above task would entail the overall design and economic analysis of circuit switched and message switched communications systems to satisfy given sets of specifications. An undertaking of this magnitude was not contemplated at this time but instead the objective has been to establish, where possible with a minimum of effort, general guidelines that could be used to select between the two switching concepts in different operational situations.

Analysis in Part I (Subsection 1.4) revealed that while message switching centers generally cost considerably more to acquire and operate than circuit switching centers (even when many fewer line and trunk terminations are involved), message switching permits some savings in the area of backbone trunking and subscriber equipment. If we for the time being ignore subscriber terminal cost variations, it may be asked under what sets of circumstances will transmission plant savings with message switching more than offset the increased cost of the switching facilities. The answer to this question involves knowing the ratio of the switching center costs for the two switching concepts, the ratio of the transmission facility costs, and the manner in which total system cost with either concept breaks down between switching and transmission facilities. To pursue this line of investigation, it

is convenient to introduce a limited amount of mathematical notation as follows:

$X_M$ = cost of switching facilities with message switching

$T_M$ = cost of transmission facilities with message switching

$X_C$ = cost of switching facilities with circuit switching

$T_C$ = cost of transmission facilities with circuit switching

We then define the following ratios:

$$r_X = \frac{X_M}{X_C} \qquad r_T = \frac{T_C}{T_M} \qquad \gamma_C = \frac{T_C}{X_C}$$

where in general:

$$r_X > 1 \quad \text{and} \quad r_T > 1 .$$

For any given pair of values for $r_X$ and $\gamma_C$, $r_T$ must equal some minimum value to break even by utilizing message switching. For all larger values of $r_T$, a net savings will be effected.

Figure 2-1 expresses the relationship between $r_X$, $\gamma_C$, and the break-even value for $r_T$. For example, if message switching centers cost three times (or more) circuit switching centers ($r_X \geq 3$) and the circuit switched network alternative calls for twice the total discounted dollar outlay over the life of the system for transmission facilities as for switching centers ($\gamma_C = 2$), reference to Figure 2-1 discloses that no amount of savings in the transmission plant can cause message switching to be the more economical alternative. On the other hand, if message switching centers cost two times their circuit

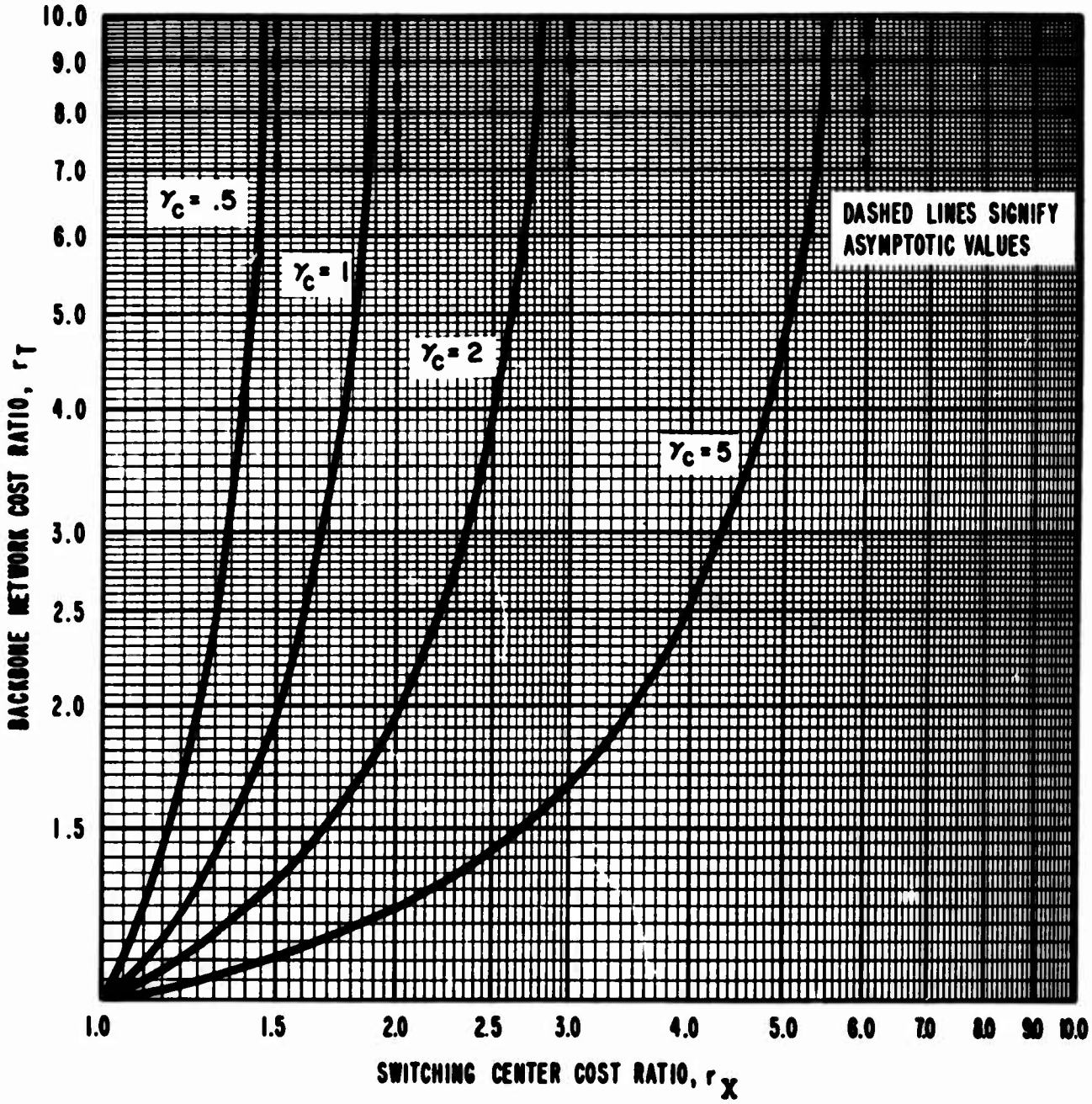



Figure 2-1. Relationship of System Element Cost Ratios to Break Even with Message Switching

switching counterpart ($r_X = 2$), it is necessary for message switching to cut transmission costs in half to break even.

The remainder of this section will be devoted to exhibiting actual values of $r_T$ (assuming costs are proportional to trunk group capacity) that resulted from simulating a spectrum of operational environments. For these recorded values of $r_T$, the reader may refer to Figure 2-1 to see which values of $r_X$ and $\gamma_C$ achieve a net savings with message switching. The reader may then determine whether, in his judgment, such values are realistic.

Again it should be re-emphasized that the above approach represents an oversimplification of the interrelationships among all pertinent factors. Naturally the degree to which message switching can effect savings in subscriber terminal equipment should be taken into account if significant. Also the relative ability of message switching and circuit switching to provide special operational capabilities (identified in Section 4.0 of Part I) and the cost implications of implementing such features warrants further investigation. Finally, it is worthwhile to note the possibility that an optimal (i.e., meeting performance objectives at a minimum cost) configuration could take the form of a hybrid system utilizing both circuit and store-and-forward switching elements. Nevertheless, the examination of these facets of the problem is considered beyond the scope of the current study.

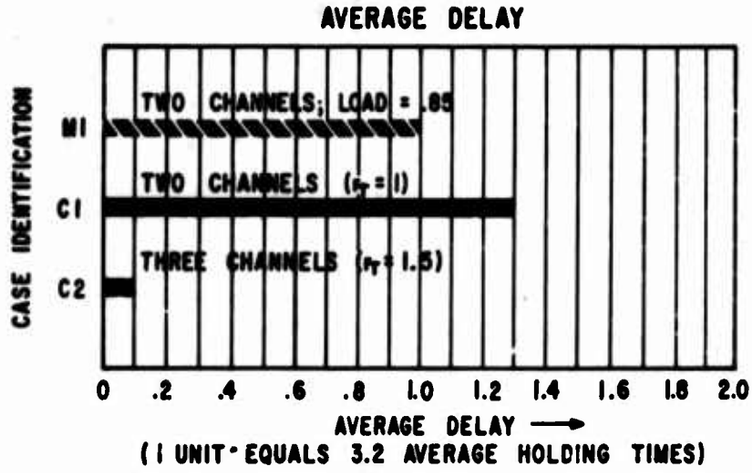
Investigation of the relative trunk capacity required by circuit switching and message switching was carried out by exercising the STARCON model. Equal volumes of traffic were introduced into both types of networks and the amount of trunking capacity needed in each case to experience equivalent average and ninety-percentile delays was determined. This comparison was conducted for different sets of values for parameters defining the operational environment. These parameters included channel loading, the number of channels per trunk group with message switching, the number of tandem trunks

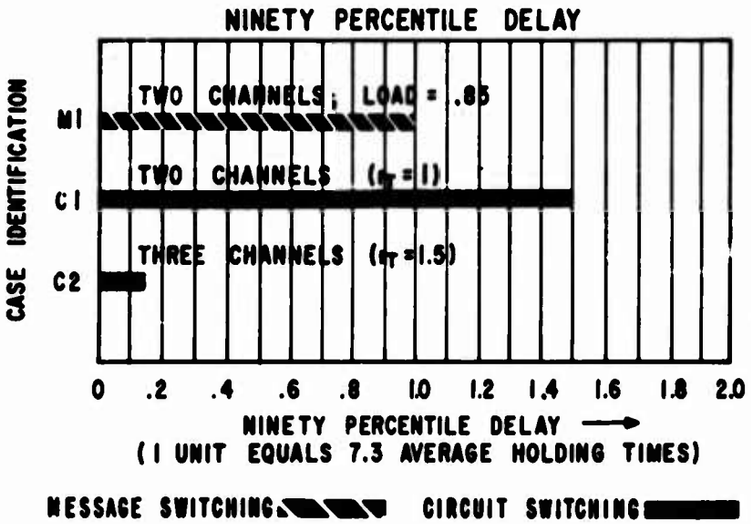
traversed per message, traffic transient specification, the type of message length distribution, the trunk transmission speed-up factor with message switching, and the retrial interval and connection delay interval with circuit switching. Definitions for these terms are provided in Part I.

