

PARAMETRIC ANALYSIS OF STATISTICAL COMMUNICATION NETS*

BY

H. FRANK (*University of California, Berkeley*) AND S. L. HAKIMI (*Northwestern University*)

Abstract. The existing traffic within the branches of a communication net can often be assumed to be normally distributed random variables. A natural problem is to determine the probability that a particular flow rate between a pair of stations can be attained. If this probability is too small, it is necessary to improve the net with minimum cost. In this paper, analysis techniques on which effective synthesis procedures can be based are developed. An exact method for evaluating the flow rate probability is obtained as well as upper and lower bounds. Monte Carlo techniques are applied and the flow rate is seen to be approximately normally distributed. A method of finding the approximate mean and variance of the flow rate is given, as well as a Uniformly Most Powerful Invariant Statistical test.

I. Introduction¹. Let G be a communication network with v vertices v_1, v_2, \dots, v_v , and b branches b_1, b_2, \dots, b_b . It is natural to assume that the existing flows or traffic within the branches of G are random variables with possibly known probability distributions. Given a pair of stations v_i and v_j of G , we may want to determine the *probability* that a flow rate F_{ij} of at least R units can be established between v_i and v_j . When this probability (written $\text{Prob} \{F_{ij} \geq R\}$) is found, it may be that it is too small to meet the demands on the system. We must then increase the *capacities* of the branches of G (with minimum cost) until a prespecified probability level p_0 is reached.

The problem of finding $\text{Prob} \{F_{ij} \geq R\}$ in the case where the *joint probability density* of the network branch flow vector \mathfrak{F} is known has been solved by the authors [1]. There, the authors give a solution which requires the inversion of a multidimensional characteristic function. On the other hand, no information concerning the probability distributions of the branch flows may be available, but a set of n time observations of these flows may be known. In this case, a Uniformly Most Powerful level α test for testing the hypothesis

$$H_1 : p = \text{Prob} \{F_{ij} \geq R\} \geq p_0 \quad (1a)$$

against

$$H_2 : p < p_0 \quad (1b)$$

can be given [2]. Furthermore, if hypothesis H_1 is rejected, a procedure to increase the branch capacities of G with minimum cost until H_1 is accepted is also developed in [2]. This synthesis procedure is optimum in several senses: it makes full use of the available data, minimizes the probability of error, and minimizes total (linear) cost.

We can write $\text{Prob} \{F_{ij} \geq R\}$ in closed form when the density of \mathfrak{F} is known [1].

*Received September 26, 1966; revised manuscript received June 22, 1967. This work was supported by the U. S. Air Force Office of Scientific Research and the U. S. Army Research Office, Durham, N. C., Grant No. DA-ARO-D-31-124-G576.

¹The notation used in this paper is identical to the notation of our previous work [1], [2].

However, the computational problems encountered in evaluating the necessary expressions are formidable, even for a digital computer. It appears unlikely that, in general, practical *parametric* synthesis techniques based on this method of analysis are obtainable. The purpose of this paper is to develop the mathematical apparatus on which efficient synthesis techniques can be based. The application of this apparatus to the optimum synthesis of statistical communication nets is discussed in another paper [3].

II. Probability computations with normally distributed branch flows. One of the most common assumptions in probabilistic approaches to physical problems is the "normality" (Gaussian) assumption. This assumption is usually based on physical observations and some application of the Central Limit Theorem. In our problem, we will assume that the branch flows are normally distributed random variables. This assumption is justifiable on the following grounds: (1) In the case of telephone systems, branch flow has been observed to be Poisson distributed [4]. The normal approximation to a Poisson variable is usually adequate for most purposes. (2) A branch flow is actually the sum of a large number of independent random variables, since each flow consists of the contributions of a large number of subscribers. These independent variables may be considered to assume two values, "0" and "1". Then, it can be shown (see Theorem 6.9.3 [5]) that the limit distribution of the standardized sum of the variables is normal, and thus, if a large number of users have access to the same line, the flow distribution is approximately normal.

