ESD-TR-69-362

MTR 952

Extra File Copy

ESD ACCESSION LIST
Call No. AL 67705
Copy No. 1 of 2 cys.

# THE NUMBER OF MESSAGES THAT CAN BE STORED ON A DRUM

R. A. Gildea and D. F. Votaw, Jr.

NOVEMBER 1969

Prepared for

DEPUTY FOR COMMUNICATIONS SYSTEMS
ELECTRONIC SYSTEMS DIVISION
AIR FORCE SYSTEMS COMMAND
UNITED STATES AIR FORCE
L. G. Hanscom Field, Bedford, Massachusetts

ESD RECORD COPY
RETURN TO
SCIENTIFIC & TECHNICAL INFORMATION DIVISION
(ESTI), BUILDING 1211

This document has been approved for public release and sale; its distribution is unlimited.

Prepared by
THE MITRE CORPORATION
Bedford, Massachusetts
Contract AF19(628)-68-C-0365

AD0698762

When U.S. Government drawings, specifications, or other data are used for any purpose other than a definitely related government procurement operation, the government thereby incurs no responsibility nor any obligation whatsoever; and the fact that the government may have formulated, furnished, or in any way supplied the said drawings, specifications, or other data is not to be regarded by implication or otherwise, as in any manner licensing the holder or any other person or corporation, or conveying any rights or permission to manufacture, use, or sell any patented invention that may in any way be related thereto.

Do not return this copy. Retain or destroy.

ESD-TR-69-362

MTR 952

# THE NUMBER OF MESSAGES THAT CAN BE STORED ON A DRUM

R. A. Gildea and D. F. Votaw, Jr.

NOVEMBER 1969

Prepared for

DEPUTY FOR COMMUNICATIONS SYSTEMS
ELECTRONIC SYSTEMS DIVISION
AIR FORCE SYSTEMS COMMAND
UNITED STATES AIR FORCE
L. G. Hanscom Field, Bedford, Massachusetts

This document has been approved for public release and sale; its distribution is unlimited.

Prepared by
THE MITRE CORPORATION
Bedford, Massachusetts
Contract AF19(628)-68-C-0365

## FOREWORD

This Technical Report was prepared by The MITRE Corporation, Bedford, Massachusetts, under ESD Contract No. F19628-68-C-0365 for the Automatic Digital Switch System Contract No. AF 19(628)-4936.

## REVIEW AND APPROVAL

This technical report has been reviewed and is approved.

JOSEPH M. DAHM, 1/Lt., USAF
Project Engineer

## ABSTRACT

Messages arriving at a store-and-forward switch are stored on a drum. Each message stored takes up a number of blocks of drum storage that depends on the message length in characters. It is assumed in this report that the arriving messages are drawn at random from a message population; thus the number, say N, of messages that fill the drum is a random variable. It is shown, by means of renewal theory, that the distribution of N is approximately a normal distribution whose mean and variance are simple functions of the drum size and the parameters of the message population. Illustrative examples are given. An estimate of the expected fraction of wasted characters of drum storage is also given.

## TABLE OF CONTENTS


| | | Page |
|---|---|---|
| SECTION I | INTRODUCTION | 1 |
| SECTION II | THE PROBLEM AND ITS SOLUTION | 3 |
| SECTION III | THE 'BRACKETING' POPULATIONS | 7 |
| SECTION IV | NUMERICAL ILLUSTRATIONS | 10 |
| SECTION V | COMPARISON OF RESULTS | 15 |
| SECTION VI | SUPPLEMENTARY REMARKS | 19 |
| APPENDIX | | 21 |
| REFERENCES | | 28 |


## LIST OF ILLUSTRATIONS


| Figure Number | | Page |
|---|---|---|
| 1. | Limiting Values of $E_N$ (1A-27,931) | 9 |
| 2. | Effect of Drum Size on the Expected Number of In-Transit Messages (L = Drum Size) (1A-27,932) | 17 |
| 3. | Effect of Drum Size on the Probability Distribution of the Number of In-Transit Messages (L = Drum Size) (1A-27,930) | 18 |


## LIST OF TABLES


| Table Number | | Page |
|---|---|---|
| I. | Narrative Population | 13 |
| II. | Values of $E_N$ and $E_N \pm 3\,\sigma_N$ for Combinations of Five Message Populations and Two Drum Sizes | 14 |

## SECTION I

## INTRODUCTION

Messages arriving at a store-and-forward message switch are stored on a drum. The number of stored messages -- also called in-transit messages -- is related to the capacity of the drum and to the lengths of the arriving messages. In this report we regard the arriving messages as being drawn at random from a message population. In keeping with this statistical approach, we regard the total number of stored messages as a random variable. What are the principal characteristics of the distribution of this random variable? We shall answer this question analytically and give numerical illustrations of our results.

Our analytical results and the elementary theory underlying them are presented in Section II. The distribution of the number of stored messages is given in Section II for a special case involving a particularly simple choice of drum size and message population. In Section III we consider two extreme message populations which 'bracket' all other message populations considered in this report. In Section IV numerical examples of our results are given for combinations of two different drum sizes and five different message populations. Section V gives a comparison of results; and some supplementary remarks are given in Section VI. The Appendix gives

a derivation of our main results and gives basic formulas used in Sections II - VI.

## SECTION II

## THE PROBLEM AND ITS SOLUTION

Consider a drum having a certain number, say L, of blocks of storage. (For example, L might be 8192.) Each block has a certain number of characters. Let w denote the length (in characters) of an arriving message ($w \geq 1$). Any message stored on the drum takes up an integral number of blocks on the drum.* Let x ($x = 1, 2, \ldots.$) denote the number of blocks required to store a message of length w. The dependence of x on w is described in Appendix Section II. This dependence involves quantities $B_1$, $B_2$, $B_3$, ..., which are the 'ends' (expressed in characters) of the first, second, third, .... blocks of drum storage. It will be noted from the discussion in Appendix Section II that when w is a random variable the quantity x is also a random variable; furthermore, the probability distribution of x can be determined from the probability distribution of w together with a knowledge of $B_1$, $B_2$, ... . Each arriving message is regarded as having been drawn at random from a message population; accordingly, the number of blocks required to store the message is regarded as having been drawn at random from the population of x-values associated with the message population. (It is assumed also that all arriving messages are drawn independently.)

