REF ID:A56978

MEMO ROUTING SLIP

NEVER USE FOR APPROVALS, DISAPPROVALS, CONCURRENCES, OR SIMILAR ACTIONS

1 NAME OR TITLE S/A SST	INITIALS		CIRCULATE
ORGANIZATION AND LOCATION Mr Friedman	DATE		COORDINATION
2		✓	FILE
		✓	INFORMATION
3			NECESSARY ACTION
			NOTE AND RETURN
4			SEE ME
			SIGNATURE

REMARKS 2 Reports 342 343

FROM NAME OR TITLE NSA-314		DATE 10 Aug
ORGANIZATION AND LOCATION 17-117		TELEPHONE 60391

DD FORM 96 REPLACES NME FORM 94, 1 FEB 49 WHICH MAY BE USED. 16—66969-2 GPO

SECRET

PRELIMINARY REPORT ON AFSAM 9 DEPTH STUDY

INTRODUCTION:

One of the greatest difficulties encountered in assessing the security of cryptographic devices which operate on principles similar to that of the AFSAM 9 is the determination of the probability of depth. It has been possible to derive satisfactory a' priori estimates of the probability of depth for the class of structures wherein each setting has a unique predecessor; e.g. the AFSAM 7. However, a' priori estimates of the probability of depth are not at all adequate for structures, such as those generated by the AFSAM 9, which are comprised not only of single points (i.e. settings which have a unique predecessor), but also of critical points (e.g. branch points and origins).

Thus the only practical means of ascertaining the incidence of depth for structures containing critical points is by means of sampling. As to what specific sampling procedures should be employed, there are obviously very many theoretically sound procedures one could propose; however, whether or not a particular sampling procedure is feasible will be largely determined by the kind of high speed analytic equipment that is available.

In Part I of the report which follows is given a description and analysis of the sampling procedures that have been devised to determine the incidence of depth for AFSAM 9 traffic together with a description of the various types of statistics that can be secured by utilizing these procedures. In Part II of the report, the identification and evaluation of

those notch pattern parameters, on which the incidence of depth depends, are discussed.

Part I

The sampling procedure which is currently being used in the study of the AFSAM 9 structure utilizes the facilities afforded by PLUTO. Another sampling procedure which will make use of ATLAS II is currently being programmed and will be ready for use in the near future. The first half of Part I will describe the PLUTO program. This will be followed by a description of the ATLAS II program.

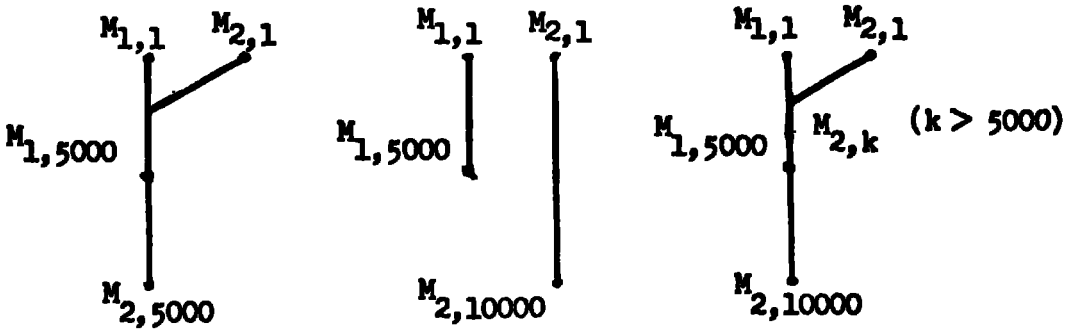
The sampling procedure on PLUTO is logically (but not chronologically) equivalent to the following steps:

(1) A quasi-random initial setting is first generated (by a method to be described later).

(2) With the wheels set to this initial setting, the AFSAM 9 is then stepped through a message of length 5000, and the 5000 th setting is then stored in a "memory."

(3) The wheels are now set to a new quasi-random initial setting, and the device is then stepped through a second message. Each subsequent setting that is generated is compared with the setting in the "memory." The results of this comparison can be characterised by the following figures, where $M_{i,j}$ is the jth setting on the ith message:

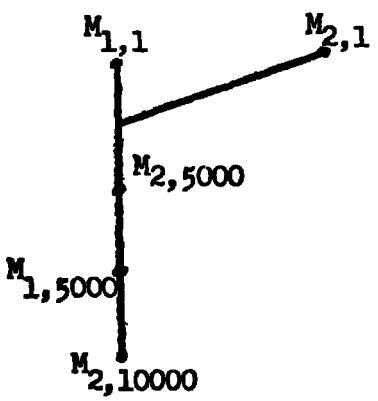
SECRET



Case I Case II Case III a,b,

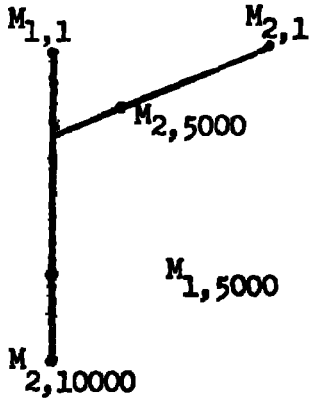
If there is a hit before the device has stepped more than 5000 times, as in Case I, the first message is in depth with the second for at least one setting. If there is no hit before the second message has been stepped through 10,000 settings, as shown by Case II, the two messages are not in depth.

(4) If there is a hit before the second message has been stepped through less than 10,000 settings, but more than 5,000 settings, as indicated by Case III, the 5000th setting of the second message is then inserted in the "memory," and each setting of the first message is now compared with the 50006h setting of the second message. If there is a hit before the first message has been stepped through 5000 settings, the two messages are then known to be in depth for at least one setting.



Case III a

Depth



Case III b

No Depth

(5) If the first and second messages are found not to be in depth, a new quasi-random initial setting is generated and this third message is then compared for depth with each of the first two messages in the manner outlined in steps (1) through (4).

(6) Sampling according to the steps prescribed in (1) through (5) is continued until some message j is found to be in depth with some message i, where $j > i$. The "statistic j" thus obtained corresponds to the number of messages that had to be sent in order to insure that the jth message was the first one yielding depth with any prior message sent in the same cryptoperiod. If an adequate number of runs are made, say n, yielding the distribution of the j's, this statistic can then be utilized as a reliable criterion to determine the amount of traffic load that the AFSAM 9 can safely carry in any given cryptoperiod.

SECRET

In a similar fashion other types of statistics can also be obtained, such as a statistic corresponding to the number of messages which have to be sent in order to insure that exactly k pairs of messages be in depth for at least r settings; or a statistic corresponding to the number of messages which have to be sent in order that k n-tuples be in depth for at least r settings.

The above sampling procedure was not feasible until just recently, as the time involved in obtaining a single statistic of the type j required about 120 hours. The sampling process has now been almost completely automatized by devising a program which enables PLUTO to generate its own "quasi-random" initial settings. At the present time PLUTO can obtain a j statistic in about 45 minutes.

Concerning the sampling procedure currently being used on PLUTO, the feature which is most likely to incur adverse criticism is the method by which initial settings are generated. The initial settings of messages that are to be used in obtaining an observed value of j are generated in the following manner. In addition to the actual AFSAM 9 notch patterns, separate randomly constructed notch patterns are inserted on the wheels W-2, W-3, W-5, W-6, W-7 and will be referred to as the secondary patterns. The wheels are now initially set to some randomly selected setting (designated here as the primary setting) and stepped through message 1. After the final setting of the message has been generated, the rules of motion of the AFSAM 9 are amended by the addition of the following rules:

1. W-1, W-2, W-5 and W-6 are delayed by the simultaneous occurrence of a no-notch on the secondary pattern of W-5 and a no-notch on an auxiliary

SECRET

two point wheel which steps every time.

2. W-i is stepped by the occurrence of a notch on the secondary pattern of wheel W - (i-1), where i = 3, 4, 7, 8.

The wheels are now stepped according to the amended rules of motion for some fixed interval, say 10,000 steps. The 10,000 th setting arrived at under the amended rules of motion is now defined as the initial setting for message .2. By making this procedure iterative, a set of initial settings can thus be generated.

It is apparent that the above procedure will ultimately result in the generation of a cycle, so that after a certain point in the iteration the process will become periodic. There is therefore a limit to the number of initial starts that can be generated by the above procedure. Since it is easy to ascertain at just what point the process becomes periodic, this limit is determined and is never exceeded.

Inasmuch as the AFSAM 9 notch patterns are held constant throughout any given run, it is the secondary patterns and the primary setting which determine what initial settings are used in securing a statistic. It is therefore important that a new random primary setting and a new random set of secondary patterns be selected for each statistic that is to be obtained.

In regard to the selection of the amended rules of motion referred to in the aforementioned procedure for generating initial settings, the selection was not completely arbitary but was governed by the following factors:

(a) It is desirable that the amended rules of motion be easy to implement on PLUTO.