Before proceeding, we must realize the limitations of the normality assumption. If the random flow F_k in branch b_k is normally distributed with mean μ_k and variance σ_k^2 (written $N(\mu_k, \sigma_k^2)$), the probability that an additional flow x may be established in b_k is

$$\text{Prob} \{c_k - F_k \geq x\} = 1 - \int_0^x \frac{1}{(2\pi)^{1/2} \sigma_k} \exp \left\{ -\frac{1}{2} \frac{[y - (c_k - \mu_k)]^2}{\sigma_k^2} \right\} dy, \quad (2)$$

where c_k is the capacity of b_k . The physical constraint that flow F_k lies in the interval $[0, c_k]$ requires that the mean and variance of F_k is such that the "tails" of the approximation are negligible. In other words,

$$\int_{-\infty}^0 \frac{1}{(2\pi)^{1/2} \sigma_k} \exp \left\{ -\frac{1}{2} \frac{[y - \mu_k]^2}{\sigma_k^2} \right\} dy \doteq 0, \quad k = 1, 2, \dots, b, \quad (3a)$$

and

$$\int_{c_k}^{\infty} \frac{1}{(2\pi)^{1/2} \sigma_k} \exp \left\{ -\frac{1}{2} \frac{[y - \mu_k]^2}{\sigma_k^2} \right\} dy \doteq 0. \quad k = 1, 2, \dots, b. \quad (3b)$$

If the above equations do not hold, we must deal with *truncated* distributions (Sec. 19.3 of [6]). The implications of this statement are that even though the *demand* on a line is normally distributed, the actual distribution of branch flow may not be. For example, this is the case when a large number of people randomly sample a line of small capacity according to a normal distribution with a large mean. Although the demand on the line is normal, we can expect that the flow in the line will always be close to the line's capacity.

We have not yet completely specified the probabilistic structure of the model since we have not stated the relationship between F_1, F_2, \dots, F_b . In this treatment, we will *not* assume that the F_i 's are independent. Instead, we will make the more general assumption that the joint distribution of $\mathcal{F} \triangleq (F_1, F_2, \dots, F_b)'$ may be approximated

by a b -dimensional, nonsingular normal distribution [6] with mean vector \mathbf{u} and variance-covariance matrix Σ (written $N(\mathbf{u}, \Sigma)$). Therefore, the probability density of \mathcal{F} is

$$p(f_1, f_2, \dots, f_b) = \frac{1}{(2\pi)^{b/2} |\Sigma|^{1/2}} \exp \left\{ -\frac{1}{2}(\mathbf{f} - \mathbf{u})' \Sigma^{-1}(\mathbf{f} - \mathbf{u}) \right\} \tag{4}$$

where $\mathbf{f} = (f_1, f_2, \dots, f_b)'$, $|\Sigma|$ represents the determinant of Σ , and $'$ indicates transpose.

The available capacity C_k of branch b_k is a random variable given by $C_k \triangleq c_k - F_k$. Thus, it is easy to see that the random capacity vector $\mathbf{C} = (C_1, C_2, \dots, C_b)'$ is $N(\mathbf{c} - \mathbf{u}, \Sigma)$. We want to find $\text{Prob} \{F_{i,j} \geq R\}$. If \mathbf{C} were a fixed vector, the maximum flow $F_{i,j}$ would be given by Ford and Fulkerson's Max Flow-Min Cut Theorem [7] which states that flow $F_{i,j} \geq R$ is attainable in G if and only if

$$F_{i,j} = \min (|A_1|, \dots, |A_q|) \geq R \tag{5}$$

where A_1, \dots, A_q are the set of basic cut-sets of G which separate v_i and v_j , and $|A_k|$ is the value of A_k obtained by adding the capacities of the branches in A_k . However, in our problem C_1, \dots, C_b are random variables and consequently $|A_1|, \dots, |A_q|$ are also random. Clearly,

$$\begin{aligned} \text{Prob} \{F_{i,j} \geq R\} &= \text{Prob} \{ \min (|A_1|, \dots, |A_q|) \geq R \} \\ &= \text{Prob} \{ |A_1| \geq R, \dots, |A_q| \geq R \}. \end{aligned} \tag{6}$$

Therefore, we must compute the joint density of the random vector $|\mathbf{A}| \triangleq (|A_1|, \dots, |A_q|)'$.