---

*Incidentally, the storage of messages in blocks usually leads to <u>storage waste</u>. For further discussion of this subject see Appendix Section III.

The arriving messages are stored on the drum one after the other. Let the sequence of lengths (in blocks) of these messages be $x_1$, $x_2$, ... . Let N be the largest integer such that

$$x_1 + x_2 + \ldots + x_N \leq L.$$

We call the quantity N the number of messages that 'fill the drum.' In general this number will vary from one 'filling of the drum' to another; and since the quantities $x_1$, $x_2$, ... are regarded as drawn at random from a population, the quantity N is a random variable. Let F(n) be the probability that N is less than or equal to n (n = 0, 1, 2, ...); F(n) is the so-called cumulative distribution function of N. An exact formula for F(n) is given in Appendix Section I. We shall denote the mean and variance of N by $E_N$ and $\sigma^2_N$, respectively.

Let $\mu$ and $\sigma^2$ be the mean and variance of x. If $\sigma$ is finite and L is large, we see from the results stated in the last paragraph of Appendix Section I that $E_N$ is approximately $L/\mu$, $\sigma_N$ is approximately $(\sigma/\mu)\sqrt{L/\mu}$, and that the distribution of N is approximately normal with mean $L/\mu$ and standard deviation $(\sigma/\mu)\sqrt{L/\mu}$.

By way of illustration, suppose the distribution of x is as follows:

$$g(x) = 1/8 \qquad (x = 1, 2, \ldots, 8).$$

(The function g(x) is the probability function of x; thus g(1) is the probability that x = 1, g(2) is the probability that x = 2, etc.).

It will be noted that the function, $g(x) = 1/8$, is uniform over the blocks 1, 2, ..., 8. The mean, $\mu$, of x is

$$\mu = 1(1/8) + 2(1/8) + \ldots + 8(1/8) = 36(1/8) = 4.5 \ .$$

The variance, $\sigma^2$, of x is

$$\sigma^2 = 1^2(1/8) + 2^2(1/8) + \ldots + 8^2(1/8) - \mu^2$$
$$= (204)\ (1/8) - (4.5)^2 = 5.25 \ .$$

Let L be 8192. It is reasonable to regard this value of L as large in relation to $\mu$; thus we shall use the normal approximation mentioned above. We have

$$E_N \doteq (8192)/(4.5) = 1820,$$
$$\sigma^2_N \doteq (1820)\ (5.25)/(4.5)^2 \doteq 472,$$
$$\sigma_N \doteq \sqrt{472} \doteq 21.7 \ .$$

The value, 1820, obtained above for $E_N$ is approximately the 50 percent point (also called the <u>median</u>) of the distribution of N. Roughly speaking, half the time the number of messages that fill the drum will be less than or equal to 1820 and half the time the number will be greater than 1820. Since 1820 is both the mean and the median of the distribution, we have ample reason to regard it as the 'center' of the distribution.

The standard deviation, $\sigma_N$, which was found to be about 22 in our example, is an indicator or measure of the 'spread' (i.e., the variability) of the distribution of N.

The following remarks are intended to explain how $\sigma_N$ relates to the spread of the distribution.

With regard to any normal population, 99.87 percent of the population lies below the point that is three standard deviations above the mean, and .13 percent lies below the point that is three standard deviations below the mean; hence, the percent of the population lying between these two points is 99.74 (= 99.87 - 0.13). Accordingly, about 99.74 percent of the population of N-values lies between the two numbers $E_N - 3\,\sigma_N$ and $E_N + 3\,\sigma_N$. (Note that the length of the interval between these two numbers equals $6\,\sigma_N$; thus, nearly all the population lies in an interval whose length is $6\,\sigma_N$.) With regard to our numerical example we can say that approximately 99.74 percent of the population of N-values lies between 1755 ($\doteq E_N - 3\,\sigma_N$) and 1885 ($\doteq E_N + 3\,\sigma_N$).

In the example given above, the distribution $g(x) = 1/8$ ($x$ = number of blocks required for storage) is too idealized to indicate fully the nature of message populations encountered in practice. Several distributions which are of more practical interest will be studied in Section IV. To provide a useful frame of reference for the results presented in Section IV, we introduce in Section III two extreme distributions of x; these two extremes yield bounds on all the distributions of the drum-filling number, N, that are considered in Section IV.

## SECTION III

## THE 'BRACKETING' POPULATIONS

Consider a message population in which every message requires one block of drum storage. Call this population $C_S$. Consider another message population in which every message requires nine blocks of drum storage. Call this population $C_T$. The mean and standard deviation of $C_S$ are equal to 1 and 0, respectively (i.e., for $C_S$ $\mu = 1$ and $\sigma = 0$). The mean and standard deviation of $C_T$ are equal to 9 and 0, respectively (i.e., for $C_T$ $\mu = 9$ and $\sigma = 0$). Both populations have the unusual feature that they have no variability (i.e., no spread).