SECRET

(b) In comparison with the actual AFSAM 9 structure, the structure generated by the amended rules of motion should not be comprised of a disproportionately large number of confluences.

(c) It has been shown in [1] that the cycle structure of the AFSAM 9 is comprised of 108 disjoint sets of structures, where the selection of the initial setting of W-1, W-2, W-5 and W-6 determines which of the structures is being used. It was desired that depth studies be confined to a particular one of the 108 mutually exclusive sets into which the structure of the AFSAM 9 can be resolved, rather than to study the entire structure. Therefore, a set of amended rules of motion in order to be satisfactory must maintain the relationship of the settings of W-1, W-2, W-5 and W-6 necessary to this chosen one of the structures.

Inasmuch as it is desirable that any method of generation of initial settings simulate a random process, one would like to know if it is statistically feasible to assume that the initial settings generated by the procedure described above could arise by random sampling from the population at large. To determine whether or not the hypothesis that the generated initial settings could have been obtained by random selection can be statistically rejected, one could devise any number of tests to exhibit the degree to which the method of generation of initial settings is characterized by mathematical randomness. However, it should be pointed out that even though one were able to reject statistically the hypothesis that the generated initial settings are mathematically random, the method of generation might still be acceptable. That is, a set of initial settings may fail to exhibit

characteristics of mathematical randomness and yet be dispersed so haphazardly over the entire AFSAM 9 structure that their use does not invalidate any statistical analyses pertaining to depth studies. All that is required here as regards the conditions of selection of initial settings is that there should be no marked element of preference or bias that would tend toward the inclusion or exclusion of certain local areas of the AFSAM 9 structure.

Suppose, then, one were to compare several prominent characteristics (such as precentage of frequency of occurrence of the structures, frequency of incidence of certain entry points, expected length of lead-in for each of several entry points) of the statistical picture of a particular AFSAM 9 structure obtained from N mathematically random initial starts with these same characteristics of the statistical picture obtained from N initial settings that have been generated by the method in question. If it were found that in both cases these characteristics were substantially the same, it would then seem quite reasonable to regard the method of "pseudo-random" generation as satisfactory for the purpose of depth study.

A comparison test of this type, based on a sample of 1000 initial starts, has been applied to one of the 108 sets of structures generated by the AFSAM 9 for an arbitrarily chosen set of notch patterns. The results of this test are described in Appendix A.

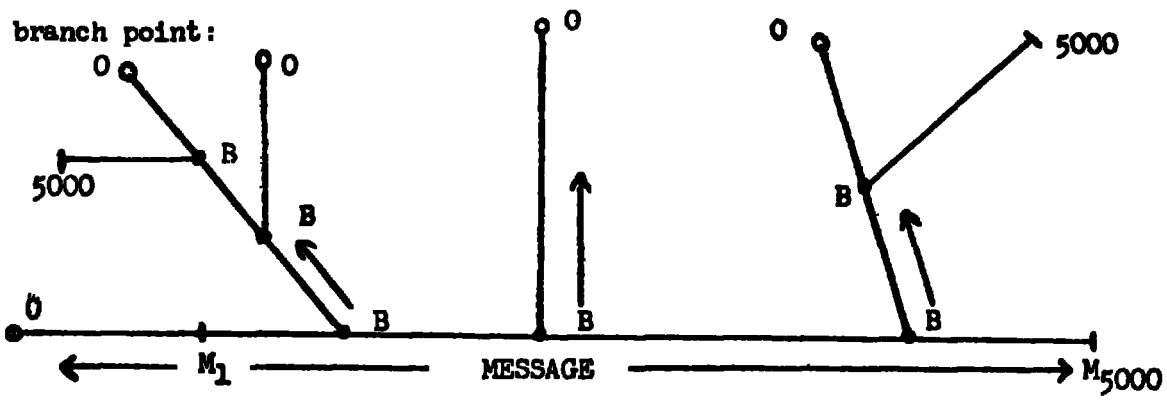
The ATLAS II procedure involves a more direct approach to the solution of the depth problem, simply by answering the question, "How many messages can be sent which will be in depth with a given message?" This number can be found by reproducing that portion of the cycle structure which contains

~~SECRET~~

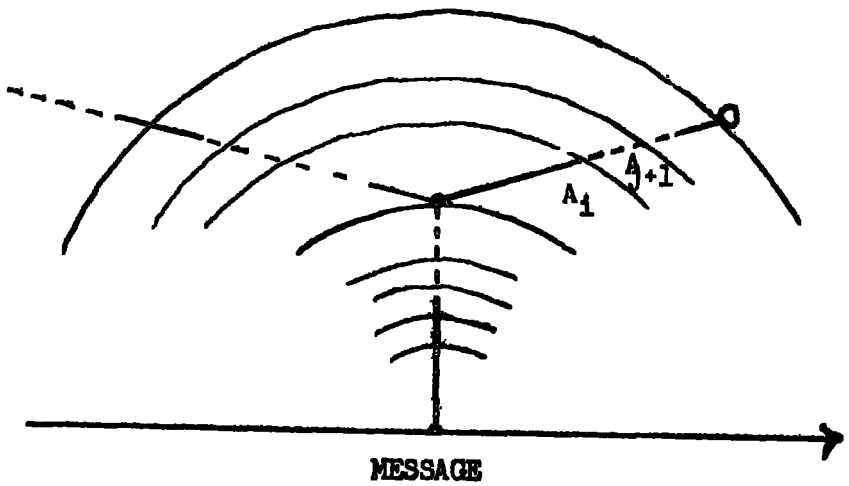
all paths leading into the given message and for which no path exceeds the maximum message length, say 5000. In order to simplify the problem and thus decrease machine time, the assumption has been made that the probability of depth with an indicator system (restricting the number of possible message starting points) is <u>not</u> significantly greater than the probability of depth with no restrictions. At this writing there has been no conclusive evidence denying this assumption.

A more detailed description of the ATLAS program follows: In a manner analogous to a cycle study, whereby a picture of a particular cycle structure is desired, consider the message as a "cycle" of length 5000. The entry points to this "cycle" are then all the branch points on the message together with the initial or starting point of the message. Then the lead-inswill be the paths taken by other messages such that depth occurs.

The method of obtaining this sub-structure is a two-phase procedure. The first phase involves normal (forward) stepping of the device from some setting M_1 to produce the message as far as setting M_{5000}. The second phase uses reverse (backward) stepping starting from the branch points and M_1 and continuing 5000 points or until an origin terminates the stream. This can be shown graphically in the following diagram, where O represents an origin, B a branch point:



In order to simplify a description of the ATLAS procedure, consider the following diagram, where an arc centered at the message branch point is drawn through each successive point in the structure (backward) such that a point on arc A_i is i settings removed from the message branch point.



Two or more points on the same arc have in common the 3-wheel CCM setting contained in the AFSAM 9 setting. Furthermore, the motion of the remaining five wheels proceeding from a branch point (neccessarily of multiplicity two) in one direction is the complementary motion of the opposite direction. These two statements constitute the basis of the counting procedure, whereby successive CCM settings account for the i associated with A_i along one stream (direction), and the number of such streams f_i intersecting or terminating on A_i is

$$f_i = f_{i-1} + B_{i-1} - O_{i-1}$$

where B_{i-1} and O_{i-1} are the number of branch points and origins respectively along A_{i-1}. Hence the totality of points in this sub-structure is

$$T = \sum_{i=1}^{4999} f_i$$

SECRET

If one calls the stream originating (backwards) from the message "machine number 1", or m_1, then in determining the two predecessors of a branch point, let the predecessor that is reached by moving W_j (one of the 5 wheels not contained in the three CCM wheels) originate a new machine, m_2, while the other predecessor will continue along m_1. Continue this procedure for each branch point. Then at some arc A_i there will be f_i machines still active in the backward generation and counting procedure. Therefore the procedure is first to move the CCM wheels; secondly (for each m_1) to determine whether the point is a single point, branch point or origin (simultaneously finding the motion pattern); finally to move the wheels if necessary. If m_1 is at a single point, no "machine" change is necessary. A branch point involves setting up a new "machine", while an origin causes that "machine" to be made available for a new branch when needed.

The preceding paragraphs show how T, the total number of points from which a depth of at least one can occur, is found. In general one is looking for T_D, the number of points yielding a depth of at least $D > 1$:

$$T_D = \sum_{1}^{5000-D} f_i$$

Then the total sub-structure will give $N_D = \Sigma T_D + 5000-(D-1)$ points which will result in a depth of D, where the summation is over the entry points to the message, say M_1, where $i \leq 5000-(D-1)$. Other counts to be tabulated are the number of branch points and origins in the sub-structure.