The random vectors $|\mathbf{A}|$ and \mathbf{C} are related by the basic cut-set matrix $B = [b_{ij}]$ (B is a $q \times b$ matrix of "0's" and "1's" such that $b_{ij} = 1$ if and only if branch b_i is in cut-set A_j) through the equation

$$|\mathbf{A}| = B\mathbf{C}. \tag{7}$$

Thus, $|\mathbf{A}|$ is obtained from \mathbf{C} by means of a *linear* transformation. It is well known that linear transformations of normal variables are themselves normal variables.² In fact, since \mathbf{C} is $N(\mathbf{c} - \mathbf{u}, \Sigma)$, $|\mathbf{A}|$ can be shown to be $N(B(\mathbf{c} - \mathbf{u}), B\Sigma B')$ [6].

If the rank of B (over the real field) is r and $r < q$, the variance-covariance matrix $B\Sigma B'$ will be singular (positive semidefinite). This means that, over the real field, some of the cut-sets may be expressed as linear combinations of others. To see this, consider the graph G_1 shown in Fig. 1. The basic cut-set matrix of G_1 is

$$B = \begin{matrix} & \begin{matrix} b_1 & b_2 & b_3 & b_4 & b_5 & b_6 & b_7 & b_8 \end{matrix} \\ \left. \begin{matrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 & 1 & 0 & 1 & 1 \end{matrix} \right\} & \begin{matrix} A_1 \\ A_2 \\ A_3 \\ A_4 \\ A_5 \\ A_6 \\ A_7 \end{matrix} \end{matrix} \tag{8}$$

²This gives us a further basis for our initial normality assumption. Even if the branch flows are not normal variables, the cut-set variables will, under mild conditions, tend to normal variables. Furthermore, as the complexity of the net increases, the number of branches in a cut-set increases and so the approximation improves.

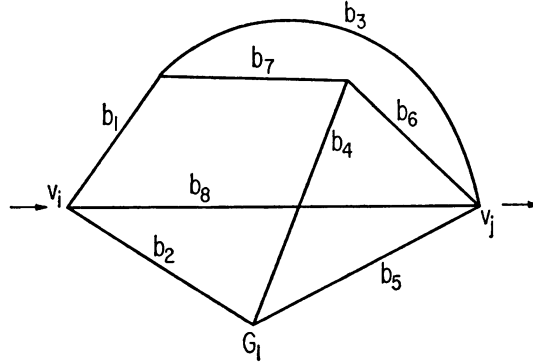


FIG. 1. Graph \$G_1\$.

and it is easily seen that \$|A_7|\$ can be expressed as a linear combination of \$|A_1|\$, \$|A_2|\$, and \$|A_3|\$. In fact

$$|A_7| = |A_2| + |A_3| - |A_1|. \tag{9}$$

In this case, \$[B\Sigma B']^{-1}\$ does not exist. Then, we can reorder the \$|A_i|\$ such that the rows of \$B\$ corresponding to \$A_1, \dots, A_r\$ are linearly independent. There exist constants \$a_{k,i+h}\$ (\$k = 1, \dots, r; h = 1, \dots, q - r\$) such that

$$|A_h| = \sum_{k=1}^r a_{k,i+h} |A_k|, \quad h = 1, \dots, q - r. \tag{10}$$

Let \$\hat{B}\$ be the \$r \times b\$ matrix with respect to the reordered \$A_i\$'s whose rows correspond to \$A_1, A_2, \dots, A_r\$. Then

$$\begin{Bmatrix} |A_1| \\ \vdots \\ |A_r| \end{Bmatrix} = \hat{B}C \tag{11}$$

and we can write the probability that the maximum flow is at least \$R\$ as

$$\begin{aligned} \text{Prob} \{F_{ii} \geq R\} &= \int \cdots \int_{\Omega} \frac{1}{(2\pi)^{r/2} |\hat{B}\Sigma\hat{B}'|^{1/2}} \\ &\quad \cdot \exp \left\{ -\frac{1}{2} [\mathbf{x} - \hat{B}(\mathbf{c} - \mathbf{u})]' [\hat{B}\Sigma\hat{B}']^{-1} [\mathbf{x} - \hat{B}(\mathbf{c} - \mathbf{u})] \right\} d\mathbf{x}, \end{aligned} \tag{12}$$

where \$\mathbf{x} = (x_1, \dots, x_r)'\$, \$d\mathbf{x} = dx_1 dx_2 \cdots, dx_r\$, and \$\Omega\$ is the convex region defined by the inequalities

$$\begin{aligned} x_i &\geq R, \quad i = 1, \dots, r \\ \sum_{k=1}^r a_{k,i+h} x_k &\geq R, \quad h = 1, \dots, q - r \end{aligned}$$

This expression is complicated, but the number of integrals to be evaluated is \$r\$ and \$r \leq b\$ (the number of branches of \$G\$).