It is obvious that when all the arriving messages are drawn from $C_S$ the drum-filling number, N, equals L; similarly, when they are all drawn from $C_T$ N equals L/9 (approximately). These conclusions imply that when the sampling is from $C_S$ the distribution of N has '0 spread' and that when the sampling is from $C_T$ the distribution of N also has 0 spread. It will be instructive to see what our formulas for $E_N$ and $\sigma_N$ (given in Section II) yield when we apply them to sampling from $C_S$ and to sampling from $C_T$. In the case of $C_S$ we have $\mu = 1$ and $\sigma = 0$; hence $E_N = L/\mu = L$ and $\sigma_N = (\sigma/\mu)\sqrt{L/\mu} = 0$. In the case of $C_T$ we have $\mu = 9$ and $\sigma = 0$; hence $E_N = L/9$ and $\sigma_N = 0$. What these formulas tell us is: in the case of $C_S$ the population of N-values has 0 spread and all its values are concentrated at L;

in the case of $C_T$ the population of N-values has 0 spread and all its values are concentrated at L/9 (approximately). It should be noted that the results obtained by using the formulas agree perfectly with the common-sense conclusions stated in the first part of this paragraph.

Graphs of $E_N = L/\mu$ (for $\mu = 1$ and $\mu = 9$) are given in Figure 1. The upper graph is associated with $C_S$ and the lower one with $C_T$. It will be apparent from the discussion in Section IV and Section V that these two graphs are bounds (see Figures 2 and 3 on pages 19 and 20).

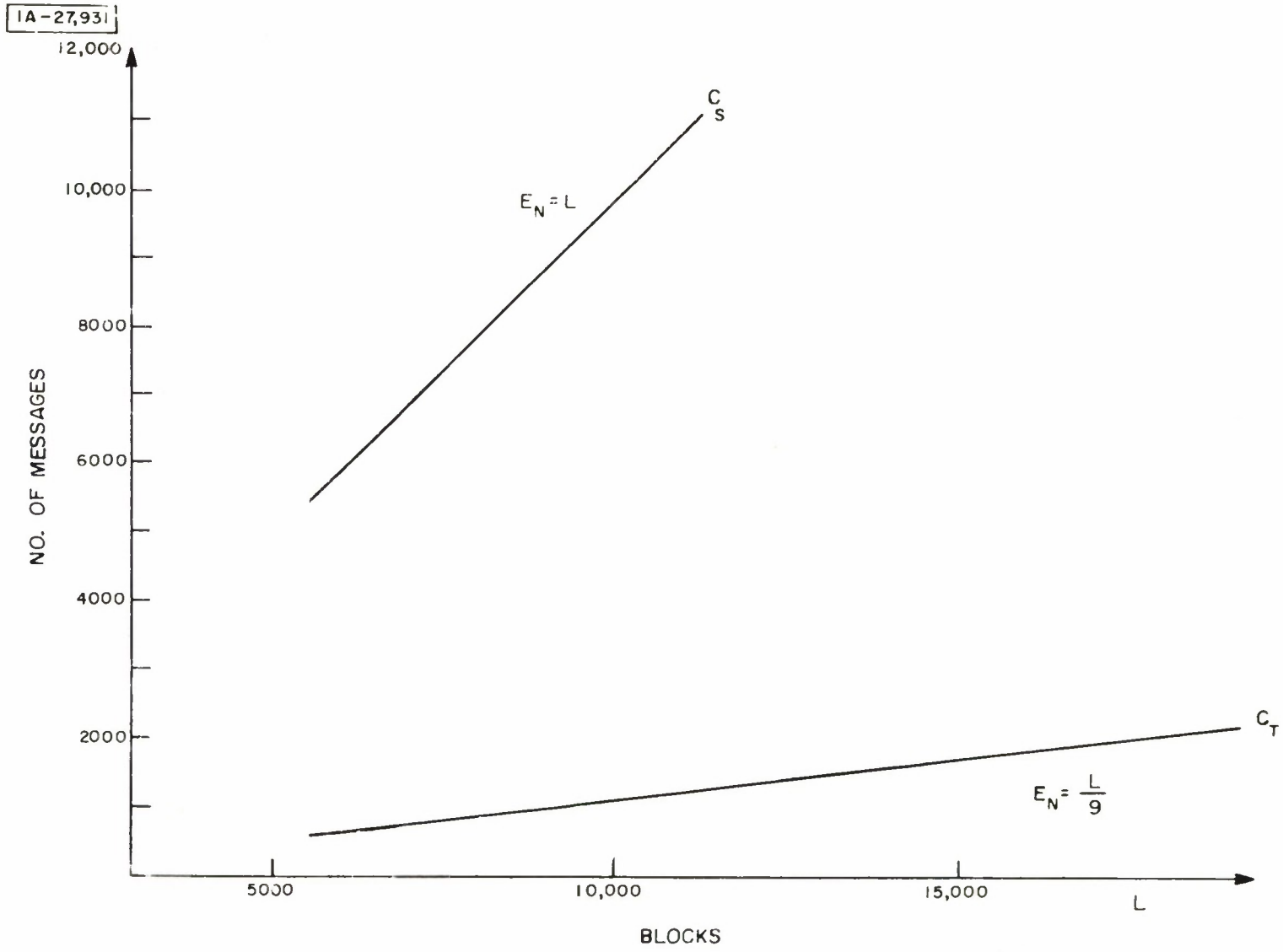



Figure 1. Limiting Values of $E_N$

## SECTION IV

## NUMERICAL ILLUSTRATIONS

The main results given in this section are numbers indicating the center and spread of the distribution of the number of in-transit messages. These results are given for all combinations of five distinct message populations and two widely different drum sizes. As indicated in Table II, the populations are denoted by $C_0$, $C_1$, $C_2$, $C_3$ and $C_4$; the drum sizes are 8192 and 16384.

In specifying the number of blocks required to store a message we have used the following values for $B_1$, $B_2$, $B_3$, etc., which are the ends (in characters) of the drum blocks:*

$$B_1 = 602$$

$$B_2 = 602 + 672 = 1274$$

$$B_3 = 1274 + 672 = 1946$$

$$\vdots$$

$$B_9 = 5306 + 672 = 5978.$$

(Note that we are assuming that no message will require more than nine blocks of drum storage.)

---

*For further discussion of the quantities $B_1$, $B_2$, $B_3$, etc., see Section II and Appendix Section II.