Let the total number of points in the structure be designated by N. With each of several sets of notch patterns, a sufficient number of runs will give enough statistical data to yield $p = N_D/N$, which can be used to estimate the

SECRET

true probability of depth.

Part II

In any empirical approach to the problem of determining the incidence of depth, the question quite naturally arises as to whether one should confine his study to several sets of notch patterns that have been deliberately selected so as to amplify or exaggerate certain attributes of the notch pattern population at large; or whether one should employ the method of random selection. If enough information concerning the relation between the notch pattern parameters and the genealogy of the AFSAM 9 structure is available, the former procedure would seem to be the more expedient. It therefore is desirable to devote some effort to the problem of identifying and evaluating those notch pattern parameters of which the incidence of depth is a function.

The most important attribute affecting the incidence of depth is the number of branch points that are contained in the motion structure generated by a set of notch patterns. Let N (B) be the function which relates the number of branch points to the notch pattern parameters. The CCM settings distribute themselves into three classes of size $(36)^3/3$, since (21,15) = 3. One of the criteria for a branch point is that W-3 be on a minus (-). Hence the number of settings of the CCM which may yield branch points is $(36)^2 15/3 = 6480$. The 16 branch point patterns on 5 wheels give

$$B_4 B_7 B_8 \left[4B_5 (15+C_6) + 4B_6 (15+C_5) \right]$$

where C_i is the number of no-notches preceeded by no-notches (—) on W-i. Hence

$$N(B) = 6480\, B_4\, B_7\, B_8 \left[4B_5\,(15+C_6) + 4B_6\,(15+C_5)\right].$$
$$= 25{,}920\, B_4\, B_7\, B_8 \left[B_5\,(30-B_6) + B_6\,(30-B_5)\right]$$
$$= 51{,}840\, B_4\, B_7\, B_8 \left[15\,(B_5+B_6) - B_5\, B_6\right]$$

In the above expression B_i $(i = 4,5,6,7,8)$ corresponds to the number of blocks of which the notch pattern on wheel W_i is comprised. The number of blocks is defined as one half the number of changes of sign. Because of the criteria that are used in defining whether or not a given notch pattern is an admissable one, N (B) is independent of the patterns on W-1, W-2 and W-3, and the B_i are limited to the range of values $5 \leqslant B_i \leqslant 15$. It is apparent that the function N (B) is maximized when:

(1) $B_4 = B_7 = B_8 = 15$

(2) $B_5 = 15$ and/or $B_6 = 15$

and is minimized when:

(1) $B_i = 5$ $(i = 4,5,6,7,8)$

Therefore, the number of branch points in any AFSAM 9 structure cannot exceed 118,098,000,000; nor can the number of branch points be less than 2,430,000,000. If the maximum number of branch points that is realizable is defined as 100 percent saturation, it then follows that the totality of branch points can never be less than 2.05 percent of saturation.

Statistics relevant to the incidence of depth which have been secured for a particular set of notch patterns would be much more significant if the probability with respect to a given notch parameter of selecting the

set is known. For example, if a set of patterns is selected which generates a structure with the number of branch points classified as 48-49 percent saturation, it would be very useful to be able to estimate the probability of securing by random selection a set of patterns for which the number of branch points would have a classification which exceeds 49 percent saturation.

Because the degree of branch point saturation is the most important attribute affecting the incidence of depth, it is expedient to devote some time to the problem of deriving a probability distribution which can be used to determine the probability of selecting a set of patterns for which the degree of saturation is greater than x percent. Inasmuch as the number of branch points is, in terms of the notch pattern parameters, a function only of the number of blocks, the problem of obtaining such a distribution obviously entails a means of counting or estimating the number of admissable notch patterns which have i blocks ($5 \leqslant i \leqslant 15$). The problem of securing such counts is not exactly trivial, but could be programmed for ATLAS II. However, the probability distribution for the population of patterns which satisfy only that portion of the criterion of admissability which requires that there be no more than five contiguous notches or no-notches can be obtained directly from known results obtained in a report by B. Harris, NSA-314, "On the Number of Cyclic Binary Patterns". This distribution appears on page number 15.

DISTRIBUTION A

i	P(i)	Regrouped P(i)	Theoretical Frequency
5	0.000		
6	0.003	.038	48.
7	0.035		
8	0.144	.144	182.
9	0.282	.282	355.
10	0.294	.294	370.
11	0.172	.172	217.
12	0.058	.058	72.
13	0.011		
14	0.001	.012	15.
15	0.000		
			1259.

The working hypothesis was now made that the above distribution for the population of patterns which are only partially admissible does not essentially differ from the distribution for the population of admissable patterns and could therefore be used to estimate the proportion of admissable patterns which have i blocks. This estimate was then incoprorated in to an ATLAS program to obtain the desired probability distribution, which is given in Appendix B.

Empirical results which lend credence to the working hypothesis referred to in the preceding paragraph were secured in the following manner. The distribution of the number of admissable patterns which have i blocks was obtained for a previously constructed and available sample of 1258 fully admissable patterns. (For a description of the method by which these admissable patterns were constructed, see reference [2] at the end of this report). This sample distribution together with the computation of χ^2 is given on page 16.

SECRET

DISTRIBUTION B

i	f_o	Regrouped f_o	P(i)	$(f_o - f_t)^2 \div f_t$
5	0			
6	4	47	.037	.021
7	43			
8	185	185	.147	.049
9	359	359	.285	.045
10	372	372	.296	.011
11	195	195	.155	2.230
12	83	83	.066	1.681
13	17			
14	0	17	.014	.643
15	0			4.679

From a table of χ^2 it is seen that when the number of degrees of-freedom is 6 a value of χ^2 as great or greater than 4.779 would occur purely as a result of chance in more than two out of three random samples. Therefore the working hypothesis that distribution A is not essentially different from distribution B is entirely consistent with the above results.

An examination of the distribution in Appendix B shows that the probability of securing by random selection a set of admissable patterns having a percentage of saturation greater than 50 percent is less than 0.0005. If, therefore, it were known that the only important attribute affecting the incidence of depth is the number of branch points comprising a structure, statistics relevant to the incidence of depth could then be secured by the procedures described in Part I for several sets of patterns having, say, from 48 percent to 50 percent saturation. These statistics could then be used as criteria to determine the amount of traffic load that can be safely tolerated, since the probability of a situation arising for which these saturation criteria are not satisfied would be less than 0.0005.

The identification and evaluation of other notch pattern parameters which affect the incidence of depth are therefore matters of practical importance. Before attempting to identify any other pattern-parameters which might affect the incidence of depth, however, it is first necessary to determine how the genealogical characteristics of a general structure of the AFSAM 9 type are related to the incidence of depth. Those characteristics of the structure which do significantly affect the incidence of depth may perhaps then be expressed as some function of the pattern parameters, as has already been done for the function N (B).

An investigation of the characteristics of a general structure reveals that there are three distinct types or classes of branch points. Those branch points which have as their two predecessors (either immediate or removed by one or more single points) two branch points are here classified as type I. Those branch points which have as their two predecessors two origins are classified as type II, and those which have as their two predecessors one origin and one branch point as type III.

In terms of these three classes of branch points, the two limiting distributions for any tree comprised of N critical points are:

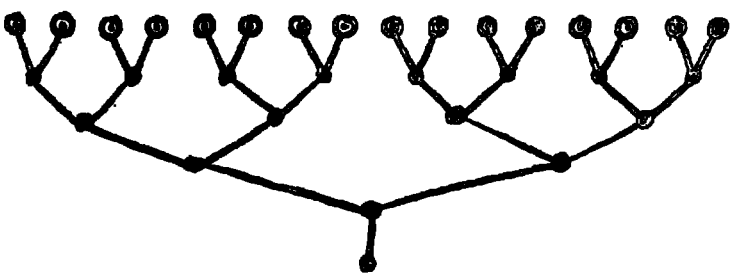
	Distribution I	Distribution II
Type I	N/2	1
Type II	N/2	1
Type III	0	N-2

SECRET

(The configuration comprised of an n-cycle branch point together with its entire associated off-cycle network of predecessors is here referred to as a "tree".)

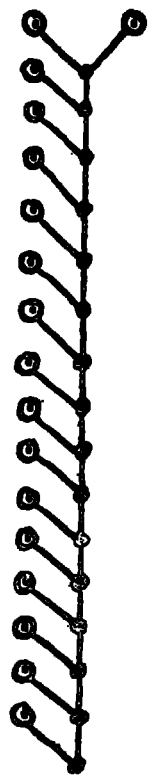
From the illustrations of two trees which are given below, it is apparent that for structures comprised solely of critical points the function F (P), corresponding to the probability of a pair of messages being in depth, is maximized for the set of trees having the limiting distribution I.