We can also express \$\text{Prob} \{F_{ii} \geq R\}\$ as the product of integrals of independent normal densities. If the rank of \$B\Sigma B'\$ is \$r\$, each \$|A_k|\$ can be written as the sum of \$r\$ in-

dependent normal variables, say Z_1, \dots, Z_r . Variables Z_1, \dots, Z_r are related to $|A_1|, \dots, |A_r|$ by means of a linear transformation. The transformation can be found by using the method of Jacobi [8]. First, we Gauss reduce the matrix $\hat{B}\Sigma\hat{B}'$ to the upper triangular matrix T . Then, it can be shown that

$$\begin{Bmatrix} Z_1 \\ \vdots \\ Z_r \end{Bmatrix} = T'^{-1} \begin{Bmatrix} |A_1| \\ \vdots \\ |A_r| \end{Bmatrix} \quad \text{or} \quad \begin{Bmatrix} |A_1| \\ \vdots \\ |A_r| \end{Bmatrix} = T' \begin{Bmatrix} Z_1 \\ \vdots \\ Z_r \end{Bmatrix} \tag{13}$$

For $k > r$, since we can write $|A_k|$ as a linear combination of $|A_1|, \dots, |A_r|$, with the help of Eq. (13) we can easily write $|A_k|$ as a linear combination of Z_1, \dots, Z_r . From Eq. (13), we can find the means and variances of $Z = (Z_1, \dots, Z_r)'$ which is $N(T'^{-1}\hat{B}(\mathbf{c} - \mathbf{u}), T'^{-1}\hat{B}\Sigma\hat{B}'T^{-1})$. Note that $T'^{-1}\hat{B}\Sigma\hat{B}'T^{-1}$ is a diagonal matrix. If we then standardize these variables, we obtain

$$\text{Prob} \{F_{ij} \geq R\} = \int \phi(y_1) dy_1 \int_{\Omega'} \phi(y_2) dy_2 \cdots \int \phi(y_r) dy_r, \tag{14}$$

where $\phi(y)$ is the standard normal density function and Ω' is the convex region defined by a set of inequalities of the form

$$\sum_{k=1}^r s_{ik}y_k \geq k_i \quad j = 1, 2, \dots, q,$$

and the k_i 's and s_{ik} 's are known constants.

To compute $\text{Prob} \{F_{ij} \geq R\}$, we must evaluate a probability integral of a multi-dimensional normal distribution. This problem is of considerable interest to statisticians and a number of papers have been written on the subject. An excellent review of the progress in this area is given in a paper by S. Gupta [9]. We will briefly summarize some special results.

If $|A|$ is $N(B(\mathbf{c} - \mathbf{u}), B\Sigma B')$, the correlation matrix of $|A|$ is a $q \times q$ matrix

$$P^{|A|} = [\rho_{ij}^{|A|}], \tag{15}$$

where

$$\rho_{ij}^{|A|} = \frac{\beta_{ij}}{\beta_{ii}^{1/2}\beta_{jj}^{1/2}},$$

and β_{ij} is the $(i - j)$ th entry of $B\Sigma B'$. For example, if Σ is the identity matrix, the $(i - j)$ th entry of $P^{|A|}$ is

$$\rho_{ij}^{|A|} = \frac{n_{ij}}{n_{ii}^{1/2}n_{jj}^{1/2}} \tag{16}$$

where n_{ij} is the number of branches in $A_i \cap A_j$.