The population $C_0$ is uniform and has the following range: 70 characters $\leq w \leq$ 5400 characters. The probability function $g(x)$ associated with this population is, therefore,

$$g(1) = \frac{537}{5331}, \; g(2) = g(3) = \ldots = g(8) = \frac{672}{5331}, \; g(9) = \frac{94}{5331} .$$

The populations $C_2$, $C_3$, and $C_4$ are, so to speak, 'composite' populations since they are formed by distinct 'mixtures' of two basic populations. We shall first explain what the basic populations are and then indicate how they are 'mixed' to form $C_2$, $C_3$, and $C_4$.

One of the basic populations is $C_1$, which consists of narrative messages (i.e., 'people-to-people' messages); we shall call this the <u>narrative population</u>. The other basic population is derived from very long data messages between computers. These very long data messages are broken up into shorter, fixed-length segments. The fixed-length segments constitute our second basic population of messages, and we assume that they are all 5300 characters in length. We shall call this basic population the <u>data population</u>.

The characteristics of the narrative population are given in Table I. Note that its mean ($\mu$) and the standard deviation ($\sigma$) are, respectively, about 2.627 and 2.154. The probability function $g(x)$ associated with the narrative population can be regarded as simply given. (The function used here was actually estimated from data on the lengths, in characters, of narrative messages. These data indicated that the mean length (in characters) of the narrative

messages was 1358.) As regards the data population, it is apparent from the description above that all its messages are eight blocks long; thus, for the data population $\mu = 8$ and $\sigma = 0$.

Populations $C_2$, $C_3$, and $C_4$ are distinct mixtures of the narrative and data populations. $C_2$ is a (90% - 10%) mixture -- i.e., a mixture consisting of nine parts of the narrative population and one part of the data population. $C_3$ is a (71% - 29%) mixture; and $C_4$ is a (10% - 90%) mixture. The means and standard deviations of $C_2$, $C_3$, and $C_4$ have been calculated by making use of the appropriate formulas in Appendix Section IV.

Table II gives $E_N$, $E_N - 3\,\sigma_N$, and $E_N + 3\,\sigma_N$ for each combination of $C_0$, $C_1$, $C_2$, $C_3$, and $C_4$ and two different drum sizes. Table II also gives the quantities m, $\mu$, and $(\sigma/\mu)$ for $C_0$, $C_1$, $C_2$, $C_3$, and $C_4$.

Table I

Narrative Population

| Block | Character Range | Estimate of Total Population Fraction | Estimate of (TPF x Block No.) | Estimate of (TPF x (Block No.)$^2$ |
|---|---|---|---|---|
| 1 | 1 - 602 | .42082 | .42082 | .42082 |
| 2 | 603 - 1274 | .26652 | .53304 | 1.06608 |
| 3 | 1275 - 1946 | .07822 | .23466 | .70398 |
| 4 | 1947 - 2618 | .04472 | .17888 | .71552 |
| 5 | 2619 - 3290 | .04250 | .21250 | 1.06250 |
| 6 | 3291 - 3962 | .04400 | .26400 | 1.58400 |
| 7 | 3963 - 4634 | .05250 | .36750 | 2.57250 |
| 8 | 4635 - 5306 | .04100 | .32800 | 2.62400 |
| 9 | 5307 - 5978 | .00972 | .08748 | .78732 |
| | Total | 1.00000 | 2.62688 | 11.53672 |

$\hat{\mu} = 2.62688$

$\hat{\sigma}^2 = 4.63622$

$\hat{\sigma} = 2.154$

Table II

Values of $E_N$ and $E_N \pm 3\,\sigma_N$ for Combinations of Five Message Populations and Two Drum Sizes

| Population | m = Mean Number Characters | $\mu$ = Mean Number Blocks | $\sigma$ | $\frac{\sigma}{\mu}$ | L = 8192 | | | L = 16384 | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | $E_N$ | $E_N - 3\,\sigma_N$ | $E_N + 3\,\sigma_N$ | $E_N$ | $E_N - 3\,\sigma_N$ | $E_N + 3\,\sigma_N$ |
| $C_0$ Uniform (70-5400) | 2735 | 4.671 | 2.30 | .493 | 1754 | 1692 | 1816 | 3508 | 3420 | 3596 |
| $C_1$ Narrative | 1358 | 2.627 | 2.154 | .820 | 3118 | 2980 | 3256 | 6237 | 6042 | 6432 |
| $C_2$ 90%(Narrative population) 10%(Data population)* | 1752 | 3.164 | 2.60 | .822 | 2589 | 2463 | 2715 | 5177 | 5000 | 5354 |
| $C_3$ 71%(Narrative population) 29%(Data population)* | 2501 | 4.185 | 3.04 | .726 | 1957 | 1861 | 2053 | 3914 | 3776 | 4052 |
| $C_4$ 10%(Narrative population) 90%(Data population)* | 4906 | 7.463 | 1.75 | .234 | 1098 | 1074 | 1122 | 2195 | 2162 | 2228 |

*Data population: $m = 5300$, $\mu = 8$, $\sigma = 0$.

## SECTION V

## COMPARISON OF RESULTS

Several graphs are given in this section which permit one to easily compare and integrate many of the results presented in the previous sections.

The graphs in Figure 2 show $E_N$ as a function of the drum size, L, for the seven populations $C_S$, $C_0$, $C_1$, $C_2$, $C_3$, $C_4$, and $C_T$. It is noteworthy that the graphs for $C_0$, $C_1$, $C_2$, $C_3$, and $C_4$ lie between those for $C_S$ and $C_T$.