## 2.1 Single Trunk Group

The results of comparing the performance of message switching and circuit switching, when an equivalent amount of traffic is offered to a single trunk group, is shown in Figures 2-2 through 2-5. Relative performance is measured by examining the average and ninety percentile delay encountered with each switching technique. Although a single trunk group does not constitute a network and merely represents an elementary situation, it is felt that gross relationships among variables would be indicative of more generalized situations. In addition, these results, when compared to what is experienced with a more complex configuration, show more clearly the effects of introducing successive tandem stages. When the above factors are considered along with the knowledge that the rate with which delays are generated and processed by the simulation with a single trunk group (and in some instances the speed with which convergence occurs) is significantly greater than with a multiple node configuration, it becomes reasonable to initially study this idealized situation.

By referring to Figures 2-2 through 2-5, one may observe the following phenomena. Message switching effects a more substantial reduction in message delays (with the same amount of trunking) when (1) the number of trunk channels per trunk group is low, (2) channel loading is high, and (3) there is significant variability in message length. In fact, for the case examined in

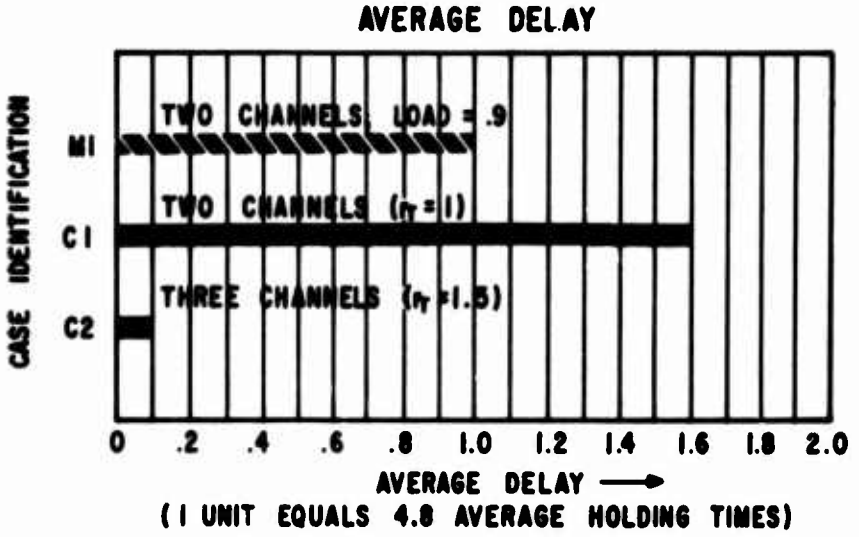


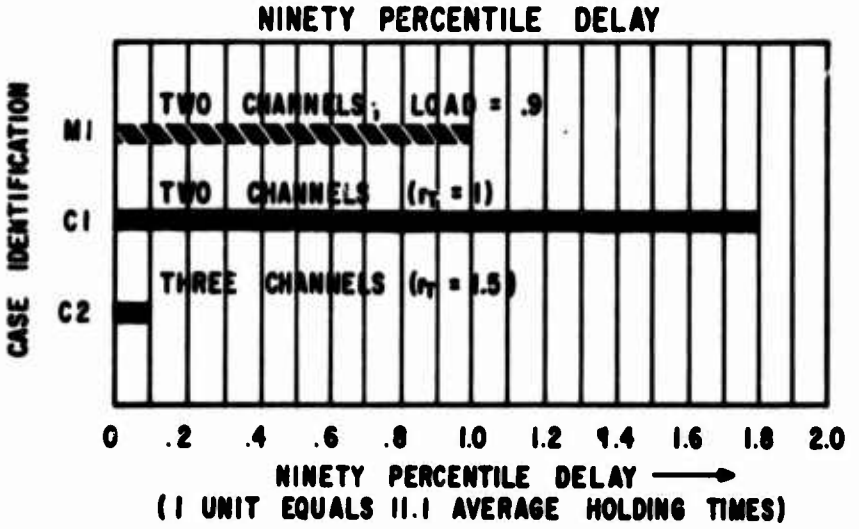




CASE SPECIFICATION

| | M1 | C1 | C2 |
|---|---|---|---|
| MESSAGE LENGTH DISTRIB. | EXPON | EXPON | EXPON |
| TRAFFIC VOLUME OFFERED AND CARRIED | 1.7 Erl. | 1.7 Erl. | 1.7 Erl. |
| CHANNEL LOADING | .85 | .85 | .57 |
| NUMBER OF CHANNELS PER TRUNK GROUP | 2 | 2 | 3 |
| RETRIAL INTERVAL | — | .3 | .3 |

Figure 2-2. Single Trunk Group (Small Number of Channels per Group)

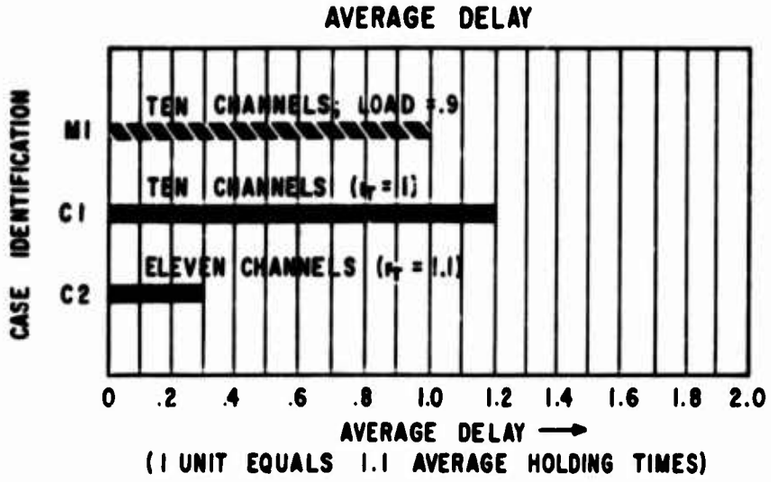


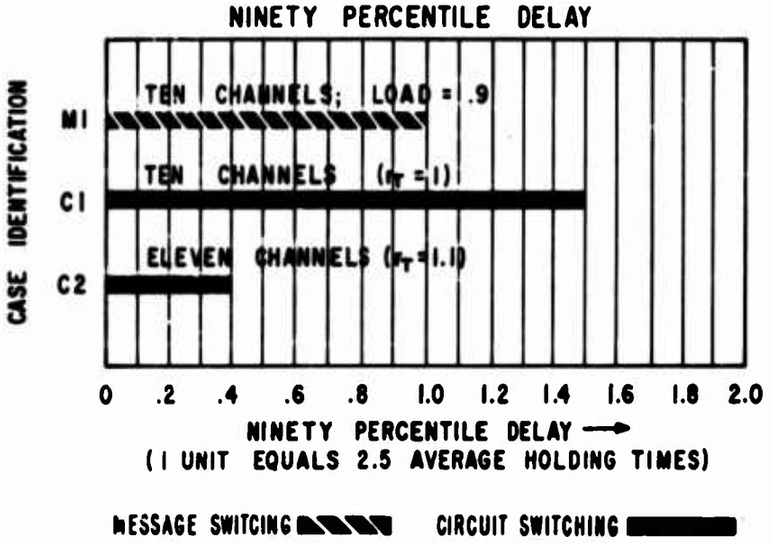




CASE SPECIFICATION

| | M1 | C1 | C2 |
|---|---|---|---|
| MESSAGE LENGTH DISTRIB. | EXPON | EXPON | EXPON |
| TRAFFIC VOLUME OFFERED AND CARRIED | 1.8 Erl. | 1.8 Erl. | 1.8 Erl. |
| CHANNEL LOADING | .9 | .9 | .6 |
| NUMBER OF CHANNELS PER TRUNK GROUP | 2 | 2 | 3 |
| RETRIAL INTERVAL | — | .3 | .3 |

Figure 2-3. Single Trunk Group (Small Number of Channels per Group)