For simplicity, let us standardize $|A_1|, \dots, |A_q|$ and consider the random variables Y_1, \dots, Y_q defined by

$$Y_i = (|A_i| - E |A_i|)/(\text{Var } |A_i|)^{1/2}, \quad i = 1, 2, \dots, q. \tag{17}$$

Then, if $\rho_{\min} = \min_{i,j} \rho_{ij}$ and $\rho_{\max} = \max_{i \neq j} \rho_{ij}$, the following upper and lower bounds can be shown to hold³ for

³Usually the condition $\rho_{\min} \geq 0$ is required, but in our case Y_1, \dots, Y_q are always nonnegatively correlated.

$$\text{Prob } \{|A_1| \geq R, \dots, |A_q| \geq R\} = \text{Prob } \{Y_1 \geq R_1, \dots, Y_q \geq R_q\}, \quad (18)$$

where

$$R_i \triangleq \frac{R - E |A_i|}{(\text{Var } |A_i|)^{1/2}} = \frac{R - \sum_{j=1}^b b_{ij}(c_j - \mu_j)}{(\beta_{ii})^{1/2}},$$

$$\int_{-\infty}^{\infty} \left[\prod_{i=1}^q \Phi \left(\frac{(\rho_{\min})^{1/2} y - R_i}{(1 - \rho_{\min})^{1/2}} \right) \right] \phi(y) dy \leq \text{Prob } \{F_{ij} \geq R\}$$

$$\leq \int_{-\infty}^{\infty} \prod_{i=1}^q \left[\Phi \left(\frac{(\rho_{\max})^{1/2} y - R_i}{(1 - \rho_{\max})^{1/2}} \right) \right] \phi(y) dy, \quad (19)$$

when $\Phi(\cdot)$ is the cumulative distribution function of the standard normal variable.

Suppose that the correlation coefficients can be written as

$$\rho_{ij} = \alpha_i \alpha_j \quad \forall i, j \quad (i \neq j). \quad (20)$$

This condition holds if, for example, the number of branches in $A_i \cap A_j$ is constant. Then, an *exact* expression for $\text{Prob } \{F_{ij} \geq R\}$ is

$$\text{Prob } \{F_{ij} \geq R\} = \int_{-\infty}^{\infty} \left[\prod_{i=1}^q \Phi \left(\frac{\alpha_i y - R_i}{(1 - \alpha_i^2)^{1/2}} \right) \right] \phi(y) dy. \quad (21)$$

If the number of branches n_{ij} in $A_i \cap A_j$ is not constant, but small compared to the n_{ii} , the preceding expression gives an excellent approximation to the actual probability.

As another special case, suppose that the expected value of each cut-set is equal to a constant, say t . Thus

$$\sum_{j=1}^b b_{ij}(c_j - \mu_j) = t, \quad i = 1, \dots, q. \quad (22)$$

We may want to find the probability that the maximum flow exceeds the expected value t . This problem reduces to one of finding the probability of the positive quadrant, and has been attacked by several authors. For example, McFadden gives the approximation

$$\text{Prob } \{F_{ij} \geq t\} \doteq 2^{-q} \left(1 + \frac{2}{\pi} \sum_{i>j \geq 1} \arcsin \rho_{ij} + \frac{4}{\pi^2} \sum_{h>k > i > j \geq 1} \rho_{ij} \rho_{hk} + \rho_{ik} \rho_{jh} + \rho_{ih} \rho_{jk} \right). \quad (23)$$

To clarify the above ideas, again consider the network G_1 shown in Fig. 1. The number of branches in $A_i \cap A_j$, $i, j = 1, \dots, 7$, are

$$\begin{aligned} n_{11} &= 3 \\ n_{12} &= 2 \quad n_{22} = 4 \\ n_{13} &= 2 \quad n_{23} = 1 \quad n_{33} = 4 \\ n_{14} &= 2 \quad n_{24} = 3 \quad n_{34} = 2 \quad n_{44} = 5 \\ n_{15} &= 2 \quad n_{25} = 2 \quad n_{35} = 3 \quad n_{45} = 2 \quad n_{55} = 5 \end{aligned}$$

$$\begin{aligned}
 n_{16} &= 1 & n_{26} &= 2 & n_{36} &= 2 & n_{46} &= 3 & n_{56} &= 3 & n_{66} &= 4 \\
 n_{17} &= 1 & n_{27} &= 3 & n_{37} &= 3 & n_{47} &= 3 & n_{57} &= 3 & n_{67} &= 3 & n_{77} &= 5
 \end{aligned}$$