The graphs in Figure 3 show, on a different scale, the function $E_N$ again. It will be noted, however, that in Figure 3 the graphs of $E_N$ for $C_0$, $C_1$, $C_2$, $C_3$, and $C_4$ lie inside shaded bands. The purpose of any of these bands is to show the spread of the probability distribution of N abouts its center, $E_N$. Each band extends from $E_N - 3\,\sigma_N$ to $E_N + 3\,\sigma_N$. No shaded band is given for $C_S$ or $C_T$ since in each of these two cases the probability distribution of N has 0 spread (see Section III).

Incidentally, none of the graphs in Figure 2 or Figure 3 is directly associated with the uniform population of x-values considered in Section II; however, the graph of $E_N$ for this case would be extremely close to the graph of $E_N + 3\,\sigma_N$ for $C_0$ (see Figure 3). (In fact, the two graphs would differ at most by about the thickness

of a pencil line.) The reason that the two graphs would nearly coincide rests in the fact that the probability function $g(x)$ is almost the same in the two cases. The $g(x)$ used in Section II is $g(x) = 1/8$ $(x = 1, 2, \ldots, 8)$; and the $g(x)$ for $C_0$ (see Section IV) is only slightly different.

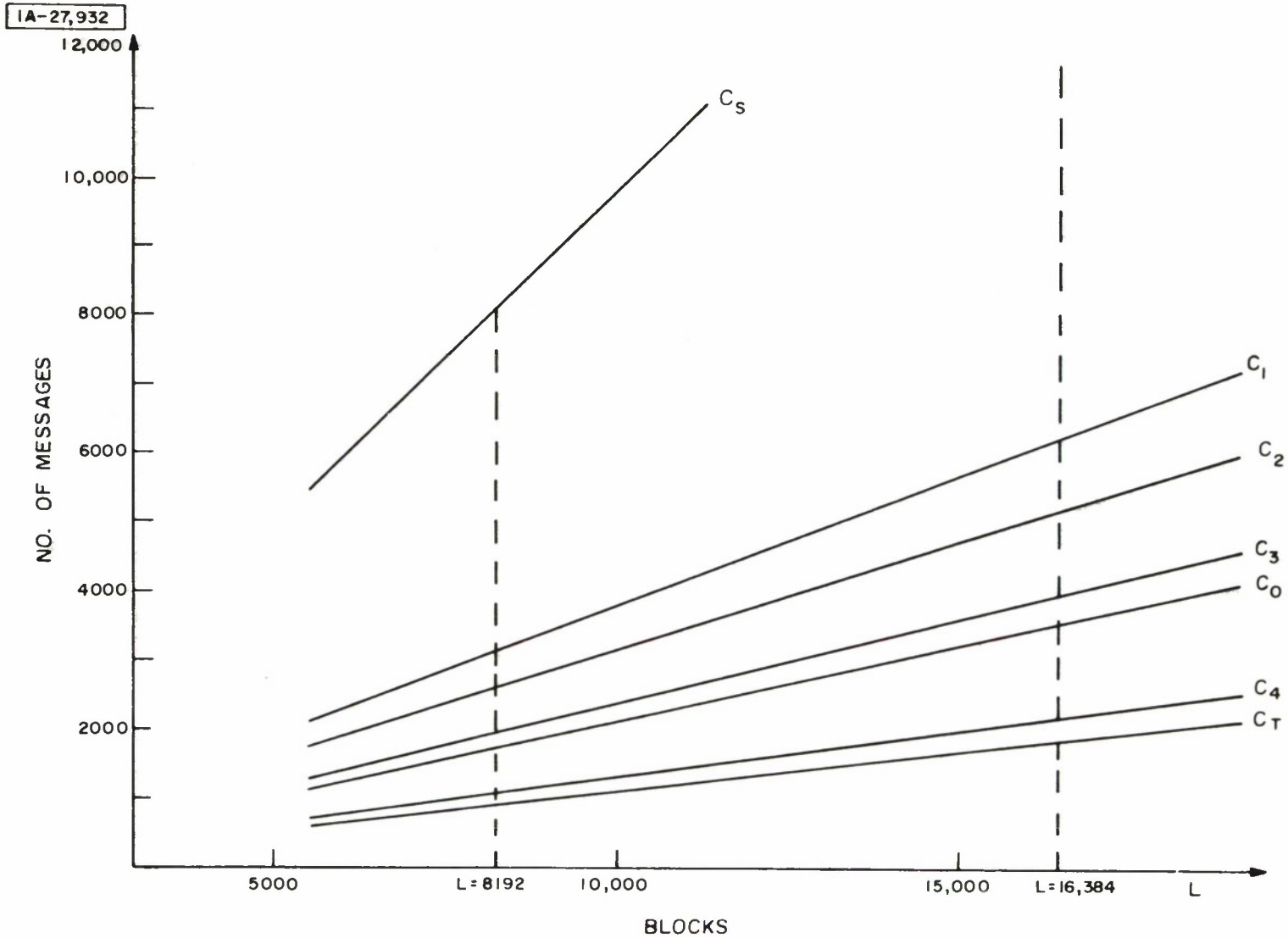



Figure 2. Effect of Drum Size on the Expected Number of In-Transit Messages (L = Drum Size)