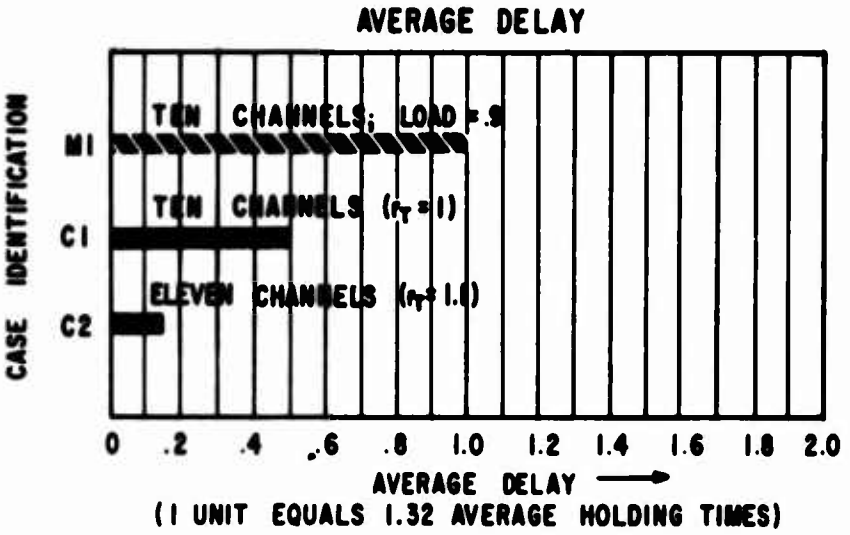


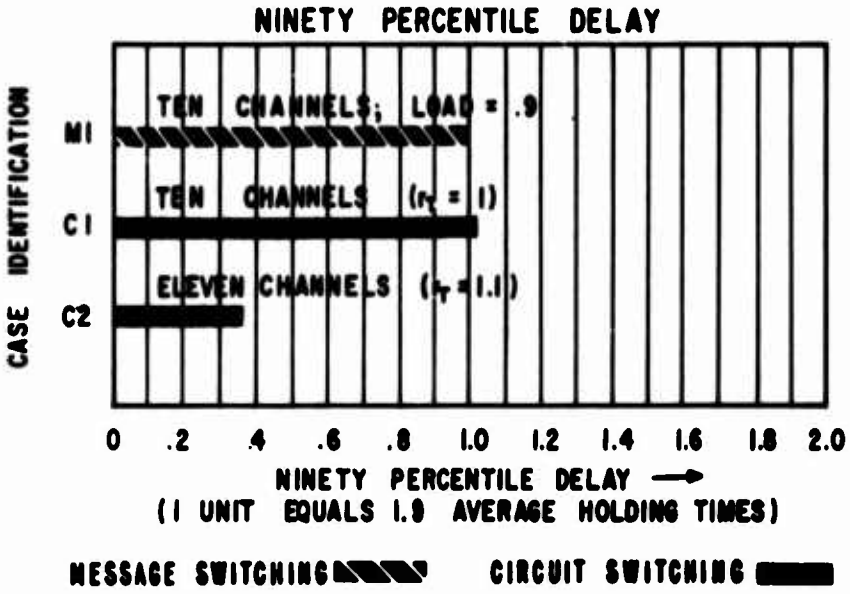




CASE SPECIFICATION

| | M1 | C1 | C2 |
|---|---|---|---|
| MESSAGE LENGTH DISTRIB. | EXPON | EXPON | EXPON |
| TRAFFIC VOLUME OFFERED AND CARRIED | 9.0 Erl. | 9.0 Erl. | 9.0 Erl. |
| CHANNEL LOADING | .9 | .9 | .82 |
| NUMBER OF CHANNELS PER TRUNK GROUP | 10 | 10 | 11 |
| RETRIAL INTERVAL | — | .33 | .33 |

Figure 2-4. Single Trunk Group (Moderate Number of Channels)




CASE SPECIFICATION

| | M1 | C1 | C2 |
|---|---|---|---|
| MESSAGE LENGTH DIST | CONSTANT | CONSTANT | CONSTANT |
| TRAFFIC VOLUME OFFERED AND CARRIED | 9.0 Erl | 9.0 Erl | 9.0 Erl |
| CHANNEL LOADING | .9 | .9 | .82 |
| NUMBER OF CHANNELS PER TRUNK GROUP | 10 | 10 | 11 |
| RETRIAL INTERVAL | — | .33 | .33 |

Figure 2-5. Single Trunk Group (Constant Message Length)

which all messages had a constant length, circuit switching actually outperformed message switching (owing to the forced hold doctrine[1] observed by the latter switching concept).

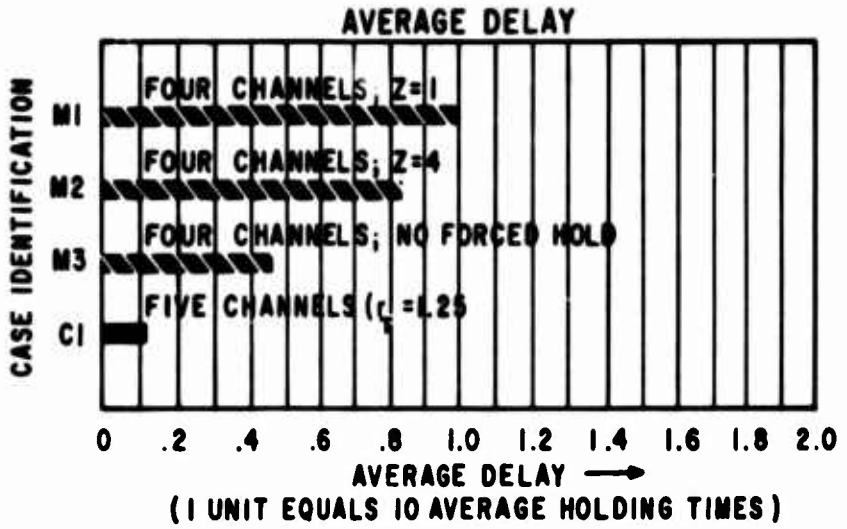
For all the single trunk group cases investigated, circuit switching achieved a smaller average and ninety percentile delay when the number of channels in the trunk group exceeded by one the number used with message switching. This is equivalent to an increase in transmission capacity of fifty percent ($r_T = 1.5$) when message switching used two trunks, and an increase of only ten percent ($r_T = 1.1$) when the message switching configuration had ten trunks in the trunk group. Such minimal savings in trunking requirements through using message switching are not expected to produce a net saving in overall system cost.

## 2.2 Multiple Node/Tandem Trunk

The ring configuration was used to investigate a twelve node configuration in which messages traverse three tandem trunks in passing from the originating to the destination subscriber. Four channels per trunk group were used when monitoring message switching performance; five channels with circuit switching. Both systems, however, were studied while supporting the same traffic volume under constant as well as transient loading conditions.

With a constant traffic load, the performance of message switching was observed without trunk transmission speed-up, with a maximum speed-up (for a four channel system keeping trunk group capacity constant) and with an operational doctrine which did not require forced-hold (i.e., Model One). The results are shown in Figure 2-6. Performance was improved by the

[1] Each message entering a switching center must be fully stored before it may be released to an idle outgoing trunk.




CASE SPECIFICATION

| | MESSAGE LENGTH DISTRIB | TRAFFIC VOL. OFFERED & CARRIED | MODEL NUMBER | TRUNK SPEED-UP, Z | RETRIAL INTERVAL | CHANNEL LOADING | NUMBER OF CHANNELS PER TRUNK GROUP |
|---|---|---|---|---|---|---|---|
| M1 | EXPON | 3.6 Erl | 2 | 1 | – | .9 | 4 |
| M2 | ↓ | ↓ | 2 | 4 | – | .9 | 4 |
| M3 | ↓ | ↓ | 1 | – | – | .9 | 4 |
| C1 | ↓ | ↓ | 3 | – | .33 | .72 | 5 |

Figure 2-6. Multiple Node/Tandem Trunk (Twelve Node Ring; Three Tandem Trunks Traversed)

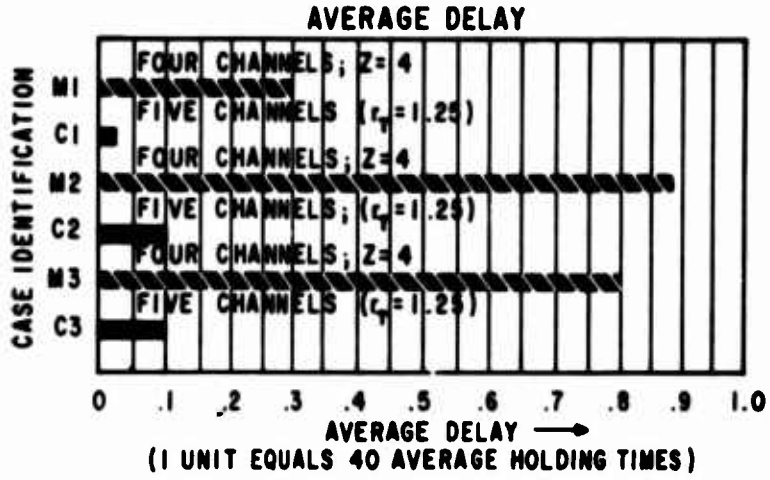
latter two operational features but in no instance did average or ninety percentile delays fall below what was achieved when circuit switching was used with only a twenty-five percent increase ($r_T = 1.25$) in trunking capacity. It is interesting to note that when it was no longer necessary to completely store messages at a switching center before releasing them to appropriate idle channels with message switching, the average delay was more than cut in half for the set of parameter values considered.

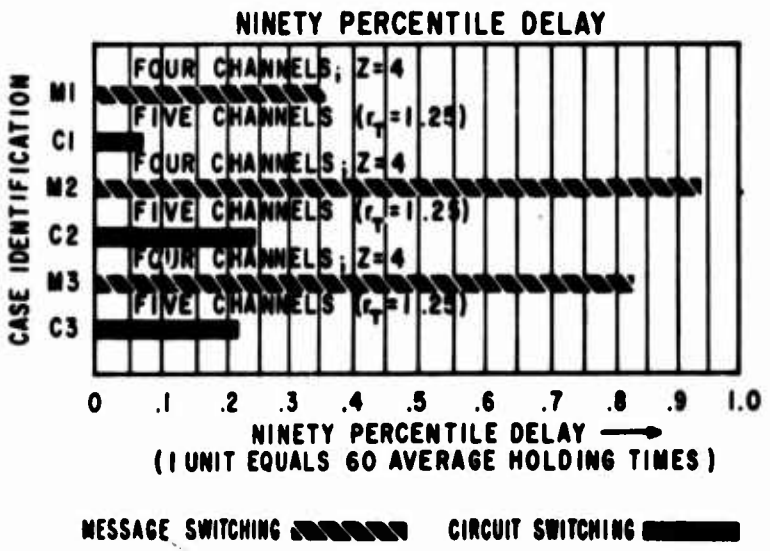
Message delays were observed when the two networks were subjected to time variant traffic loading conditions in which the intensity and duration of transient load peaks were altered for different cases. In all cases examined, however, the circuit switched network demonstrated a vastly superior performance (see Figure 2-7). Thus, for the entire spectrum of operational situations considered above, circuit switching, with only a twenty-five percent greater trunking capacity, yielded a much more favorable traffic delay experience.

## 2.3 General Conclusions

Message switching has been frequently proposed as an economical switching technique for data handling communications systems because of the large savings realizable in transmission facilities when compared to circuit switching trunking requirements. Before conducting this investigation, the author was unsuccessful, however, in locating quantitative data to support this thesis. It was suspected that a significant portion of the postulated savings was illusory since historical applications of these switching techniques have dramatized their performance under dissimilar sets of operating conditions.[1]

[1] Circuit switching has been traditionally used in the field of telephony where low trunk utilization is required to provide a high grade of service (i.e., a low fraction of blocked calls). Message switching, on the other hand, has been used exclusively to support record communications where it is, in general, not necessary to achieve an end-to-end connection almost instantaneously most of the time.