Distribution I: N = 32



TYPE I TREE

Distribution II: N = 32



TYPE II TREE

SECRET

The function $F(P_I)$ has been formulated for the particular Distribution I case for which $N = 2^x$:

$$F(P_I) = \frac{2}{N^2} \left\{ \left[3 \quad (2)^{L-2} - 1 \right] N - (2^{L-1} - 1) \cdot 2^{L-1} \right\}$$

In the above expression L is the length of the message and N is the weight of the tree, i.e. the total number of points or settings of which the tree is comprised. The function $F(P_I)$ gives the exact probability of depth for messages of length L when N is a power of two, and is a reasonably good approximation for all other admissable values of N. (Trees having the limiting distribution I require that $N \equiv 0 \pmod 4$, while those having the limiting distribution II require that $N \equiv 0 \pmod 2$).

The corresponding function $F(P_{II})$ has been formulated for all admissable values of N, and appears below:

$$F(P_{II}) = \frac{4}{N^2} \left[L(N+2) - L^2 - (N+1) \right]$$

It is apparent that when $L = 2$,

$$F(P_I) = F(P_{II}) = \frac{4}{N^2} \left[N - 1 \right],$$

and that when $L \geq 3$, $F(P_1) > F(P_{II})$. Moreover, when L is approximately 50 or more and $N \gg L$, the ratio

$$\frac{F(P_I)}{F(P_{II})} = \frac{1}{2} \left[\frac{(3 \cdot 2^{L-2} - 1) N - (2^{L-1} - 1) 2^{L-1}}{L(N+2) - L^2 - (N+1)} \right] \gg 1$$

It would thus appear that the type of branch points of which a structure

is predominantly comprised is a factor which does significantly affect the incidence of depth.

It has been conjectured that the distribution of single points between branch-origin segments and branch-branch segments may materially affect the probability of depth. In accordance with this conjecture, it is believed that the most favorable distribution consists of distributing single points as uniformly as possible between branch-origin segments of the tree. Therefore, one may conditionally regard the distribution of single points as a factor affecting the incidence of depth.

As regards this third factor, it can be demonstrated that the distribution of single points between branch-branch segments of the structure cannot be made to differ substantially from the distribution of single points between branch-origin segments; so that it is impossible to distribute the single points in the ideal manner described above. To illustrate this fact, consider the functions $F_{BO}(C_i,W_j,X_k)$, $F_{BB}(C_i,W_j,X_k)$, $G_{BBB}(C_i,W_j,X_k,Y_l)$ and $G_{OBO}(C_i,W_j,X_k,Y_l)$, where F_{BO} determines the number of branch points which have exactly one immediate predecessor that is an origin, F_{BB} determines the number of branch points which have exactly one immediate predecessor that is a branch point, G_{BBB} determines the number of branch points whose two immediate predecessors are branch points, and G_{OBO} determines the number of branch points whose two immediate predecessors are origins. It can be proven that the function corresponding to the number of branch points whose two immediate predecessors are a branch point and an origin is always null for AFSAM 9 structures and therefore need not be considered. The functions F_{BO}, G_{OBO},

SECRET

F_{BB}, and G_{BBB} are expressed below in terms of the notch parameters C_i, W_j, X_k and Y_l, where C_i corresponds to the number of blocks on W_i, W_j corresponds to the number of 010 configurations on W_j, X_k corresponds to the number of 101 configurations on W_k, and Y_l corresponds to the number of 000 configurations on W_l. In all of the following expressions H_{00} is the function corresponding to the number of 00 configurations in the dilated output stream of W-3.

$$F_{BO} = H_{00} \{X_4(15-C_5) W_6 X_7 C_8 + X_4 C_5 (15-C_6) C_7 C_8 + C_4 (15-C_5) W_6 W_7 X_8$$
$$+ C_4 C_5 (15-C_6) C_7 X_8 + C_4 C_5 (15-C_6) X_7 C_8 + X_4 C_5 (15-C_6) W_7 W_8$$
$$+ C_4 (15-C_5) X_6 C_7 C_8 + X_4 (15-C_5) X_6 C_7 W_8 + C_4 (15-C_5) C_6 X_7 C_8$$
$$+ C_4 X_5 (15-C_6) C_7 C_8 + W_4 (15-C_5) C_6 W_7 W_8 + W_4 X_5 (15-C_6) C_7 W_8$$
$$+ W_4 W_5 (15-C_6) X_7 X_8 + C_4 W_5 (15-C_6) W_7 X_8 + W_4 (15-C_5) C_6 C_7 C_8$$
$$+ C_4 (15-C_5) C_6 C_7 X_8\} - G_{OBO}$$

$$G_{OBO} = H_{00} \{X_4 Y_5 W_6 X_7 X_8 + X_4 W_5 Y_6 W_7 X_8 + W_4 Y_5 W_6 W_7 X_8$$
$$+ W_4 W_5 Y_6 X_7 X_8 + W_4 X_5 (15-C_6) X_7 W_8 + X_4 X_5 (15-C_6) W_7 W_8$$
$$+ W_4 (15-C_5) X_6 W_7 W_8 + X_4 (15-C_5) X_6 X_7 W_8\}$$

SECRET

SECRET

$$F_{BB} = H_{00} \{X_4 C_5 (15-C_6) X_7 C_8 + C_4 X_5 (15-C_6) C_7 X_8 + W_4 (15-C_5) C_6 X_7 C_8$$
$$+ C_4 (15-C_5) X_6 C_7 X_8 + X_4 (15-C_5) X_6 C_7 C_8 + C_4 (15-C_5) C_6 W_7 X_8$$
$$+ W_4 X_5 (15-C_6) C_7 C_8 + C_4 C_5 (15-C_6) W_7 X_8 + X_4 C_5 (15-C_6) C_7 W_8$$
$$+ C_4 W_5 (15-C_6) X_7 C_8 + W_4 (15-C_5) C_6 C_7 W_8 + C_4 (15-C_5) W_6 X_7 C_8$$
$$+ C_4 (15-C_5) C_6 C_7 C_8 + X_4 (15-C_5) W_6 W_7 W_8 + C_4 C_5 (15-C_6) C_7 C_8$$
$$W_4 W_5 (15-C_6) W_7 W_8\} - G_{BBB}$$

$$G_{BBB} = H_{00} \{X_4 W_5 Y_6 X_7 W_8 + W_4 Y_5 W_6 X_7 W_8 + X_4 Y_5 W_6 W_7 W_8 + W_4 W_5 Y_6 W_7 W_8$$
$$+ X_4 X_5 (15-C_6) X_7 X_8 + W_4 (15-C_5) X_6 X_7 X_8 + X_4 (15-C_5) X_6 W_7 X_8$$
$$+ W_4 X_5 (15-C_6) W_7 X_8\}$$

The function H_{00} has not been explicitly formulated in terms of the notch pattern parameters of W-1, W-2 and W-3; however, empirical study strongly suggests that H_{00} is relatively independent of the patterns on W-1 and W-2 and is primarily a function of the number of 00 configurations in the pattern on W-3. In any case, for any given set of patterns it is easy to determine by actual count the exact value of H_{00}.

A cursory examination of the functions F_{BO}, F_{BB}, G_{BBB}, and G_{OBO} reveals that it is impossible to design a set of patterns for which F_{BO} differs appreciably from F_{BB} and for which G_{BBB} differs from G_{OBO}. Therefore the ideal distribution of single points referred to above, viz, that no single points occur between branch-branch segments of the structure, is physically unattainable, and cannot even be approximated.

SECRET

However, although the distribution of single points between branch-branch segments cannot be made to differ radically from the distribution of single points between branch-origin segments, it is possible by maximizing or minimizing G_{OBO} or G_{BBB} to cause the number of branch points having as their two immediate predecessors two branch points or two orgins to vary over a considerable range. It has already been demonstrated that, for N (B) held constant, the incidence of depth is very much greater for structures characterized by a type I limiting distribution than for those characterized by a type II limiting distribution. Therefore, it would seem quite logical to hypothesize that, with N (B) held constant, the incidence of depth for a general structure (one that is comprised of a much greater number of single points than of critical points) can be significantly increased or decreased by increasing or decreasing the number of branch points having as their two immediate predecessors two branch points.

If the aforementioned hypothesis is valid, it then follows that the function G_{BBB} is an important parameter affecting the incidence of depth. To determine the statistical feasibility of the above hypothesis the following test was made: Two sets of patterns were constructed so that the structure generated by Pattern I was comprised of the same number of critical points, viz., 48 percent - 49 percent saturation, as that generated by Pattern II. Moreover, Pattern I was designed so as to make the function G_{BBB} relatively small, while Pattern II was designed to make G_{BBB} relatively large. Pattern I and Pattern II are listed in Appendix C. Reference to the distribution in Appendix B indicates that the probability of selecting at random a set of patterns which generates a structure in which the critical points comprise more than 49 percent saturation is less than 0.0002. Therefore, the

parameter N (B) is very much larger for Patterns I and II than it would ordinarily be for an operational AFSAM 9 device. The actual value of N (B) for Pattern I was 57,241,313,280 branch points and for Pattern II was 56,862,000,000 branch points.