Consequently, if the variance-covariance matrix of the branch flows is the identity matrix, the correlation matrix of the cut-set vector $|A|$ is

$$P^{1A1} = \begin{bmatrix} 1 & \frac{1}{3^{1/2}} & \frac{1}{3^{1/2}} & \frac{2}{15^{1/2}} & \frac{2}{15^{1/2}} & \frac{1}{2(3)^{1/2}} & \frac{1}{15^{1/2}} \\ & 1 & \frac{1}{4} & \frac{3}{2(5)^{1/2}} & \frac{1}{5^{1/2}} & \frac{1}{2} & \frac{3}{2(5)^{1/2}} \\ & & 1 & \frac{1}{5^{1/2}} & \frac{3}{2(5)^{1/2}} & \frac{1}{2} & \frac{3}{2(5)^{1/2}} \\ & & & 1 & \frac{2}{5^{1/2}} & \frac{3}{2(5)^{1/2}} & \frac{3}{5} \\ & & & & 1 & \frac{3}{2(5)^{1/2}} & \frac{3}{5} \\ & & & & & 1 & \frac{3}{2(5)^{1/2}} \\ & & & & & & 1 \end{bmatrix} \tag{24}$$

and $\rho_{\min} = \frac{1}{4}$, $\rho_{\max} = 3/2(5)^{1/2}$. Therefore,

$$\begin{aligned}
 &\int_{-\infty}^{\infty} \left[\prod_{i=1}^7 \Phi\left(\frac{1}{3}y - \frac{2}{3^{1/2}}R_i\right) \right] \phi(y) dy \\
 &\leq \text{Prob} \{F_{ij} \geq R\} \\
 &\leq \int_{-\infty}^{\infty} \left[\prod_{i=1}^7 \Phi\left(\left(\frac{3}{2(5)^{1/2} - 3}\right)^{1/2} y - \frac{R_i}{(1 - 3/2(5)^{1/2})^{1/2}}\right) \right] \phi(y) dy \tag{25}
 \end{aligned}$$

III. Monte Carlo simulation with normally distributed flows. The computation of the maximum flow rate probability, even in the case of normally distributed branch flows, seems to be quite tedious. The nature of our exact results indicates that we should search for a qualitative picture of the probabilistic terminal capacity behavior. We can generate such a picture using Monte Carlo techniques.

Basically, we simulate the system a large number of times by randomly generating a set of branch flow vectors $\{\mathfrak{F}(k); k = 1, \dots, n\}$. We then apply the Max Flow-Min Cut Theorem via the relation

$$F_{ij}(k) = \min_{\text{over rows}} B(\mathbf{c} - \mathfrak{F}(k)) \tag{26}$$

to find the maximum flow through the graph at the k th simulation. We repeat this process n times and then form the Empirical Distribution Function $S_n(z)$ defined by

$$S_n(z) = \frac{1}{n} \sum_{k=1}^n h_z(F_{ij}(k)), \quad (27)$$

where

$$h_z(U) = \begin{cases} 0, & U \geq z \\ 1, & U < z. \end{cases}$$

Thus, $nS_n(z)$ is the number of $F_{ij}(k)$ that are smaller than z . By Glivenko's Theorem (Theorem 10.10.1 of [5]), we know that $S_n(z)$ converges as $n \rightarrow \infty$, in a very strong sense, to the true distribution of F_{ij} . Hence, if we pick n large enough, we can plot a distribution function which will be an excellent approximation to the theoretical distribution of F_{ij} .

A computer program was written to perform the above simulation. For simplicity, the branch flows were assumed to be identically and independently⁴ distributed normal variables (the effect of truncation was then included). Branch capacity was treated as a variable parameter, and for a fixed capacity vector $\mathbf{c} = (c_1, \dots, c_b)'$, a graph of $1 - S_n(z)$ consisting of 100 sample values of maximum flow, was drawn. In some cases, the capacity of each branch was taken to be a constant c , then c was varied over a wide range of values. In other cases, the capacities of a subset of branches of G were varied while the other branches were held fixed. Network G_1 shown in Fig. 1 and network G_2 shown in Fig. 2 were two of the graphs analyzed in this manner. In each case, the means of the branch flows were taken to be 3 and the variances were assumed to be unity. Branch capacity was then varied between $c = 3$ and $c = 10$. Figures 3-8 show some of the results of the simulation. One remarkable (and surprising) result that emerged from the simulation was that *in every case, $S_n(z)$ could be accurately approximated by a cumulative normal distribution function*. This observation is highly significant and is illustrated in Figs. 9 and 10 where approximations to a number of the curves that appear in Figs. 3-8 are given. The consequences of this observation are examined in the following sections.