CASE SPECIFICATION

| TRANSIENT LOADING | TRANSIENT PERIOD (NO. OF CONSEC. MSSGS.) | CHANNEL LOADING | AVERAGE (E.I.) OFFERED & CARRIED | TRAFFIC VOL. TANDEM TRUNKS TRAVERSED | NO. OF CHANNELS PER TRUNK GROUP | TYPE OF SWITCHING | TRANSMISSION SPEED-UP & RETRIAL INTRVL |
|---|---|---|---|---|---|---|---|
| M1 | 1.170 | 100 | .927 | 3.71 | 3 | 4 | MSSG 4 |
| C1 | .936 | 100 | .742 | 3.71 | | 5 | CIRC .50 |
| M2 | 3.215 | 200 | .900 | 3.60 | | 4 | MSSG 4 |
| C2 | 2.570 | 200 | .720 | | | 5 | CIRC .33 |
| M3 | 1.930 | 400 | .900 | | | 4 | MSSG 4 |
| C3 | 1.540 | 400 | .720 | | | 5 | CIRC .33 |

Figure 2-7 Multiple Node/Tandem Trunk (Transient Response)

Thus the STARCON Model was designed, implemented, and exercised. Delays experienced with the two switching techniques were compared under a variety of situations which included variations in channel loading, the number of tandem trunks traversed by messages, the size of the trunk groups, the distribution of message lengths, and the nature of traffic transients offered to the system. In all cases examined, circuit switching achieved shorter delays if the number of channels per trunk group exceeded by one the number used with message switching.

A preliminary cost comparison of high-speed automated message and circuit switching centers (contained in Part I), that examined how individual components of the switching function were executed, led to the conclusion that message switching centers cost considerably more, even when they accommodate many fewer line and trunk terminations. For the modest trunk savings that message switching provided in the cases examined by this investigation, it is not expected that these savings would fully compensate for the switching center cost differential.

This study by no means constitutes an exhaustive investigation. Other configurations (defined by network topology and routing disciplines) and message distribution patterns naturally should be examined. The cost implications of message switching on subscriber terminal equipment also require evaluation. In addition, the cost of providing special operational features with each type of communication system warrants further study. Nevertheless, this study does demonstrate that relative trunk utilization is not a dominant consideration in selecting between the two switching techniques.

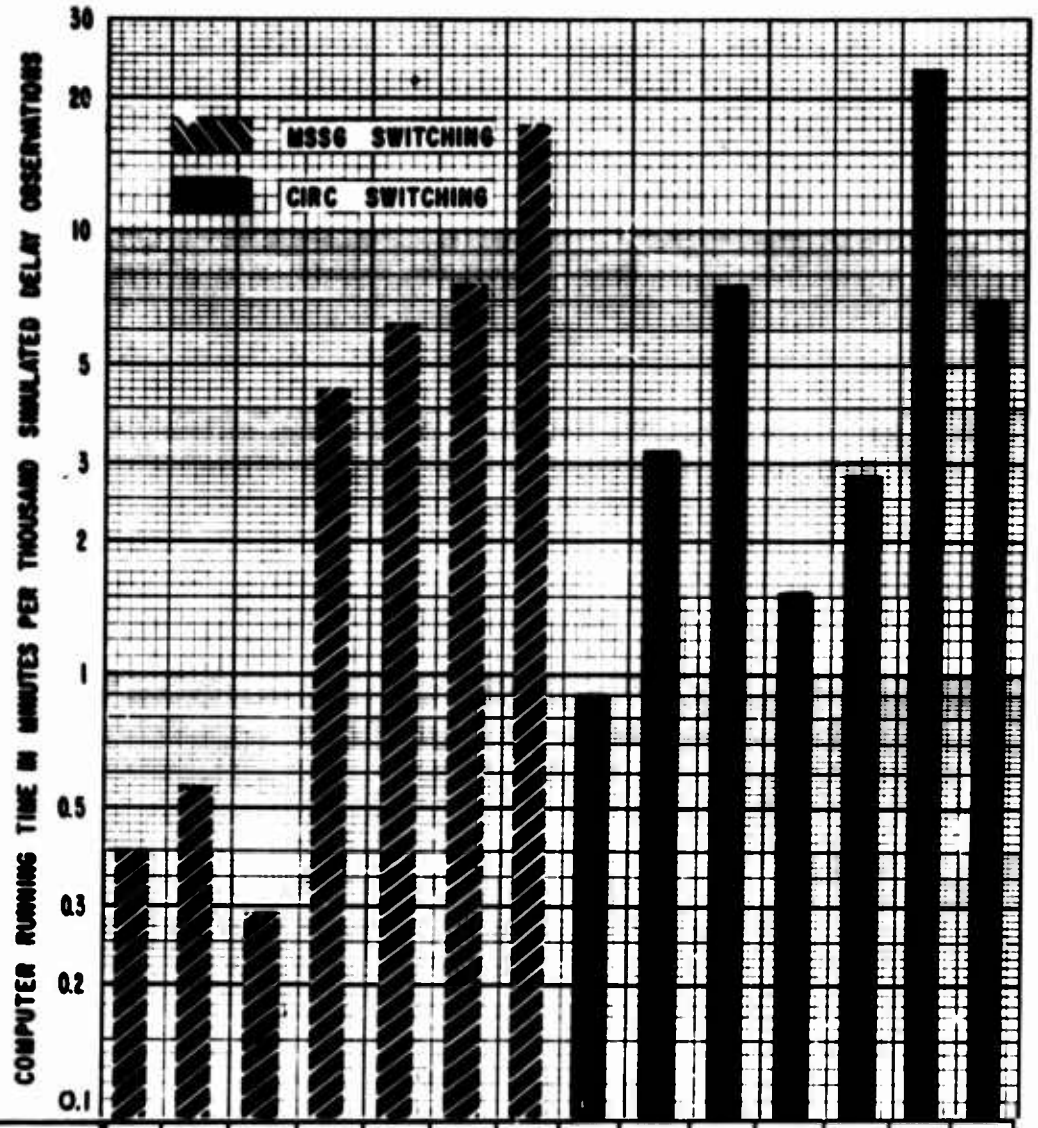
## 3.0 COMPUTER RUNNING TIME EXPERIENCE

In order for simulation to be an effective tool for aiding design decisions, the computer processing cost for carrying out the desired production runs should not be excessive considering the value of what is hoped to be learned.[1] The production time required for investigating a given situation is a function of (1) the desired estimation accuracy of the performance parameter being measured, (2) the computer speed, (3) the program design (including the type of language used, the efficiency of the compiler/assembler, etc.), (4) the complexity of the configuration, the routing and forwarding logic, and the detail with which they are represented, and (5) the numerical values of the parameters defining the situation being investigated.

The simulation runs recorded in this document were carried out on a CDC 1604 machine with a 32K word core. The program was written in FORTRAN IV. To acquaint the reader with the rate at which message delays were generated and processed by STARCON on the above machine for a spectrum of values for input parameters, Figure 3-1 has been prepared. Figure 3-1 shows that the processing rate is sensitive to the number of tandem trunks that individual messages traverse and the number of switching nodes in the total configuration, to the channel loading (particularly for circuit switching for the parameter ranges examined), to the number of channels per

---

[1] A related consideration is the avoidance of exceeding core capacity by the combination of the resident program and the parameters of the in-transit messages. The central processor's magnetic core was not saturated in any cases run to date with STARCON as runs were found to be "convergence limited."




| CASE DESIGNATOR | M1 | M2 | M3 | M4 | M5 | M6 | M7 | C1 | C2 | C3 | C4 | C5 | C6 | C7 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| TYPE OF SWITCHING | ← MESSAGE[1] → | | | | | | | ← CIRCUIT[2] → | | | | | | |
| CHANNEL LOADING | 0.9 | 0.95 | 0.95 | 0.9 | 0.95 | 0.95 | 0.95 | 0.8 | 0.9 | 0.9 | 0.9 | 0.8 | 0.9 | 0.9 |
| NUMBER OF CHANNELS PER TRUNK GROUP | 2 | 2 | 5 | 2 | 2 | 5 | 5 | 2 | 2 | 2 | 5 | 2 | 2 | 5 |
| NUMBER OF TANDEM TRUNKS TRAVERSED BY MSSG. | 1 | 1 | 1 | 3 | 3 | 3 | 4 | 1 | 1 | 1 | 1 | 3 | 3 | 3 |
| NUMBER OF NODES IN CONFIGURATION | 1 | 1 | 1 | 12 | 12 | 12 | 16 | 1 | 1 | 1 | 1 | 12 | 12 | 12 |
| MESSAGE LENGTH DISTRIBUTION | ← EXPONENTIAL → | | | | | | | | | | | | | |
| RETRIAL INTERVAL | - | - | - | - | - | - | - | 0.35 | 0.35 | 0.1 | 0.35 | 0.35 | 0.35 | 0.35 |

[1] MODEL 2
[2] MODEL 3

Figure 3-1. Sample Generation and Processing Rate of the STARCON Model with a CDC 1604 Machine for Different Values of Input Parameters

trunk group,[1] and to the retrial interval with circuit switching. It is important to realize that the actual degree of sensitivity of the processing rate to a given input parameter (i.e., the partial derivative of the processing rate with respect to the given input parameter) is a function of the values of all of the input parameters cited above (including itself) and, of course, the particular model being investigated. For the cases exhibited in Figure 3-1 (which considers only a fairly limited range for most input variables), it is seen that the processing rate is spread over two orders of magnitude ranging from twenty-eight hundredths of a minute per thousand delay observations with case M3 to twenty-three minutes per thousand delay observations with case C6.

The processing rate given by Figure 3-1, the speed of convergence, and the adopted accuracy criterion all affected the production time consumed by each case simulated in Section 2.0. It was discovered that in many instances when cases ran slowly, convergence also occurred at a reduced rate causing the production cost to yield an accurate delay estimate to become prohibitive. The phenomenon precluded the investigation of a multi-node configuration with traffic loading in excess of 0.9.

---

[1] Running time, in general, goes up as the number of channels per group is decreased (channel loading remaining constant); however, exceptions arose in cases M5 and M6.