It will be recalled at this point that the cycle structure of the AFSAM 9 can be resolved into three pair-wise disjoint sets of structures C_i (i=1,2,3) and that each set C_i can in turn be resolved into 36 pair-wise disjoint subsets $C_{i,j}$ (i=1,2,3, j=0,1,2,....35). It is easy to deduce that the number of branch points is the same for each one of the 108 sets of structures $C_{i,j}$. However, the function G_{BBB} will in general differ for each $C_{i,j}$. The function H_{00} will, of course, take on three values, i.e. a value for each of the three sets of structures C_i where $C_i = \sum_{i=0}^{35} C_{i,j}$ (i=1,2,3). For both Pattern I and Pattern II, it was decided to confine observations to the class of structures $C_{1,0}$. For Pattern I, $H_{00}(C_1)$ = 2625, $G_{BBB}(C_1)$ = 90,578,250 (0.47 percent of all branch points in C_1) and $G_{BBB}(C_{1,0})$ = 2,529,630 (0.48 percent of all branch points in $C_{1,0}$) For Pattern II, $H_{00}(C_1)$ = 5195, $G_{BBB}(C_1)$ = 3,991,754,880 (21.06 percent of all branch points in C_1) and $G_{BBB}(C_{1,0}$ = 110,911,080 (21.07 percent of all branch points in $C_{1,0}$.)

The distribution of the j - statistic for a sample of size 30 was next secured for the two sets of patterns, I and II, where in both cases observations were limited to the set of structures $C_{1,0}$. (The j - statistic has already been defined in Part I as the number of messages that had to be sent in order to insure that the jth message was the first one yielding depth with any prior message.) These two distributions are given in Appendix D.

SECRET

The means for the above two sample distribution were now compared to determine whether both distributions could have been drawn from the same parent-population. Assuming that the two underlying distributions are normal with identical variances, one may test the hypothesis that the two means are equal by using "Student's" t distribution.

The measurement needed for testing the above hypothesis is computed from the relation

$$\sigma_D = \sqrt{\frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}}$$

$$\sigma_D = \sqrt{\frac{(6950)}{30} + \frac{(7272)}{30}}$$

$$= \sqrt{474} = 21.8$$

The value of D, the difference between the two means, is 201.7 - 160.5, or 41.2. This value of D is to be judged with reference to a hypothetical value of zero. Accordingly for T (the discrepancy expressed as a normal deviate) we have

$$T = \frac{41.2}{21.8} = 1.89$$

If the true value of D were zero, a discrepancy as great as this or greater would occur as a result of chance more than 5 times out of 100 as a result of random fluctuations of sampling. Therefore, even with a standard significance of 1 out of 20, one cannot reject the hypothesis that the two sample

SECRET

distributions of the statistic j could have been drawn from the same parent population.

However, one cannot conclude on the basis of the results of the above test that the function G_{BBB} is not a parameter affecting the incidence of depth. For if the probability F (P) of a pair of messages/being in depth is indeed a function of N (B) and of G_{BBB}, i.e. F (P) $= \phi[N(B), G_{BBB}]$, then the above test simply provides empirical evidence to support the statement that

$$\left.\frac{\partial F(P)}{\partial G_{BBB}}\right| \sim 0$$

N (B) = 48 percent saturation

It may very well be that if a similar test had been conducted for the case where N (B) $\leq$ 30 percent saturation

$\frac{\partial F(P)}{\partial G_{BBB}}$ might have assumed significant proportions.

There is no logical reason to assume that

$\frac{\partial F(P)}{\partial G_{BBB}}$ and $\frac{\partial F(P)}{\partial N(B)}$ are linear.

Therefore another test to determine the effect of G_{BBB} on the incidence of depth when N (B) $\sim$ 20 or 30 percent saturation has been conducted. In any case, since for N(B) max, $G_{BBB} \approx 0$, it is apparent that if G_{BBB} is a parameter affecting the incidence of depth, it is not a very significant parameter when N (B) $\geq$ 48 percent saturation.

From patterns IV and V the j-statistic was obtained. A sample size of 40 was used with a 22.53 percent saturation. The same theory that was used on

SECRET

pattern I and pattern II will be used to obtain the following results in the comparison of patterns IV and V

$$\sigma_D = \sqrt{\frac{\sigma_4^2}{N_4} + \frac{\sigma_5^2}{N_5}}$$

$$= \sqrt{\frac{(122.7575)^2 + (142.452)^2}{40}}$$

$$= \sqrt{884.690}$$

$$= 29.74$$

The value of D is computed:

$$M_5 - M_4 = 279.93 - 206.75 = 73.18$$

thus:

$$T = \frac{73.18}{29.74} = 2.46$$

From this T value one finds when compared with the normal curve or distribution, that this value places us at the 2 percent level of confidence of approximation of the normal curve. Since this sample of j-statistics so nearly fits the normal one may conclude that with the larger sample and the cut to 22.53 percent saturation that this sample does come from the parent population.

To provide some idea of just how sensitive a parameter N (B) is, a distribution of j statistics for a sample of size 20 was obtained. This distribution, labeled distribution III, is given in Appendix D. Notch pattern set III (listed in Apendix C) for which N (B) = 22.53 percent saturation, was issued to obtain this distribution. It will be noted that the mean for distribution III is significantly greater than that for distributions I and II.

A more detailed analysis of the patterns I, II, IV, V appears in Appendix E.

CONCLUSIONS:

Logical deduction supplemented in some instances by statistical inference strongly suggests that when N (B) is relatively large, i.e. N (B) > 50 percent saturation, the only important parameter affecting the incidence of depth is N (B). However, it is quite probable that G_{BBB} becomes a significant parameter when N (B) assumes values that are, in the probability sense, more normal, viz. 15 percent - 30 percent saturation. Additional statistical tests will have to be conducted before any conclusions can be made regarding this last possibility.

Because the incidence of depth is so dependent upon the value of N (B), it is suggested that the AFSAM 9 be utilized in such a manner as will preclude the possibility of N (B) ever exceeding, say 25 percent or 30 percent saturation. This could be easily accomplished if it were not objectionable to extend further the definition of admissability for notch patterns. For example, we could define an admissable "set" of patterns as one for which N (B) < 30 percent saturation for all possible subsets comprised of eight patterns. Only sets of patterns which are admissible in the above sense would then be distributed to each user, so that it would be impossible for any user to select from his set of patterns a subset of eight patterns for which N (B) > 30 percent saturation.

The above redefinition of admissibility would permit greater volume of traffic flow and would thus increase the efficiency of the device, provided, of course, that the security afforded by the AFSAM 9 is a function of notch pattern identification rather than of recovery.

~~SECRET~~

In this study no attempt was made to determine whether the incidence of depth is appreciably greater when message starts are restricted by an indicator system. The only data relevant to this question that are currently available are the statistical pictures in Appendix A. If probability of depth is not independent of the proposed indicator system, it would seem logical to conjecture that the alignment of the alphabet ring with respect to the notch pattern sequence is significantly affecting the incidence of depth.

To determine whether the alphabet ring alignment is a depth parameter, the following test is proposed. For a suitable set of notch patterns two distributions of a j - statistic are obtained. One distribution is obtained for the case for which the alignment of the alphabet rings minimizes the number of indicator settings that are origins. The other distribution is obtained for the case for which the alignment of the alphabet rings minimizes the number of indicator settings that are origins. The mean of the two distributions could then be compared to determine whether their difference is significant.

Whether or not it will be necessary to undertake any investigations to determine the degree of dependence of F (P) on the proposed indicator system has not yet been determined.

1 "Some Genealogical Characteristics of the Cycle Structure of the AFSAM 9 For a Particular Set of Notch Rings": R. P. Murphy and Warren Lotz, C84.2

2 "Notch Rings Suitable for Use in AFSAM 7 and AFSAM 9": Marguerite D. Newell, C90.0214

SECRET

Robert P. Murphy
NSA-314
March 1955

This paper was in draft form when Mr. Murphy resigned. Its preparation for publication has been the result of a co-operative effort of Marvin Bass, NSA-314, Bernard Harris, NSA-314 and Lt. William H. Cornelius, Jr., NSA-314.

~~UNCLASSIFIED~~

APPENDIX A

In order to ascertain whether the clustering of message starts in various parts of the cycle structure exhibit any "non-random" characteristics when quasi-random or indicator starts are employed, a sample of 1,000 random, 1,000 quasi-random, and 1,000 indicator starts were obtained. These were grouped according to points at which their successors entered a cycle. The frequencies of each group are presented in Table I.

To test the hypothesis that these three samples may be regarded as three samples from the same population, we construct the likelihood ratio test for homogeneity of samples from a multinomial distribution.