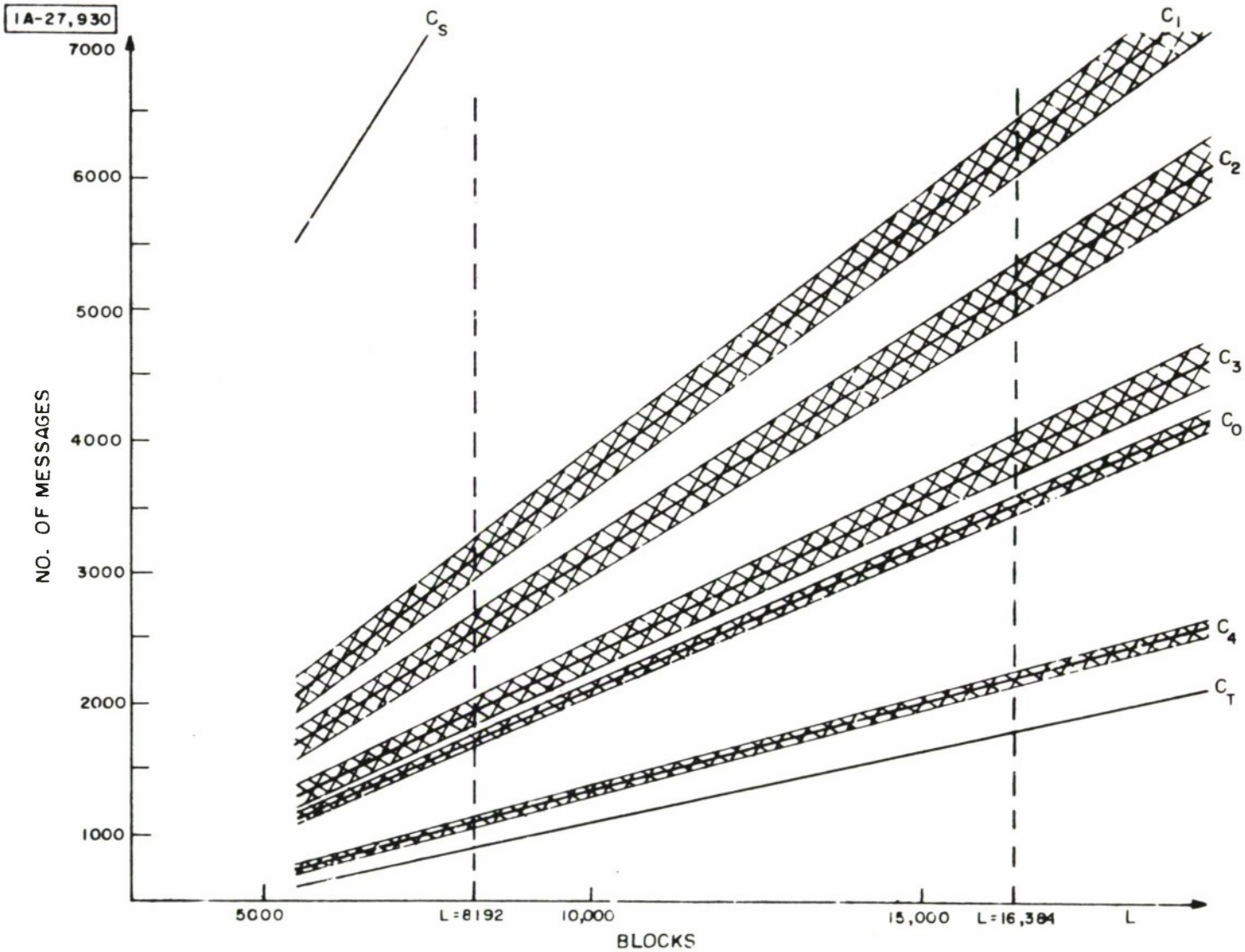



Figure 3. Effect of Drum Size on the Probability Distribution of the Number of In-Transit Messages (L = Drum Size)

## SECTION VI

## SUPPLEMENTARY REMARKS

The normal distribution that is an approximation to the distribution of the drum-filling number, N, is determined by only two parameters -- namely, the mean, $E_N = L/\mu$, and the variance, $\sigma^2_N = (L/\mu)(\sigma/\mu)^2$. Obviously these two parameters depend on only three quantities: $\mu$, $\sigma$, and L. We see, therefore, that only two characteristics of the message population influence the distribution of N; these two are the mean, $\mu$, and the standard deviation, $\sigma$.

It is of interest to compare $\mu$ and $\sigma$ with regard to their effects on the distribution of N (assuming L fixed). We see that: (1) $\mu$ influences the mean, $E_N$, of the distribution, while $\sigma$ does not; (2) both $\mu$ and $\sigma$ influence the standard deviation, $\sigma_N$. In connection with (2) it should be noted that for each of the populations considered in this report $\sigma$ is smaller than $\mu$; thus in each case $\sigma_N$ is smaller than the square root of $E_N$.

## APPENDIX

### I. The Distribution of the Size of a Sample Under a Simple Constraint on the Sample Sum

Let $X_1$, $X_2$, ... be a sequence of mutually independent random variables having the same cumulative distribution function, say $G(x)$, where $x \geq 0$. Such a sequence is called a renewal process (see Reference [1], p. 245). The quantities $X_1$, $X_2$, ..., can be regarded as obtained by drawing values at random one by one from a population of values of $X$, where

$$\Pr(X \leq x) = G(x) \qquad (x \geq 0). \qquad (A.1)$$

(The notation "$\Pr(X \leq x)$" means "the probability that $X \leq x$.") Let $X_1 + X_2 + \ldots + X_n$ be denoted by $S_n$, and let $L$ be any positive number. Suppose now that as the values $X_1$, $X_2$, ... are drawn we take note of how many have been drawn when, for the first time, the sum of the values drawn is greater than $L$; let $N$ denote this number minus 1. For example, if $S_1 < L$, $S_2 < L$, $S_3 < L$, and $S_4 > L$, then the value of $N$ would be $3(= 4 - 1)$. $N$ is a function of random variables, and it is also a random variable.

Let $F(n)$ denote the cumulative distribution function of $N$; i.e.,

$$F(n) = \Pr(N \leq n) \qquad (n = 0, 1, 2, \ldots) \qquad (A.2)$$

A useful expression for $F(n)$ is provided by the lemma stated below.

Lemma:

$$F(n) = \Pr(S_{n+1} > L) \qquad (n = 0, 1, 2, \ldots). \qquad (A.3)$$

Proof: It follows from the definition of $N$ that:

if $S_{n+1} > L$, then $N \leq n$;

and

if $N \leq n$, then $S_{n+1} > L$.