## 4.0 ON THE VALIDITY OF A CLASS OF ANALYTICAL APPROXIMATIONS

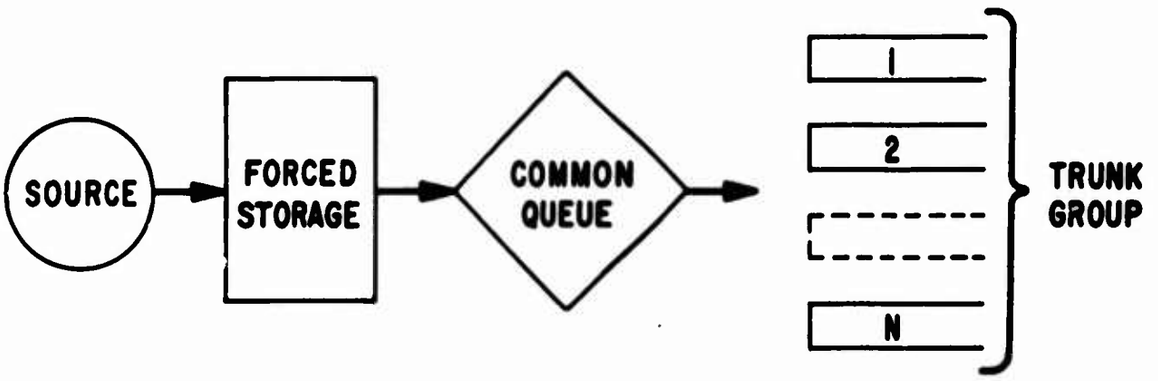
The single stage queueing problem with a Poisson (infinite) source, a finite nu ıber of parallel exponential servers, and an arrival ordered queue discipline has a well known solution for the steady-state delay (in the queue) distribution. Since Burke [3] has shown that this system exhibits a Poisson output (i.e., discharge rate), it has been possible to investigate analytically a multi-stage system by merely summing individual single stage average delays and by convolving all of the tandem waiting time and service time distributions to obtain the overall system delay distribution.

Nevertheless, it should be stressed that in order for the familiar expression for the waiting time distribution to be applicable for the single stage system, the request epochs (i.e., the message arrivals) and service (i.e., message hold time) distributions must be statistically independent. This implicit condition is subtly violated in store-and-forward communication systems because an individual message presents an unchanging hold time (i.e., message length) to each tandem trunk through which it passes. The amount by which this phenomenon alters the average and ninety percentile end-to-end message delay from what would be experienced with independent source and service distributions depends on, among other factors, the configuration topology and the traffic loading matrix.[1] The general analytical solution for this situation has not as yet been obtained, and simulation is required to determine whether the independent assumption leads to a good approximation for the average delay and for the upper tail of the delay distribution under any given set of topological and traffic loading circumstances.

---

[1] The effect of this dependence on the average delay within a single network internal stage is examined by Leonard Klienrock, [12], Chapter 3.

Even with a single node store-and-forward switching configuration, moreover, the effects of the distribution interdependence are not avoided but may be felt to a very startling degree. Consider the configuration depicted below:




This is the situation simulated by Model Two when the single node case is investigated. Messages with an exponential hold time are generated by a Poisson Source. All of these messages enter a single switching center where they must be stored until the entire message is received. Thus there may be a period during which part of the message is stored even though idle channels exist. After the message is completely received by the switching center, it is offered to an idle trunk channel if one exists; otherwise it enters a first come, first served common queue which feeds the trunk group. One is interested in the average message delay from the time the message emerges from the source until it initially enters an outgoing trunk and is begun to be received at the distant end.

The forced storage stage is equivalent to an infinite parallel server system and it has been shown by Mirasol [ 13 ] that the output of such a system, when driven by a Poisson Source is also Poisson. Thus, it would immediately appear plausible to analyze both stages separately, namely a M/M/$\infty$ followed by the familiar M/M/N (classification nomenclature per D. G. Kendall)[1], and combine the results by elementary techniques (i. e., the sum of the average delays and the convolution of the delay distributions).

If we consider a ten channel trunk group, 0.9 channel loading, and (for convenience) an average hold time of unity, one is tempted to assume not only that the average delay encountered in the forced storage stage is one (since all messages are delayed a time period equal to their message length), but that the average delay in the queueing stage is 0.67 (see Descloux [ 8 ] ). In actuality, because of the statistical dependence of the distributions in the queueing stage, the average delay there, determined by simulation, is only 0.084. Thus, the analytical approximation introduces an error of 69.89%. This result should serve to underscore the potential danger associated with a non-rigorous application of analytical techniques to deceptively tractable queueing situations.[2] It may be fair to conclude that, in the absence of rigorous mathematical treatment, simulation is frequently the only alternative which offers a credible solution. Analytical approximations which offer no reliable means to bound their accuracy may produce misleading results.

---

[1] See [ 18 ] p. 25; here the first M signifies a Poisson source, the second M an exponential service time, and N refers to the number of parallel servers in a single stage queueing system.

[2] In this case, however, the dependency disappears if a constant hold time is considered.

## 5.0 ON THE ATTAINMENT OF STATISTICAL CONVERGENCE

In order for the results from a digital simulation to be meaningful, the effects on these results of statistical fluctuations occurring within a Monte Carlo model should be negligible. Stated another way, if different series of random numbers were to be used for a simulation run (each series, however, obeying the same probabilistic laws), the simulation output should not be materially affected.

With STARCON, we are interested in obtaining values for certain population parameters. Two such parameters are the average message delay and the ninety percentile delay. These population parameters were estimated by using, respectively, the average and ninety percentile of a sample which is composed of delay observations generated by the simulation.[1] The larger the sample, the greater the chance that such an estimate is within a given proximity of the actual, steady-state, population value.

Once the desired statistical accuracy of the estimate is defined, two factors come into play (given the process is stationary) in determining how large the sample must be. Assuming that one is interested in obtaining an estimate of the mean delay which is within a stipulated percentage of the true value, one of these factors becomes simply the ratio of the steady-state standard deviation for a single delay observation and the steady-state mean, $\sigma_x/\mu$. If all of the individual observations in the sample are independent (or at least negligibly correlated), no further knowledge of the model's properties is required to compute the appropriate sample size. For in such a case (1) employing the concept that the standard deviation of the sample average is diminished by the square root of the number of observations in the sample, n, and (2) invoking the Central Limit Theorem

---

[1] These are not the best theoretical estimators although the most practical; see [12], p. 20.

allows one to immediately compute the sample size needed to satisfy a given statistical accuracy criterion. Nevertheless, if the individual observations are correlated, knowledge of the serial correlation coefficients as a function of separation, $\rho_k$, is necessary.[1] In this case, the standard deviation of the sample average becomes of the form:

$$\sigma_{\bar{x}} = \frac{\sigma_x}{n^{1/2}} \left\{ 1 + \left[ 2 \sum_{k=1}^{n-1} \left(1 + \frac{k}{n}\right) \rho_k \right] \right\}^{1/2}$$

which, for large n, asympotically becomes equal to

$$\sigma_{\bar{x}} = \sigma_x n^{-1/2} \left\{ 1 + 2 \sum_{k=1}^{\infty} \rho_k \right\}^{1/2}$$

Thus, when large positive values of $\rho_k$ are encountered which do not decay rapidly with k, convergence will be considerably slower than in the independent situation. Moreover, with certain values of input parameters for STARCON (especially large values of channel loading), convergence occurred so slowly as to render intensive investigation economically infeasible.

An initial search by the author failed to disclose an abundance of literature devoted to evaluating $\sigma_{\bar{x}}$ in queueing situations or even to determining the $\rho_k$'s of successive delays for the most fundamental type of queueing situation; i.e., the M/M/1 case[2] with an ordered queue. Benes [2] has obtained the covariance function (in continuous time) of the "virtual" waiting time for a queue with

---

[1] Separation, k, is the sequential integer distance between pairs of waiting time observations; for consecutive waiting times, the distance is 1, etc...

[2] D. G. Kendall's classification scheme; see Saaty [18], p. 25.

Poisson arrivals, but this is not strictly applicable to our situation. For successive delays associated with the M/M/1 case, Craven [7] has evaluated the $\rho_k$'s for k = 1, 2, and 3 for limiting values of the channel load factor, $\rho$. Nevertheless for larger lag values Craven has (in the cited reference) only conjectured at the form which the correlation coefficients might take.[1] Thus, to the author's knowledge, even in this elementary M/M/1 case, computations have not been carried out and published in the past that reveal how large a sample must be so that, for example, exactly (or at least) ninety-five percent of the time the sample average delay will fall within ten percent of the true average delay (given that the sampling is started in the steady-state).

One could therefore benefit by examining some empirical data. For this purpose, consider a Poisson source feeding two parallel, exponential servers with a first come, first served queue discipline, and a server loading of 0.95. To establish a reference sample size for a simulation, let us consider what is required to attain a ninety-five percent chance that the sample average will lie within five percent of the true average under the unrealistic premise that observations are independent. Then the solution of the following equations is in order,

$$\frac{\sigma_x}{\mu} = \sqrt{\frac{2}{P(0)} - 1} \qquad \text{for M/M/R}$$

$$.05\mu = 1.96\sigma_s \qquad \text{statistical accuracy criterion}$$

$$\sigma_s = \frac{\sigma_x}{\sqrt{n}} \qquad \text{convergence of sample mean's standard deviation with independent observations}$$

[1] Immediately prior to publication, the author was furnished with some very pertinent results concerning the M/M/1 case by Daryl J. Daley (who is currently at the Statistical Laboratory of the University of Cambridge). These results are used to compute required sample sizes for different system loads in Appendix A-II, Part II.

where:

P(0) is the probability of a delay

$\sigma_s$ is the standard deviation of the sample average

n is the sample size

and this leads to a required sample size depicted in Figure 5-1. For the case being considered P(0) is 0.9256, and n becomes 1784. Computer runs each consisting of 2000 observations were therefore carried out with STARCON. In striking contrast to the computation above, out of thirteen runs made (each with different random number series), only one satisfied the stipulated accuracy goal and only three, or less than one quarter, yielded a sample average which was within ten percent of the true mean. When the results of five runs, consisting of 10,000 observations apiece, were examined, four had errors in excess of five percent, two had errors in excess of ten percent, and the root mean square error was 21%. In fact, when two computer runs were made with a sample size of 50,000 observations, one run had an error of 33%.[1] The probability of obtaining this (or greater) magnitude error, if correlation did not play a significant role, is (assuming perfect normality for the sample average) smaller than $2.5 \times 10^{-784}$. Finally, seven computer runs were made consisting of 100,000 observations apiece and individually consuming about thirty-five minutes on the CDC 1604. 1.96 times the root-mean-square error turned out to be 32% of the true mean. Two out of the seven runs produced errors in estimating the population mean that exceeded 20%. Errors in the mean estimates of four out of the seven runs were greater than 10%. When the estimates of the ninety percentile delays were examined, five out of the seven runs had errors in excess of ten percent.