Let x_{ij} be the number of elements in the ith cell in the jth sample

$$(i=1,2,\ldots,m; j = 1,2,\ldots,k)$$

Let $\sum_i x_{ij} = n_j$ and $\sum_j n_j = n$

Forming the likelihood ratio, we get:

$$\lambda = \frac{\prod_i \left(\frac{\sum_j x_{ij}}{n} \right)^{\sum_j x_{ij}}}{\prod_{i,j} \left(\frac{x_{ij}}{n_j} \right)^{x_{ij}}}$$

$$\log \lambda = \sum_i \sum_j x_{ij} \log \left(\frac{\sum_i x_{ij}}{n} \right) - \sum_i \sum_j x_{ij} \log \left(\frac{x_{ij}}{n_j} \right)$$

$$-2 \log \lambda = 2 \left[\sum_i \sum_j x_{ij} \log \left(\frac{x_{ij}}{n_j} \right) - \sum_i \sum_j x_{ij} \log \left(\frac{\sum_j x_{ij}}{n} \right) \right]$$

~~SECRET~~

$$-2\log\lambda = \chi^2_{(k-1)(m-1)} = 2\sum_i\sum_j x_{ij}\left(\log\frac{x_{ij}}{n_j} - \log\sum_j\frac{x_{ij}}{n}\right)$$

We know that this quantity is approximately distributed as χ^2 with (k-1)(m-1) degrees of freedom. (Logarithms are base e)

Applying the above test, we find $\chi^2_0 = 17.62$, and we therefore accept the null hypothesis at the .05 level of significance.

In Table II, the average lengths of lead-ins to various cycle entry points are shown. Due to the non-orthogonality of the data, a regression analysis was not carried out at this time. Since the table suggests that random starts may give rise to longer lead-ins, some analysis of this type will be carried out at some future date. The data does not suggest that results from the quasi-random starts would lead to less depth than the random starts.

TABLE I

ENTRY	RANDOM	QUASI-RANDOM	INDICATOR
AA	142	147	128
AB	194	179	173
AC	60	64	58
BA	133	143	131
BB	68	72	83
BC	30	24	24
BD	42	30	45
BE	71	72	85
Residue	260	269	273

TABLE II

ENTRY	RANDOM	QUASI-RANDOM	INDICATOR
AA	311110	295353	272378
AB	384011	386665	373349
AC	183975	198623	177730
BA	289334	282557	283168
BB	316828	309518	300674
BC	200403	168741	185079
BD	240007	206810	228950
BE	323878	321169	346601

APPENDIX B

PERCENT SATURATION	PROBABILITY OF SELECTING A SET OF PATTERNS HAVING A HIGHER PERCENT SATURATION	PERCENT SATURATION	PROBABILITY OF SELECTING A SET OF PATTERNS HAVING A HIGHER PERCENT SATURATION
00	1.000	26	0.221
01	1.000	27	0.177
02	1.000	28	0.141
03	1.000	29	0.109
04	1.000	30	0.085
05	1.000	31	0.065
06	1.000	32	0.048
07	1.000	33	0.035
08	1.000	34	0.025
09	0.999	35	0.019
10	0.998	36	0.014
11	0.994	37	0.010
12	0.986	38	0.007
13	0.974	39	0.005
14	0.952	40	0.004
15	0.923	41	0.003
16	0.879	42	0.002
17	0.828	43	0.002
18	0.766	44	0.001
19	0.697	45	0.001
20	0.621	46	0.001
21	0.546	47	0.001
22	0.476	48	0.001
23	0.399	49	0.001
24	0.331	50	0.000
25	0.281	51 - 99	0.000

APPENDIX C

	PATTERN I								PATTERN II								PATTERN III							
	1	2	3	4	5	6	7	8	1	2	3	4	5	6	7	8	1	2	3	4	5	6	7	8
0	0	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0
1	1	1	0	1	1	1	1	1	1	1	1	0	0	0	0	0	0	1	1	1	1	0	0	1
2	0	1	1	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1
3	1	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	1	0	1	1	1
4	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	0	1	1	0
5	1	1	0	1	1	1	1	1	1	1	0	0	0	0	0	0	0	1	0	1	0	1	1	0
6	1	1	1	0	0	0	0	0	1	1	0	1	1	1	1	1	1	1	0	1	1	0	0	1
7	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	1	0	1	1	0	0
8	1	0	1	1	1	1	1	1	1	0	0	1	1	1	1	1	1	0	1	1	1	1	1	1
9	0	0	0	1	1	1	1	1	0	0	0	0	0	1	1	0	0	0	1	0	1	0	1	1
10	0	1	1	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	0	0	1	0	1	0
11	1	1	0	0	0	0	0	0	1	1	1	0	1	1	0	0	1	1	0	1	0	0	1	0
12	0	0	1	1	1	1	1	1	0	0	1	1	1	1	1	1	1	0	0	0	1	1	0	1
13	0	1	0	1	1	1	1	1	0	1	1	0	1	1	0	0	0	1	1	0	1	0	0	0
14	1	0	1	0	0	0	0	0	1	0	1	1	1	1	1	1	0	0	1	1	1	0	0	0
15	1	1	0	1	0	1	1	1	1	1	0	0	1	1	0	0	0	1	1	1	0	1	1	1
16	0	1	1	1	1	1	1	1	0	1	0	1	1	1	1	1	1	1	1	0	1	0	1	1
17	0	0	0	0	1	0	0	0	0	0	0	0	1	1	0	0	0	0	1	0	0	0	0	1
18	1	1	1	1	0	1	1	1	1	1	0	1	1	1	1	1	1	1	0	1	0	0	1	0
19	1	0	0	1	1	1	1	1	1	0	0	0	1	1	0	0	1	0	1	1	1	1	0	1
20	1	1	1	0	1	0	0	0	1	1	1	1	1	1	1	1	0	1	1	1	1	0	0	0
21	0	1	0	1	0	1	1	1	0	1	1	0	1	1	0	0	0	1	1	0	1	1	1	1
22	0	0	1	1	1	1	1	1	0	0	1	1	1	1	1	1	1	0	1	0	0	1	0	0
23	0	1	0	0	1	0	0	0	0	1	1	0	1	1	0	0	1	1	0	0	1	1	1	1
24	1	0	1	1	0	1	1	1	1	0	1	1	1	1	1	1	1	0	1	1	1	1	0	0
25	0	1	0	1	1	1	1	1	0	1	0	1	0	0	0	0	0	1	1	0	0	0	1	1
26	0	1	1	0	1	0	0	0	0	1	0	1	0	0	1	1	1	1	1	0	0	1	1	1
27	1	1	0	1	0	1	1	1	1	1	0	1	0	0	0	0	0	1	0	1	1	1	1	1
28	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	1	1	1	1	1	1	1	0	1
29	1	0	1	0	1	0	0	0	1	0	1	1	0	0	1	1	1	0	1	1	1	1	1	1
30	0	0	1	1	0	1	1	1	0	0	1	1	0	0	1	1	1	0	1	0	0	1	0	0
31	1	0	1	0	1	0	0	0	1	0	1	1	0	0	1	1	0	0	0	1	0	0	1	1
32	1	0	1	1	1	1	1	1	1	0	1	1	0	0	1	1	1	0	0	1	0	1	1	0
33	1	1	1	0	0	0	0	0	1	1	0	0	0	0	1	1	1	1	0	1	0	1	1	1
34	1	0	1	1	1	1	1	1	1	0	1	0	0	0	1	1	0	0	0	1	1	1	1	1
35	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	1	1	0	0	0

APPENDIX C

PATTERN IV

	1	2	3	4	5	6	7	8
0	0	1	1	1	1	1	1	1
1	1	1	1	0	0	0	0	0
2	0	1	1	1	1	1	1	1
3	1	0	1	0	0	0	0	0
4	1	1	1	1	1	1	1	1
5	1	1	0	0	0	0	0	0
6	1	1	1	1	1	1	1	1
7	0	1	0	0	0	0	0	0
8	1	0	0	1	1	1	1	1
9	0	0	0	0	0	0	0	0
10	0	1	1	1	1	1	1	1
11	1	1	1	0	0	0	0	0
12	0	0	1	1	1	1	1	1
13	0	1	1	0	0	0	0	0
14	1	0	1	1	1	1	1	1
15	1	1	0	0	1	1	0	0
16	0	1	0	1	1	1	1	1
17	0	0	0	1	1	1	1	1
18	1	1	0	1	1	1	1	1
19	1	0	0	1	1	1	1	0
20	1	1	1	1	1	1	1	1
21	0	1	1	1	1	1	1	0
22	0	0	1	1	1	1	1	1
23	0	1	1	1	1	1	1	1
24	1	0	1	1	1	1	1	1
25	0	1	0	1	1	1	1	1
26	0	1	0	1	1	1	1	1
27	1	1	0	1	1	1	1	1
28	1	1	1	1	0	0	1	1
29	1	0	1	0	0	0	0	1
30	0	0	1	0	0	0	0	1
31	1	0	1	0	0	0	0	1
32	1	0	1	0	0	0	0	0
33	1	1	0	0	0	0	0	0
34	1	0	1	0	0	0	0	0
35	1	0	0	0	0	0	0	0