The conditions $(S_{n+1} > L)$ and $(N \leq n)$ are equivalent; therefore $\Pr(S_{n+1} > L) = (\Pr(N \leq n)$, which, by (A.2), equals $F(n)$. This completes the proof.

One reason why the result in (A.3) is useful is that in some cases we can express the exact distribution of $S_{n+1}$ (for any n) in terms of tabulated functions and thereby calculate the distribution function $F(n)$. An even more important reason for the usefulness of (A.3) is that in many cases we can easily calculate an approximate distribution of $S_{n+1}$ and thereby calculate an approximation to the distribution function $F(n)$. More specifically, let $\mu$ and $\sigma^2$ be the mean and variance of X, respectively, and assume that $\sigma^2$ is finite (thus $\mu$ is finite) and that both $\sigma^2$ and $\mu$ are positive; we conclude from the Central Limit Theorem that if $n$ is large, the distribution of $\left[S_{n+1} - (n + 1)\mu\right]/(\sigma\sqrt{n+1})$ is approximately a standard normal distribution. The cumulative distribution function,

say $\Phi(z)$, associated with the standard normal distribution is

$$\Phi(z) = \left(\frac{1}{\sqrt{2\pi}}\right) \int_{-\infty}^{z} e^{-u^2/2}\, du \quad (-\infty < z < +\infty).$$

If $n$ is large, then

$$\Pr(S_{n+1} > L) \doteq 1 - \Phi\left[\frac{L - (n+1)\mu}{\sigma\sqrt{n+1}}\right] = \Phi\left[\frac{(n+1)\mu - L}{\sigma\sqrt{n+1}}\right] \quad (A.4)$$

We conclude from (A.3) and (A.4) that for large $n$

$$F(n) \doteq \Phi\left[\frac{(n+1)\mu - L}{\sigma\sqrt{n+1}}\right]. \quad (A.5)$$

Let $E_N$ be the expected value (i.e., the mean) and $\sigma^2_N$ be the variance of $N$; and assume that $X$ is a positive random variable. The following statements can be obtained directly from some results given by Takacs (see Reference [2], p. 225):

if $\mu$ is finite, then as $L \to \infty$

$$E_N/L \to 1/\mu \quad ; \quad (A.6)$$

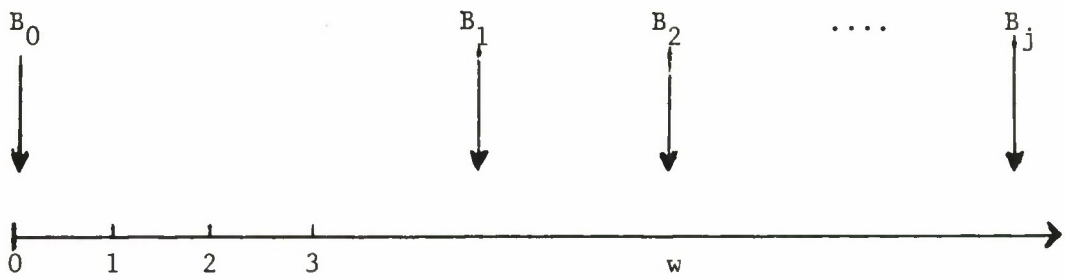
if $\sigma^2$ is finite, then as $L \to \infty$

$$\sigma^2_N/L \to \sigma^2/\mu^3 \quad , \quad (A.7)$$

and $[N - L/\mu] \,/\, [\,(\sigma/\mu)\sqrt{L/\mu}\,]$ has an asympototic standard normal distribution.

## II. Using the Character-Length Distribution to Determine the Distribution of Blocks Used

Let $w$ denote the length of a message in characters ($w \geq 1$). Any message stored on the drum takes up an integral number of blocks. The number of blocks required depends on the length, $w$, of the message. Let $B_1$ be the number of characters available in the first block of drum storage used in storing a message. Let $B_2$ be $B_1$ + the number of characters available in the second block of drum storage used in storing a message. Let $B_j = B_1 + B_2 + \ldots + B_{j-1}$ + the number of characters available in the jth block of drum storage used in storing a message ($j = 1, 2, \ldots,$), where for notational convenience we define $B_0$ as 0. The relations among $B_0$, $B_1$, ... and $w$ are indicated in the diagram below.




The quantities $B_1$, $B_2$, ... are the upper ends of the blocks available for storage of a message on the drum. If $w \leq B_1$, then one block is required to store the message; if $B_1 < w \leq B_2$, then two blocks are required to store the message; if $B_2 < w \leq B_3$, then three blocks are required to store the message; etc. Let $x$ denote the

number of blocks required to store a message of length $w$. $x$ is a function of $w$, and we can express this function as follows:

$$\text{if} \quad B_{j-1} < w \leq B_j, \text{ then } x = j \qquad (j = 1, 2, \ldots). \qquad (A.8)$$

Assume that $W$ is selected at random from a population of message lengths and let $r(w)$ be the probability function of $W$. The mean value, say $m$, of $W$ can be expressed as follows:

$$m = \sum_{w=1}^{K} wr(w), \qquad (A.9)$$

where $K$ is the maximum possible message length in characters.