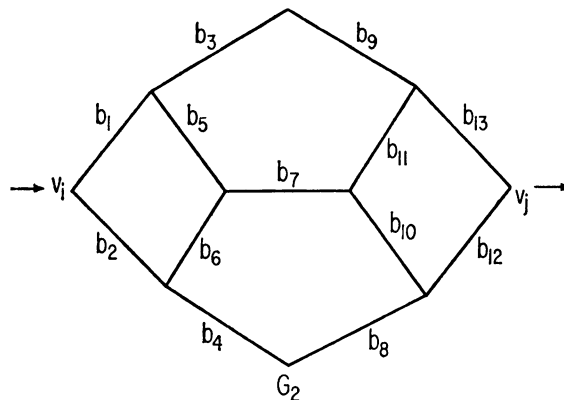


FIG. 2. Graph G_2 .

⁴The independence of branch flows does not change the generality of the simulation since the flow limitation is imposed by the cut-sets which are always dependent (see Eq. (12)).

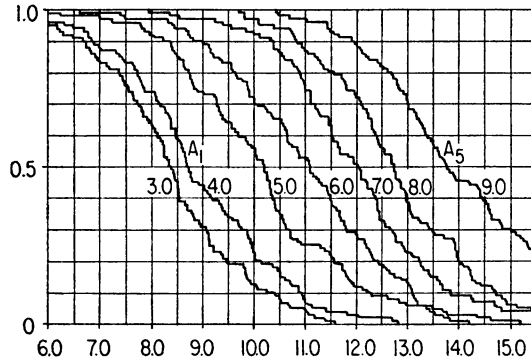


FIG. 3. Branch b_8 of G_1 is varied from $c_8 = 3.0$ to $c_8 = 9.0$. All other branch capacities equal 7.

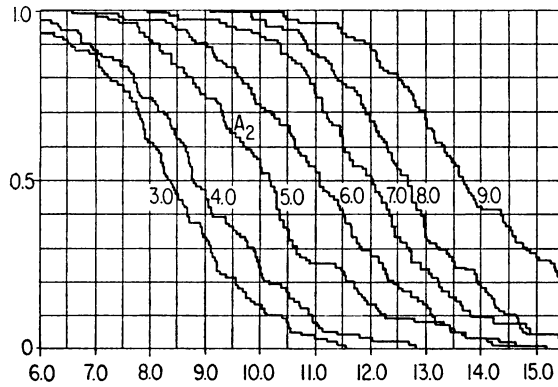


FIG. 4. Branch b_1 of G_1 is varied from $c_1 = 3.0$ to $c_1 = 9.0$. All other branch capacities equal 7.

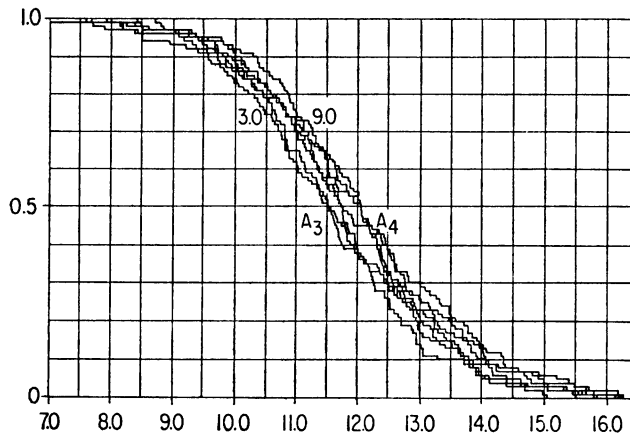


FIG. 5. Branch b_7 of G_1 is varied from $c_1 = 3.0$ to $c_7 = 9.0$. All other branch capacities equal 7.

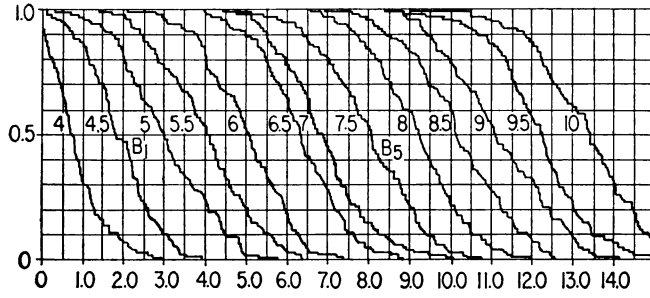


FIG. 6. All branches of G_2 are varied from $c = 4$ to $c = 10.0$.