PATTERN V

	1	2	3	4	5	6	7	8
	0	1	1	1	1	1	1	1
	1	1	0	1	1	1	1	1
	0	1	1	0	1	1	0	0
	1	0	0	0	0	0	0	0
	1	1	1	1	0	0	1	1
	1	1	0	1	1	1	1	1
	1	1	1	0	1	1	0	0
	0	1	0	0	1	1	0	0
	1	0	1	1	0	0	1	1
	0	0	0	1	0	0	1	1
	0	1	1	0	1	1	0	0
	1	1	0	0	1	1	0	0
	0	0	1	1	1	1	1	1
	0	1	0	1	0	0	1	1
	1	0	1	0	0	0	0	0
	1	1	0	0	1	1	0	1
	0	1	1	1	1	1	1	1
	0	0	0	1	1	1	1	0
	1	1	1	0	0	0	0	1
	1	0	0	0	0	0	0	1
	1	1	1	1	1	1	1	0
	0	1	0	1	1	1	1	1
	0	0	1	0	1	1	0	1
	0	1	0	0	0	0	0	0
	1	0	1	1	0	0	1	1
	0	1	0	1	1	1	1	1
	0	1	1	1	1	1	1	0
	1	1	0	0	1	1	0	1
	1	1	1	1	0	0	1	1
	1	0	1	1	0	0	1	0
	0	0	1	1	1	1	1	1
	1	0	1	0	1	1	0	0
	1	0	1	1	0	0	1	1
	1	1	1	1	0	0	1	0
	1	0	1	1	1	1	1	1
	1	0	0	0	0	0	0	0

(Both patterns have the same saturation as pattern 3, viz 22.53 percent)

APPENDIX D

(j - statistics for messages of length 5000)

	DISTRIBUTION I	DISTRIBUTION II	DISTRIBUTION III
1	234	230	197
2	203	199	381
3	189	154	355
4	305	283	187
5	69	118	314
6	150	237	209
7	283	75	263
8	220	214	355
9	138	144	415
10	182	146	350
11	39	151	214
12	108	105	126
13	182	132	290
14	365	20	207
15	157	192	404
16	144	159	359
17	275	85	261
18	231	126	27
19	209	379	383
20	229	106	295
21	299	187	5592
22	228	48	
23	99	162	
24	156	211	
25	176	174	
26	244	364	
27	136	122	
28	118	61	
29	434	58	
30	248	173	
	6050	4815	
	$\bar{j}$ = 201.67	$\bar{j}$ = 160.50	$\bar{j}$ = 279.6
	σ = 85.28	σ = 83.37	σ = 101.8

APPENDIX D

	DISTRIBUTION IV	DISTRIBUTION V
1	113	303
2	310	138
3	90	390
4	341	59
5	173	174
6	60	99
7	252	421
8	536	355
9	154	150
10	56	273
11	273	116
12	83	40
13	106	364
14	294	208
15	291	250
16	109	296
17	161	606
18	122	286
19	157	83
20	221	304
21	200	559
22	71	199
23	361	384
24	277	304
25	212	459
26	207	29
27	113	48
28	208	378
29	485	384
30	140	98
31	177	262
32	82	278
33	53	260
34	358	336
35	183	348
36	23	439
37	461	256
38	166	449
39	380	470
40	151	342

Distribution IV: $\bar{j} = 206.75$, $\sigma = 122.7575$

Distribution V: $\bar{j} = 279.925$, $\sigma = 142.542$

SECRET

APPENDIX E

SUMMARY AND RESULTS

I. The four sets of observations J_{11}, J_{12}, J_{21}, J_{22} were plotted on probability paper. No severe deviations from normality were noted. J_{11} corresponds to pattern I, J_{12} to pattern II, J_{21} to pattern IV, and J_{22} to pattern V.

II Sample means and sample variances were computed. The fact that the four sets of observations are heteroscedastic was noted and verified by Bartlett's Test establishing significance at the 1 percent level.

III An analysis of variance was carried out by inverse weighting of variances per Theorem of David and Neyman ("Extension of the Markoff Theorem on Least Squares," Statistical Research Memoirs., 2: 105-116 (1938)).

Let $H_0: \mu_{11} = \mu_{12} = \mu_{21} = \mu_{22}$

In accordance with the linear hypothesis write :

$$Y_{ijk} = \mu + \beta_i + \gamma_j + \varepsilon_{ijk}$$

and

$$\mu_{ij} = \mu + \beta_i + \gamma_j.$$

H_0 is then $\beta_i = \gamma_j = 0$.

β_i is the effect of saturation level, γ_i is the effect of the branching pattern within saturation level.

ε_{ijk} is the residual error; furthermore, ε_{ijk} is N $(0, \sigma^2_{ij})$. For convenience let $\Sigma \beta_i = 0$, $\Sigma \gamma_i = 0$. Let $\sigma_{ij}^2 = \sigma^2 / W_{ij}$ where W_{ij} are weights used to correct for heteroscedasticity. Designate sample estimates of

$\mu, \beta_i, \gamma_j, \sigma^2_{ij}$ by M, b, C, S. $\Sigma b_i = \Sigma C_i = 0$

In the Neyman-David Theorem, the W_{ij} are known constants. In the problem, they were derived from the data, a procedure which will tend to weaken the

significance level of the test somewhat. In view of the large number of degrees of freedom for estimation of each σ^2_{ij}, it is felt that this will not be too serious.

A) To determine the effect of μ and reduction of the total sum of squares under the null hypothesis.

Compute the least squares estimate of μ:

Assume: $Y_{ijk} = \mu + \varepsilon_{ijk}$

then $\hat{\mu} = M$

$$SSE = \sum_{ijk} W_{ij} (Y_{ijk} - M)^2$$

$$SSM = TSS - SSE \qquad TSS = \sum_{ijk} W_{ij} Y_{ijk}^2$$

$$\frac{\partial SSE}{\partial M} = -2 \sum_{ijk} W_{ij} (Y_{ijk} - M) = 0$$

$$\sum_{ijk} W_{ij} Y_{ijk} = M \sum_{i,j} N_{ij} W_{ij}$$

N_{ij} is the number of observations in group J_{ij}

$$M = \frac{\sum_{ijk} W_{ij} Y_{ijk}}{\sum_{ij} N_{ij} W_{ij}}$$

SECRET

$$SSE = \sum_{ijk} W_{ij} (Y_{ijk} - M)^2$$

$$= \sum_{ijk} W_{ij} Y_{ijk}^2 - 2M \sum_{ijk} W_{ij} Y_{ijk} + M^2 \sum_{ij} N_{ij} W_{ij}$$

$$= \sum_{ijk} W_{ij} Y_{ijk}^2 - M \left[2 \sum_{ijk} W_{ij} Y_{ijk} - M \sum_{ij} N_{ij} W_{ij} \right]$$

Since $M = \dfrac{\sum_{ijk} W_{ij} Y_{ijk}}{\sum_{ij} N_{ij} W_{ij}}$

$$SSE = \sum_{ijk} W_{ij} Y_{ijk}^2 - M \left[2 \sum_{ijk} W_{ij} Y_{ijk} - \frac{\sum_{ijk} W_{ij} Y_{ijk}}{\sum_{ij} N_{ij} W_{ij}} \sum_{ij} N_{ij} W_{ij} \right]$$

$$= \sum_{ijk} W_{ij} Y_{ijk}^2 - M^2 \sum_{ij} N_{ij} W_{ij}$$

$$SSM = M^2 \sum_{ij} N_{ij} W_{ij}$$

SECRET

This gives the partition of the sum of squares under the null hypothesis. Any effect due to β and γ must provide a further reduction of SSE. Since the data is non-orthogonal we can not estimate parameters separately but must repeat the entire process of obtain simultaneous estimates.