---

[1] The error in the other was 7%.

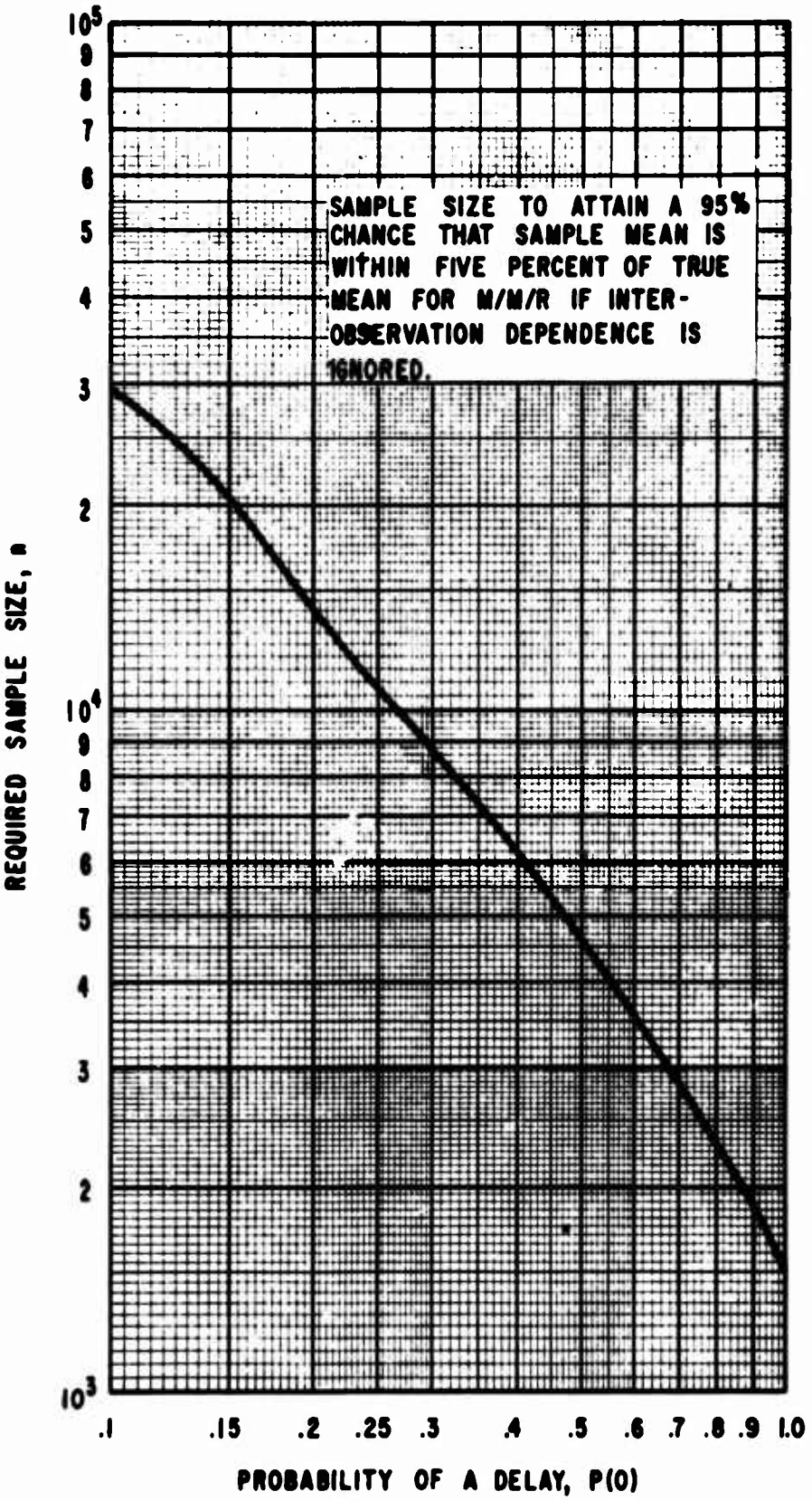



Figure 5-1. Reference Sample Size

To summarize, we have seen that even for the commonest of queueing problems[1] at the present time apparently no analytical computations have been carried out and published which reveal the accuracy with which the sample mean estimates the true steady-state average customer delay as a function of the sample size and the server facility loading. Theoretical approximations of this accuracy, which neglect interobservation dependence for large server loading seem to suffer from orders of magnitude error, based on empirical data provided by exercising the STARCON model. With heavy facility loading and an unlimited queue, the rate of convergence of the sample mean is drastically reduced; this raises the serious question whether, in these circumstances, simulation is a practical tool unless only the crudest of estimates is desired.[2]

When convergence occurs extremely slowly, this fact in itself has real world implications, however. For then the meaningfulness of the steady-state solution is open to question. The time to attain equilibrium may not be short when compared to the span of operational interest and may not occur in practice before significant changes to traffic and network parameters are experienced.

---

1 Poisson source, single exponential server, strict queue discipline.

2 Of course, if the predicted traffic patterns themselves suffer from many uncertainties, this added source of uncertainty from statistical fluctuations may not be significant.

## 6.0 ON THE MEASUREMENT OF THE STATISTICAL ACCURACY OF MONTE CARLO SIMULATION RESULTS

One frequently turns to Monte Carlo simulation when appropriate analytical techniques to describe relationships of interest in complex stochastic processes do not exist. Certain parameters of a parent population (mean and fractile delays, in our case) are often sufficient to summarize the processes performance. A simulation permits one to estimate the value of these parameters by subjecting a sample of observations, created during the simulation, to statistical analysis.

The accuracy of these estimates is a strong function of the size of the sample. While unnecessarily large samples[1] tend to produce very reliable results, they often require an oppressive investment in production time on a large scale computer. Thus, in general, it is a good practice to evaluate the penalties associated with errors of different magnitudes, establish an error criterion which is tolerable,[2] and then insure that the run length used yields an estimate that satisfies this criterion.

Thus, it behooves us to have at our disposal means to measure the accuracy of simulation results. Unfortunately owing to the intrinsic nature of stochastic processes, no reasonable inviolable bound can be computed for the error contained in the simulation result regardless of the sample size. Rather one must be content with an expression implying the "maximum likely" error, such as a confidence interval. The validity, methods of computation, and other properties of some of the accuracy indices which have been proposed in the past are discussed in this section.

---

[1] Of course in many situations, intuition fails to offer a reliable yardstick as to what length constitutes a large sample.

[2] Concepts and techniques developed by Schlaifer [19] may be helpful in this endeavor.

## 6.1 Desired Attributes of an Accuracy Index

An unbiased (or negligibly biased) index is required if the index is to provide any useful information. Let us therefore examine the meaning of bias in terms of exact confidence intervals. The reader will recall that a confidence interval is a set of statements (or methods to compute same) which are on hand at the outset of a stochastic experiment. At the conclusion of the experiment the appropriate member of the set is selected. A typical statement might read, "The parent population average delay is within the range from 2.3 to 2.9 average hold times." If ninety-five percent confidence intervals are being used, if the experiment is repeated indefinitely, and if an appropriate interval is selected for each experiment, ninety-five percent of the statements will be correct. If, however, the method of computation of the bounds leads to a biased estimate, some other fraction, which is unknown,[1] of the statements will be true. Examples of seriously biased confidence intervals will be discussed subsequently.

A second property which is desired of the accuracy index is that it be efficient, or in other words, that it takes advantage of all the information in the sample to compute the degree of accuracy achieved in a given estimate. Certain procedures for computing confidence bounds to measure the statistical accuracy in estimates obtained from a sample of size n, entail generating from ten to thirty independent[2] samples, each containing n observations, and observing the variation among the estimates produced by each sample. Such

---

[1] This assumes that if the bias were known, a new statistic would be employed that eliminates it; on the other hand, frequently the direction of the bias is known, although its magnitude is not.

[2] Recording of sample values should commence only after equilibrium is attained for each sample if steady-state population parameters are desired with the procedures.

procedures thus measure the statistical accuracy of estimates which are each based on only a small fraction of the total sample information available. Nevertheless, when one is interested in estimating a population parameter such as the steady-state mean, it is possible to measure statistical accuracy with just the single sample from which an estimate is generated. An illustration of what is meant by efficiency in this latter situation is offered in the following paragraphs.

Let us hypothesize that two different methods exist to compute unbiased ninety-five percent confidence intervals for a sample of 2000 observations where the true population mean is 2.5. Let us further repeat the experiment, each consisting of 2000 trials, ten times and theorize that the intervals produced are as shown in Table 6-1. One cannot help but note from glancing at this table that the volatility in the size of the interval computed by method 1 greatly exceeds that encountered with method 2. As a result of this, the size of the interval produced from a single simulation run by method 1 provides very little information about the potential error in an estimate from this length sample. For samples 5, 7, and 8, one is unduly alarmed by the possible error in the estimate. The interval for samples 2 and 10, on the other hand, yields an estimate which gives the impression of greater accuracy than is actually present.