Let $X$ denote the number of blocks required for storage of a message selected at random from the message population, and let $g(x)$ be the probability function of $X$. The function $g(x)$ can be determined from $r(w)$ by the following relation:

$$g(x) = \sum_{B_{x-1}+1}^{B_x} r(w) \quad (x = 1, 2, \ldots) \qquad (A.10)$$

Let $\mu$ and $\sigma^2$ be the mean and variance of $X$. By definition,

$$\mu = \sum_{x=1}^{H} xg(x), \qquad (A.11)$$

(where $H$ = max. possible value of $x$) and

$$\sigma^2 = \sum_{x=1}^{H} x^2 g(x) - \mu^2 \quad . \qquad (A.12)$$

III. Storage Waste

The use of blocks in storing messages results generally in storage waste. For example, if a message is 1280 characters long and the 'ends' of the drum blocks are those given in Section IV of the text, then the number of characters of drum storage wasted in storing it is 666 ( = 1946 - 1280). It is of interest to determine the expected number of characters actually used in storing a message that requires $x$ blocks (but not more than $x$ blocks) of storage. If $g(x)$ (see (A.10)) is greater than 0, let $m_x$ denote this expected number ($x = 1, 2, \ldots$). It can be shown that:

$$m_x = \sum_{w=B_{x-1}+1}^{B_x} w\, r(w)/g(x). \qquad \text{(A.13)}$$

If $g(x) = 0$ for some value (say $x'$) of $x$, then $m_{x'}$ is not defined -- since no message requires $x'$ blocks of storage.

When the blocks are of equal length and the maximum length of a message is small in relation to L, the expected ratio of number of wasted characters of drum storage to total number of characters of drum storage is less than $1/\mu$ approximately. ($\mu$ is defined in (A.11).) If, furthermore, each $m_x$ lies about midway between $B_{x-1}$ and $B_x$, the expected fraction is $1/(2\mu)$ approximately.

## IV. The Mean and Variance of a Composite Random Variable

Consider two random variables, say $y_A$ and $y_B$. Let $\mu_A$ and $\sigma^2_A$ be the mean and variance of $y_A$, respectively, and let $\mu_B$ and $\sigma^2_B$ be the mean and variance of $y_B$, respectively. Suppose that we form a new random variable, say, y, as follows:

(1) Use a random mechanism to first choose A or B. Let $\theta$ $(0 \leq \theta \leq 1)$ be the probability that A is chosen; thus $(1 - \theta)$ is the probability that B is chosen. (The probability $\theta$ is assumed to be known in advance.)

(2) If A is selected in step (1), choose a value of $y_A$ at random; if B is selected in step (1), choose a value of $y_B$ at random. Let the mean and variance of y be denoted by Mean (y) and Var (y), respectively. It can be shown that

$$\text{Mean } (y) = \theta\, \mu_A + (1 - \theta)\, \mu_B,$$
$$\text{Var } (y) = \theta\, \sigma^2_A + (1 - \theta)\, \sigma^2_B + \theta\, (1 - \theta)\, (\mu_A - \mu_B)^2. \quad (A.14)$$

We can regard y as a 'composite' random variable since it is formed by 'mixing' two other random variables, $y_A$ and $y_B$. The population associated with y can be regarded as a composite population formed by mixing the populations associated with $y_A$ and $y_B$. In this connection it is noteworthy (see (A.14)) that the mean of the y-population is a simple function of the mixing fraction, $\theta$, and the means of the $y_A$-population and $y_B$-population, and the variance of the y-population is a simple function of $\theta$ and the variances and means of the $y_A$- and $y_B$-populations.

## REFERENCES

[1] W. L. Smith, "Renewal Theory and Its Ramifications," Journal of the Royal Statistical Society, Series B, 19-20, 1957-58 (pp. 243-302).

[2] L. Takacs, Introduction to the Theory of Queues, Oxford University Press, New York, 1962.

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)*<br>The MITRE Corporation<br>Bedford, Massachusetts 01730 | 2a. REPORT SECURITY CLASSIFICATION<br>Unclassified<br>2b. GROUP |

3. REPORT TITLE

The Number of Messages Than Can be Stored on a Drum

4. DESCRIPTIVE NOTES *(Type of report and inclusive dates)*

N/A

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

Robert A. Gildea and David F. Votaw, Jr.

| | | |
|---|---|---|
| 6. REPORT DATE<br>25 August 1969 | 7a. TOTAL NO. OF PAGES<br>31 | 7b. NO. OF REFS<br>2 |
| 8a. CONTRACT OR GRANT NO.<br>F19628-68-C-0365<br>b. PROJECT NO.<br>c.<br>d. | 9a. ORIGINATOR'S REPORT NUMBER(S)<br>MTR-952<br>9b. OTHER REPORT NO(S) *(Any other numbers that may be assigned this report)* | |

10. DISTRIBUTION STATEMENT

This document has been approved for public release and sale; its distribution is unlimited.

| | |
|---|---|
| 11. SUPPLEMENTARY NOTES<br>N/A | 12. SPONSORING MILITARY ACTIVITY<br>Deputy for Communications Systems, Electronic Systems Division, L.G. Hanscom Field, Bedford, Massachusetts |

13. ABSTRACT

Messages arriving at a store-and-forward switch are stored on a drum. Each message stored takes up a number of blocks of drum storage that depends on the message length in characters. It is assumed in this report that the arriving messages are drawn at random from a message population; thus the number, say N, of messages that fill the drum is a random variable. It is shown, by means of renewal theory, that the distribution of N is approximately a normal distribution whose mean and variance are simple functions of the drum size and the parameters of the message population. Illustrative examples are given. An estimate of the expected fraction of wasted characters of drum storage is also shown.

DD FORM 1 NOV 65 1473

Security Classification

| 14. KEY WORDS | LINK A ROLE | LINK A WT | LINK B ROLE | LINK B WT | LINK C ROLE | LINK C WT |
|---|---|---|---|---|---|---|
| Digital Message | | | | | | |
| Digital Switch | | | | | | |
| Message Switch | | | | | | |
| Store-and-Forward Switch | | | | | | |
| In-Transit Storage | | | | | | |
| Communications | | | | | | |
| Communications Processor | | | | | | |
| Secondary Storage | | | | | | |
| Traffic | | | | | | |
| Bulk Storage (Communications) | | | | | | |
| Communications Network | | | | | | |
| Data Messages | | | | | | |
| Narrative Messages | | | | | | |
| Renewal Processes | | | | | | |
| Wasted Storage | | | | | | |
| Message Population | | | | | | |
| Message Sampling | | | | | | |