Let $H_1 = \gamma_j = 0$,

then $\mu_{11} = \mu_{12} = \mu_1$, $\mu_{21} = \mu_{22} = \mu_2$

and $Y_{ijk} = \mu + \beta_i + \varepsilon_{ijk}$

$$M' + b_i = \hat{\mu}_i$$

$$SSE' = \sum_{ijk} W_{ij} (Y_{ijk} - M' - b_i)^2$$

$$\frac{\partial SSE'}{\partial M} = -2 \sum_{ijk} W_{ij} (Y_{ijk} - M' - b_i) = 0$$

$$\frac{\partial SSE'}{\partial b_i} = -2 \sum_{jk} W_{ij} (Y_{ijk} - M' - b_i) = 0$$

$$\sum_{i,j,k} W_{ij} Y_{ijk} = M' \sum_{i,j} W_{ij} N_{ij} + \sum_{i,j} N_{ij} W_{ij} b_i$$

$$\sum_{j,k} W_{ij} Y_{ijk} = M' \sum_{j} N_{ij} W_{ij} + b_i \sum_{j} N_{ij} W_{ij}$$

SECRET

Since $b_2 = -b_1$ by virtue of $\sum_i b_i = 0$

$$\sum_{ijk} W_{ij} Y_{ijk} = M' \sum_{ij} W_{ij} N_{ij} + b_1 \sum_j (N_{1j} W_{1j} - N_{2j} W_{2j})$$

$$\sum_{j,k} W_{1j} Y_{1jk} = M' \sum_j N_{1j} W_{1j} + \sum_j N_{1j} W_{1j}$$

Solution of this set of equations provides the estimates b_1, b_2, M' for β_1, β_2, μ.

$$SSE' = \sum_{ijk} W_{ij} (Y_{ijk} - M' - b_i)^2$$

Since $M + b_i = \hat{\mu}_i$

$$SSE' = \sum_{ijk} W_{ij} (Y_{ijk} - \hat{\mu}_i)^2$$

$$= \sum_{ijk} W_{ij} (Y_{ijk}^2 - 2 Y_{ijk} \hat{\mu}_i + \hat{\mu}_i^2)$$

$$SSE' = \sum_{ijk} W_{ij} Y_{ijk}^2 - 2 \sum_{ijk} W_{ij} Y_{ijk} \hat{\mu}_i + \sum_{i,j} N_{ij} W_{ij} \hat{\mu}_i^2$$

$$SSE' = TSS - 2 \sum_{ijk} W_{ij} Y_{ijk} \hat{\mu}_i + \sum_{i,j} N_{ij} W_{ij} \hat{\mu}_i^2$$

SSE - SSE' is the reduction in sum of squares due to the introduction of b_1.

$F = \frac{\text{Reduction}}{SSE'/N_e}$ where N_e is the number of degrees of freedom due to error may be used to test the hypothesis $\beta_i = 0$.

Now consider $H_2: Y_{ijk} = \mu + \beta_i + \gamma_j + \varepsilon_{ijk}$

where β_i, not all zero and γ_j not all zero.

As before, compute the least squares estimates and the partition of the sum of squares.

$$SSE'' = \sum_{ijk} W_{ij} (Y_{ijk} - M'' - b_i' - C_j)^2$$

$$\frac{\partial SSE''}{\partial M''} = -2 \sum_{ijk} W_{ij} (Y_{ijk} - M'' - b_i' - C_j) = 0$$

$$\frac{\partial SSE''}{\partial b_i} = -2 \sum_{jk} W_{ij} (Y_{ijk} - M'' - b_i' - C_j) = 0$$

$$\frac{\partial SSE''}{\partial C_j} = -2 \sum_{ik} W_{ij} (Y_{ijk} - M'' - b_i' - C_j) = 0$$

$$\sum_{ijk} W_{ij} Y_{ijk} = M'' \sum_{i,j} N_{ij} W_{ij} + \sum_{i,j} W_{ij} N_{ij} b_i + \sum_{i,j} N_{ij} W_{ij} C_j$$

SECRET

$$\sum_{j,k} W_{ij} Y_{ijk} = M'' \sum_{j} W_{ij} N_{ij} + b'_i \sum_{j} W_{ij} N_{ij} + \sum_{j} W_{ij} N_{ij} C_j$$

$$\sum_{i,k} W_{ij} Y_{ijk} = M'' \sum_{i} W_{ij} N_{ij} + \sum_{i} W_{ij} N_{ij} b'_i + C_j \sum_{i} W_{ij} N_{ij}$$

Since $b'_2 = - b'_1$

and $C_2 = - C_1$

We get:

$$\sum_{i,j,k} W_{ij} Y_{ijk} = M'' \sum_{i,j} N_{ij} W_{ij} + b'_1 \sum_{j} (W_{1j} N_{1j} - W_{2j} N_{2j})$$

$$+ C_1 \sum_{i} (N_{i1} W_{i1} - N_{i2} W_{i2})$$

$$\sum_{j,k} W_{ij} Y_{ijk} = M'' \sum_{j} W_{ij} N_{ij} + b'_i \sum_{j} W_{ij} N_{ij} + C_1 (N_{11} W_{11} - N_{12} W_{12})$$

$$\sum_{i,k} W_{ij} Y_{ijk} = M'' \sum_{i} W_{ij} N_{ij} + b'_1 (W_{1j} N_{1j} - W_{2j} N_{2j}) + C_j \sum_{i} W_{ij} N_{ij}$$

The solution of these equations may be obtained using conventional matrix methods (e.g. Gauss - Doolittle, Crout, etc.).

The reduction in sum of squares due to error is obtained as follows:

$$SSE'' = \sum_{ijk} W_{ij} (Y_{ijk} - M'' - b'_i - C_j)^2$$

Designate $M'' + b'_i + C_j$ by $\hat{\mu}_{ij}$

$$SSE'' = \sum_{ijk} W_{ij} (Y_{ijk} - \hat{\mu}_{ij})^2$$

$$SSE'' = TSS - 2 \sum_{ijk} W_{ij} Y_{ijk} \hat{\mu}_{ij} + \sum_{ij} N_{ij} W_{ij} \hat{\mu}_{ij}^2$$

$SSE' - SSE''$ is the additional reduction in sums of squares due to error obtained by introducing C_j.

$F = \frac{\text{Reduction}}{SSE''/N_e}$ may be used to test the hypothesis $\gamma_{.j} = 0$.

ANALYSIS OF VARIANCE

Due to	d.f.	S.S.
Total Sum of Squares	140	6,760,514.731
M	1	5,209,152.502
SSE	139	1,551,362.229
Reduction due to b_1	1	96,491.847*
SSE'	138	1,454,870.382
Reduction due to C_j	1	87,352.187*
SSE"	137	1,367,518.195

* significant at .01 level.

The hypothesis $H_0: \mu_{11} = \mu_{12} = \mu_{21} = \mu_{22}$ is then rejected at the .01 level and the "F-test" indicates a significant effect due to both saturation level and branching pattern for these four sets of observations.

SUMMARY OF COMPUTATIONS

	J_{11}	J_{12}	J_{21}	J_{22}	Total
ΣY_{ijk}	4,815	6,050	8,270	11,197	30,332
ΣY_{ijk}^2	974,353	1,430,970	2,312,602	3,947,049	8,664,974
W_{ij}	1.439	1.375	.647	.480	
N_{ij}	30	30	40	40	140
$W_{ij} N_{ij}$	43.170	41.250	25.880	19.200	129.500
$\Sigma W_{ij} Y_{ijk}$	6,928.785	8,318.750	5,350.690	5,374.560	25,972.785
$\Sigma W_{ij} Y_{ijk}^2$	1,402,093.967	1,967,583.750	1,496,253.494	1,894,583.520	6,760,514.731

M			200,562		
M'			209.265		
b_i		- 28.650		28.650	
$\hat{\alpha}_i$		180.615		237.915	
M''			211.494		
b'_i		-30.286		30.286	
C_j		-26.076		26.076	
$\hat{\mu}_{ij}$	155.132	207.284	215.704	267.856	
$\hat{\mu}_{ij} W_{ij} N_{ij}$	6,697.04844	8550.465	5582,41952	5142.8352	

SSE 1,551,362.229

SSM 5,209,152.502

SSE' 1,454,870.382

SEE" 1,367,518.195

BARTLETT'S TEST

				S_i/N_i	$\log S_i/N_i$	$N_i \log S_i/N_i$
S_1 =	201,545.50	N_1	29	6949.84	8.84806	256.59374
S_2 =	210,866.50	N_2	29	7271.26	8.89467	257.94543
S_3 =	602,779.50	N_3	39	15455.88	9.64750	376.25250
S_4 =	812,672.79	N_4	39	20837.76	9.94633	387.90687
S =	1,827,864.29	N =	136	13440.18	9.50771	1293.04856

$$N \log \frac{S}{N} - \Sigma N_i \log \frac{S_i}{N_i}$$

$$= 1293.04856 - 1278.69854$$

$$= 14.35002$$

$$S_i = \Sigma X_i^2 - N \bar{X}_i^2$$

$$C = 1 + \frac{1}{6}\left(\frac{1}{29} + \frac{1}{29} + \frac{1}{39} + \frac{1}{39} - \frac{1}{136}\right)$$

$$= (.121786 - .007353)$$

$$= 1 + .01907 = 1.01907$$

$$M = 14.0815$$

M is significant at .01 level.