With method 2, as previously noted, the volatility of the size of the interval is considerably reduced. This reduction occurs when the method used to construct the bound is more efficient. If the standard deviation of the interval is small relative to its mean magnitude, it generally provides a faithful representation of the potential error in an estimate obtained from the sample size investigated. Hence, when the sample is sufficiently large to produce a population parameter estimate that will satisfy the established accuracy criterion, it is desirable to use a method to compute a confidence bound whose efficiency is great enough so that the interval's standard deviation to mean ratio (i.e., the coefficient of variation) is small.

TABLE 6-1

HYPOTHETICAL UNBIASED 95% CONFIDENCE INTERVALS
(TRUE POPULATION PARAMETER IS 2.5)

| SAMPLE NUMBER | SAMPLE ESTIMATE | METHOD 1 INTERVAL | SIZE | VERACITY | METHOD 2 INTERVAL | SIZE | VERACITY |
|---|---|---|---|---|---|---|---|
| 1 | 2.5 | 2.0 - 3.0 | 1.0 | TRUE | 2.0 - 3.0 | 1.0 | TRUE |
| 2 | 3.2 | 3.1 - 3.3 | 0.2 | FALSE | 2.6 - 3.7 | 1.1 | FALSE |
| 3 | 2.8 | 1.8 - 3.8 | 1.0 | TRUE | 2.4 - 3.2 | 0.8 | TRUE |
| 4 | 2.2 | 1.9 - 2.5 | 0.6 | TRUE | 1.7 - 2.7 | 1.0 | TRUE |
| 5 | 2.6 | 1.4 - 3.8 | 2.4 | TRUE | 2.1 - 3.1 | 1.0 | TRUE |
| 6 | 2.1 | 1.7 - 2.5 | 0.8 | TRUE | 1.7 - 2.5 | 0.8 | TRUE |
| 7 | 2.3 | 0.7 - 3.9 | 3.2 | TRUE | 1.8 - 2.8 | 1.0 | TRUE |
| 8 | 2.7 | 1.4 - 4.0 | 2.6 | TRUE | 2.1 - 3.3 | 1.2 | TRUE |
| 9 | 3.0 | 2.5 - 3.5 | 0.5 | TRUE | 2.5 - 3.5 | 1.0 | TRUE |
| 10 | 2.4 | 2.3 - 2.5 | 0.2 | TRUE | 2.0 - 2.8 | 0.8 | TRUE |

## 6.2 A Critical Review of Some Common Procedures to Estimate Accuracy

One approach, frequently used by simulation practitioners, involves the continuing surveillance of the simulation result, while the sample is increased by fixed increments, until negligible variation in its value is experienced. The strength of this method lies in its simplicity and the fact that it is not limited to measuring the accuracy in the population mean estimate. Its weakness, on the other hand, stems from the fact that it has little statistical validity. In most instances, the monitored variation will tend to diminish as the increment length becomes a smaller and smaller fraction of the total sample, regardless of whether sufficient convergence is even attained.

A method frequently employed to assess the potential likely error in using the sample average to estimate the population mean proposes computing the statistic $s/\sqrt{n}$, where:

$$s^2 = \frac{\sum_{k=1}^{n} (x_k - \bar{x})^2}{n-1}$$

$$\bar{x} = \frac{1}{n} \sum_{k=1}^{n} x_k$$

n is the sample size,

and creating a confidence interval that assumes the sample average is normally distributed and uses $s/\sqrt{n}$ as an estimate of the sample average standard deviation. This method produces an efficient, unbiased estimate providing the individual observations $x_k$, are statistically independent. Nevertheless, if correlation among observations is prevalent, a seriously biased confidence interval can result. Consider the case when the serial correlation coefficient, $\rho_k$, decreases geometrically, and thus:

$$\rho_k = \rho_1^k \quad \text{for} \quad k \geq 1.$$

If the sample size, n, is large relative to the k which reduces $\rho_k$ to negligible proportions, an unbiased estimate of the standard deviation of the sample average as a function of $\rho_1$ becomes $s_g \sqrt{n}$, where:

$$s_g = s\sqrt{\frac{1 + \rho_1}{1 - \rho_1}}$$

Thus for $\rho_1 = 0.5$ a bias of a factor of 1.73 is introduced by using $s/\sqrt{n}$ which in turn causes a nominal ninety-five percent interval in actuality to yield only a seventy-four percent level of confidence. Stated another way, instead of having one out of twenty (on the average) generated statements in error, one out of four will be false. With $\rho_1 = 0.99$, the nominal ninety-five percent interval produces a situation where nine out of ten statements will be false.

To remove this bias, the practice has developed to take the entire sample and break it up into groups, called blocks, composed of consecutive sample observations. Averages for each block are computed and used as if they were members of a new smaller sample. This will achieve a marked reduction in the amount of serial correlations encountered.[1] To estimate the standard deviation of the sample average, $s'/\sqrt{n}$ is used, where s' is the equivalent statistic to s for the new synthesized sample of block averages.

The use of the block averaging technique, while tending to remove bias, increases the volatility of the interval. Consider the situation where there is no correlation in the sample. If the individual sample members are normally distributed, the ratio of the standard deviation to the mean of the sample standard

---

[1] Even greater bias removal can be achieved, moreover, of only the averages of every other block are used to compute the bound.

deviation becomes equal to $\sqrt{\frac{2}{n-1}}$, where n is the sample size that is used to compute the interval.[1] Thus the volatility of the interval (for independent observations) increases roughly as the square root of the number of observations in each block. If we desire that two standard deviations are not more than ten percent of the interval's mean, we note that the number of blocks must be at least 800. If only thirty blocks are used, two sigma[2] amounts to over a fifty percent departure from the mean, which represents an interval range of over a factor of three. If only thirty blocks are used and the block averages have a distribution which resembles a chi-square with six degrees of freedom, two sigma[2] amounts to over a seventy-five percent departure from the mean or an interval range of over a factor of seven. With thirty blocks and exponentially distributed block averages, two sigma[2] amounts to over one hundred percent of the mean. If the block averages exhibit some residual correlation, even greater volatility is experienced.

Now when members of the original sample generated by the simulation are correlated (the condition which causes us to adopt the block averaging technique in the first place), it is true that the interval volatility increases by less than the square root of the number of observations in each block. Nevertheless, using solely block averages to rid ourselves of interval bias, creates, unnecessarily, a confidence interval which can tend to exhibit erratic behavior.

A second drawback associated with using the block average technique arises since it does not protect per se against the possibility that significant residual correlation exists in the block averages. Unless ample

---

[1] See Goldberger [ 9] , p. 99, Eq. 4.33; and Cramer [ 6] , p. 213, Eq. 17.2.3 [

[2] Since the distribution of s' in these cases is far from symmetric, creating a bound by departing equal distances from the mean is somewhat fictitious, and the reader is called on to make appropriate adjustments.

evidence exists which demonstrates that this is not the case, reliable statistical tests should be used to provide such a guarantee.

The technique adopted by the STARCON model to generate confidence intervals (refer to Part I, paragraph 3.5.1) attempted to partially overcome these two deficiencies. Nevertheless, owing to volatility in the first sample autocovariance,[1] the volatility in the final interval rendered it unsatisfactory. Also, some bias was introduced through factors noted in the referenced paragraph of Part I.

It was therefore considered of great value, by the author, to have available a technique to produce a negligibly biased confidence interval whose efficiency significantly exceeds that associated with the block average method. One technique which reduces variance has been offered by Edward J. Hannan[2] and is reproduced below.

To estimate the variance of the mean of a large[3] sample (size n) taken from a stationary time series $x_k$, consider the spectral density function:

$$f(\theta) = \frac{\sigma_x^2}{2\pi} \sum_{-\infty}^{+\infty} \rho_j e^{ij\theta}$$

The variance can be approximated by $2\pi f(0)/n$.

---

[1] Explanation for this phenomenon is provided in part in Goldberger [9] by equations 8.18 (p. 148), 8.26 (p. 150), 8.27 (p. 151), 8.29 (p. 152) using the moments for a normal distribution.

[2] Department of Statistics, Australian National University, Canberra; the method employs simplified Hamming weight function; see Hannan 11 , p. 61.

[3] Relative to the autocorrelation decay as well as absolutely. For smaller samples, one might consider the method proposed by equation (2) in Hannan [10].

A common procedure would then entail forming:

$$s_j = \sum_{k=0}^{n-j} (x_k - \bar{x})(x_{k+j} - \bar{x})$$

where:

$\bar{x}$ is the sample average

and then[1]

$$\hat{\sigma}_{\bar{x}}^2 = 2\pi \hat{f}(0)/n = n^{-2} \left[ s_o + \sum_{j=1}^{m-1} \left(1 + \cos \frac{\pi j}{m}\right) s_j \right]$$

Here $m$ must be chosen so that $\rho_j$ for all $j > m$ are negligible.[2] Also, $m << n$. If the conditions for $m$ and $n$ are satisfied, the confidence interval using $\hat{\sigma}_{\bar{x}}$ will have a standard deviation to mean ratio which is not appreciable.[3]

It is obvious that the computational burden is significantly increased when using this technique. Nevertheless, the effort involved in carrying out these numerical manipulations may easily be small when compared to that consumed by the simulation itself. In turn, one derives a more reliable accuracy index.

---

1 For alternative weighting functions to $\frac{1}{2}\left(1 + \cos \frac{\pi j}{m}\right)$, the reader is referred to pages 60-63 in Hannan [9]. The method of Daniell is the most efficient but the most arduous to compute. Also, one has the option of trying a second autoregression fit and proceeding according to the second and third paragraphs on page 74 of the same reference that begin, "Of course." The bias introduced by these various methods differ and are in general very sensitive to the shape of the spectral density near the origin frequencies.

2 One can put a lower bound on an acceptable value for m, by examining the behavior of the spectral density estimate at frequencies close to zero.

3 The coefficient of variation is roughly $\sqrt{\frac{3m}{4n}}$.

If the sample produced by the simulation contains more than one thousand observations, a compromise between the two techniques will probably prove satisfactory in most situations. The original sample can be transformed into a new sample consisting of approximately one thousand block averages. The above technique can then be applied to determine the confidence bound.

To review, we have seen that the sample variance can be used to compute a good estimate of the statistical accuracy with which the sample average approximates the equilibrium population mean if sample observations are uncorrelated. In this case, the larger the sample is, the less volatile the confidence interval will be. If, however, observations are significantly correlated, the sample variance will generally yield a seriously biased bound. If the correlation falls off very rapidly relative to the size of the sample and the sample is large, the block average technique may suffice to remove the majority of the bias and yet produce a fairly stable bound. Nevertheless, if the correlation is not so rapid, or the sample is not sufficiently large, one is generally forced to use a more powerful technique requiring more data processing if an informative confidence interval is to be created. Nevertheless, if correlation decay is slow relative to the size of the sample, in general no technique can produce an unbiased, stable interval since the information contained in the sample is inadequate. In all situations it is implicitly assumed that the sample comprises the sole source of information about the observed time series.[1]

### 6.3 Results from Using the Block Average Technique and the Procedure Incorporated in the STARCON Model in One Situation

The STARCON Model was used to simulate a one stage queueing process characterized by a single Poisson source, two parallel exponential

---

[1] For example, if the serial correlation coefficients are known to be all positive, monotonically decreasing, and eventually falling off geometrically, more informative accuracy bounds could be extracted from the sample than when more general autocorrelation functions are encountered.

servers, and a first come, first served queue discipline. With a server utilization of 0.95, thirteen samples (all starting in the null state but each using different random number sequences) were generated. The bias and volatility of the nominal ninety-five percent confidence intervals produced by each technique were examined. For both techniques, individual samples were subdivided into twenty blocks, each consisting of one hundred observations.

The results are shown in Table 6-2. The volatility of the confidence interval is expressed as twice the ratio of the root-mean-square deviation from the mean to the mean. A similar statistic is shown for the volatility of the block average first autocovariance estimate which is used in the computation of the STARCON bound.[1]

TABLE 6-2

ANALYSIS OF NOMINAL 95% CONFIDENCE INTERVAL PRODUCED FROM THIRTEEN SAMPLES EACH CONTAINING TWO THOUSAND OBSERVATIONS

| | Intervals from Block Average Technique | Intervals from STARCON Method | Adjacent Block Sample Correlation |
|---|---|---|---|
| STATEMENTS CORRECT | 31% | 62% | - |
| VOLATILITY INDEX | 73% | 127% | 66% |
| AVERAGE VALUE | - | - | .60 |

1 Refer to STARCON bound discussion in subsection 6.2, page 81.

(This page left intentionally blank)

## APPENDIX A-II

## REQUIRED SAMPLE SIZE FOR M/M/1 CASE

This Appendix presents certain results related to the convergence of the sample mean to the steady-state expected value when Monte Carlo simulation is used to generate successive delays in the queue for the M/M/1 situation, assuming monitoring begins in the steady-state. This is the situation investigated by STARCON when the model number, the node number, and the Erlangian phase number are each set to one and the transient load factor is set to zero.

The author is deeply indebted to Daryl J. Daley for furnishing recently developed expressions to evaluate the serial correlation coefficients in M/M/1 case, and in particular for a compact formula yielding the sum of the coefficients.

By using the z-transform (i.e., the generation function), it has been determined that

$$1 + 2 \sum_{k=1}^{\infty} \rho_k = \frac{4}{(2 - \rho)(1 - \rho)^2} - 1$$

Thus, to determine how large a sample must be so that with ninety-five percent probability the sample average delay will fall within five percent of the true average delay, solution of the following equations is in order:

$$\frac{\sigma_x}{\mu} = \left(\frac{2}{P(0)} - 1\right)^{1/2} \qquad \text{for M/M/R}$$

$$P(0) = \rho$$ for M/M/1

$$0.05\mu \doteq 1.96\sigma_{\bar{x}}$$ statistical accuracy criterion assuming sample mean is normally distributed

$$\sigma_{\bar{x}} \doteq \sigma_x n^{-1/2} \left[ \frac{4}{(2 - 1)(1 - \rho)^2} - 1 \right]^{1/2}$$

convergence of sample mean's standard deviation for M/M/1 given n is large

where

- $P(0)$ is the probability of a delay
- $\rho$ is the server load factor
- $n$ is the sample size
- $\mu$ is the expected steady-state delay
- $\sigma_x$ is the standard deviation of a single delay observation
- $\sigma_{\bar{x}}$ is the standard deviation of the sample mean

This leads to the values of n as a function of $\rho$ shown in the table below.

SAMPLE SIZE TO ATTAIN A 95% CHANCE THAT SAMPLE MEAN IS WITHIN FIVE PERCENT OF TRUE MEAN FOR M/M/1

| $\rho$ | n |
|---|---|
| 0.5 | $44.6 \times 10^3$ |
| 0.9 | $681 \times 10^3$ |
| 0.95 | $2.59 \times 10^6$ |
| 0.99 | $62.1 \times 10^6$ |

These results should cast some light on the economy of using Monte Carlo simulation in this and similar situations. In particular, extrapolating from Figure 3-1, Part II, reveals that even for the single node, single channel queueing configuration, a load factor of 0.95 would cause STARCON to require in excess of twenty hours on a CDC 1604 to meet the moderate statistical accuracy criterion considered in this appendix.

## NOMENCLATURE INDEX

Presented below is an alphabetically ordered list of terms used throughout the report with references to where these terms appear and are described in the text. This tabulation is provided for the reader's convenience.

## BIBLIOGRAPHY


1. Anderson, R. L., Distribution of the Serial Correlation Coefficient, Annals of Mathematical Statistics, XIII, 1 (1942).

2. Benes, V. E., On Queues with Poisson Arrivals, Ann. Math. Statist, 28 (1957) 670-77.

3. Burke, P. J., The Output of a Queueing System, Opns. Res., 4, (1956) 699-704.

4. Burington and May, Handbook of Probability and Statistics, Sandusky, Handbook Publishers (1958).

5. Coveyou, R. R., Serial Correlation in the Generation of Pseudo-Random Numbers, J. Assoc. Comp. Mach., 6, (1959) 72-74.

6. Cramer, H., Mathematical Methods of Statistics, Princeton University Press (1961).

7. Craven, B. D., Serial Dependence of a Markov Process, Aust. Math, Soc. Jnl., 5, pt. 3 (1965) 299-314.

8. Descloux, A., Delay Tables, New York, McGraw-Hill (1962).

9. Goldberger, A. S., Economic Theory, New York, Wiley & Sons (1964).

10. Hannan, E. J., The Variance of the Mean of a Stationary Process, Royal Statistical Society Journal, Series B., Vol. 19-20 (1957-8).

11. Hannan, E. J., Time Series Analysis, New York, Wiley & Sons (Methuen Monograph) 1962.

12. Kleinrock, L., Communication Nets, New York, McGraw-Hill (1964).

13. Mirasol, N. M., The Output of an M/G/$\infty$, Queueing System is Poisson, Opns. Res., 11, March-April (1963), 282-284.

14 Moran, P. A. P., The Theory of Storage, New York, Wiley & Sons (Methuen Monograph) (1959).

BIBLIOGRAPHY (Cont'd.)

15. Morse, P. M., Queues, Inventories & Maintenance, New York, Wiley & Sons (1958).

16. Owen, D. B., Handbook of Statistical Tables, Palo Alto, Addison-Wesley (1962).

17. Rotenberg, A., A New Pseudo-Random Number Generator, Journal of the ACM, 7, (1960) 75-77.

18. Saaty, T. L., Elements of Queueing Theory, New York, McGraw-Hill (1961).

19. Schlaifer, R., Probability and Statistics for Business Decisions, New York, McGraw-Hill (1959).

20. Syski, R., Congestion Theory in Telephone Systems, London, Oliver & Boyd (1960).

Unclassified

Security Classification

## DOCUMENT CONTROL DATA - R & D

*(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)*

1. ORIGINATING ACTIVITY *(Corporate author)*
The MITRE Corporation
5600 Columbia Pike
Bailey's Crossroads, Virginia

2a. REPORT SECURITY CLASSIFICATION
Unclassified

2b. GROUP
N/A

3. REPORT TITLE
A Preliminary Cost/Effectiveness Comparison of Message and Circuit Switching in Military Communications Networks Servicing Record Traffic

4. DESCRIPTIVE NOTES *(Type of report and inclusive dates)*
Summary Dates N/A

5. AUTHOR(S) *(First name, middle initial, last name)*
Stephen A. Sobel

| 6. REPORT DATE | 7a. TOTAL NO. OF PAGES | 7b. NO. OF REFS |
|---|---|---|
| June 1966 | 106 | 20 |

| 8a. CONTRACT OR GRANT NO. | 9a. ORIGINATOR'S REPORT NUMBER(S) |
|---|---|
| AF 19(628)5165 | |
| b. PROJECT NO. | MTP-28-A |
| c. 3500 | 9b. OTHER REPORT NO(S) *(Any other numbers that may be assigned this report)* |
| d. | None |

10. DISTRIBUTION STATEMENT
Distribution of this document is unlimited

| 11. SUPPLEMENTARY NOTES | 12. SPONSORING MILITARY ACTIVITY |
|---|---|
| N/A | Defense Communications Agency<br>NMCSTS Directorate<br>Arlington, Virginia |

13. ABSTRACT

An analysis is contained herein that compares circuit with message switching in Military Communications Networks that carry record traffic. Also examined are the effects of system parameters on the attainment and measurement of statistical convergence when Monte Carlo simulation is used to study delays caused by traffic congestion in communication networks. (U)

DD FORM 1 NOV 65 1473

Unclassified

Security Classification

| 14. KEY WORDS | LINK A | | LINK B | | LINK C | |
|---|---|---|---|---|---|---|
| | ROLE | WT | ROLE | WT | ROLE | WT |
| Monte Carlo | | | | | | |
| Digital Simulation | | | | | | |
| Traffic Congestion | | | | | | |
| Communication Networks | | | | | | |
| Switched Networks | | | | | | |
| Switching Techniques | | | | | | |
| Traffic Delays | | | | | | |
| Statistical Convergance | | | | | | |
| Message and Circuit Switching | | | | | | |
| Military Communications Networks | | | | | | |