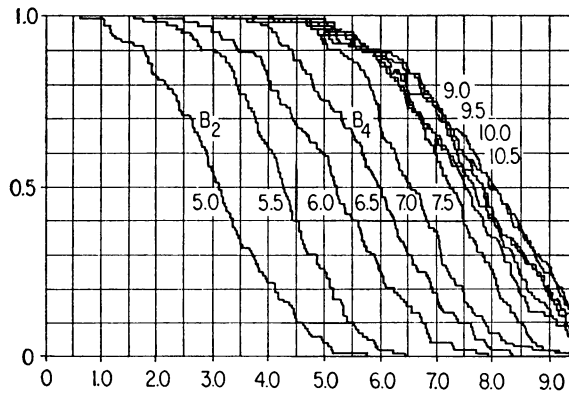


FIG. 7. Branch b_1, b_2, b_{12}, b_{13} of G_2 are varied from $c = 5$ to $c = 10.5$. All other branch capacities equal 6.

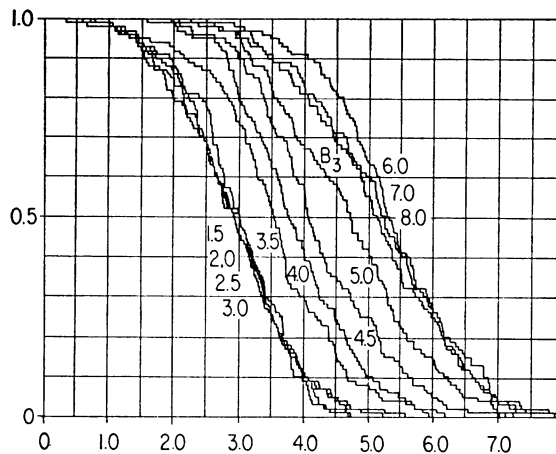


FIG. 8. Branches b_3, b_4, b_8, b_9 of G_2 are varied from $c = 1.5$ to $c = 8.0$. All other branch capacities equal 6.

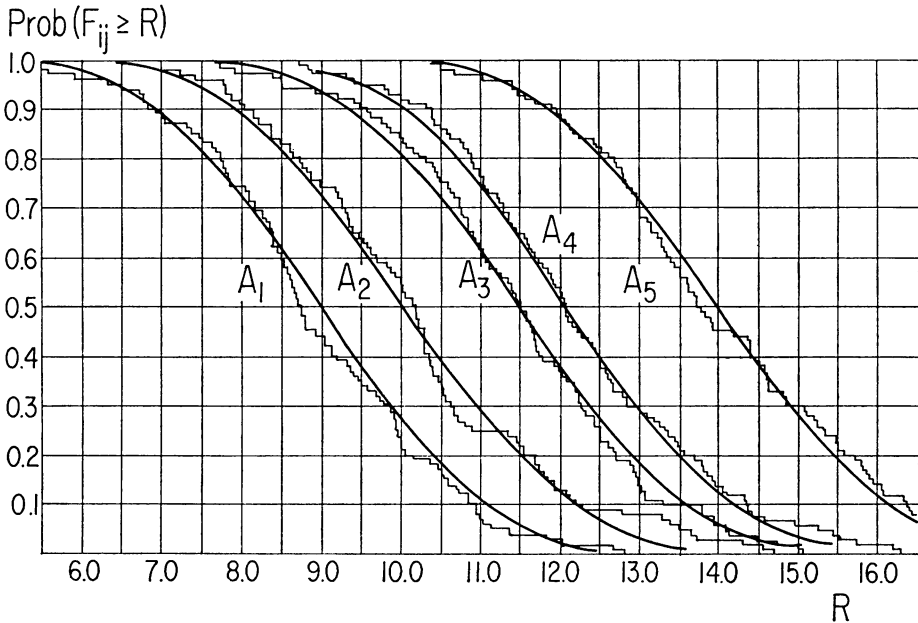


FIG. 9. The empirical curves labeled A1–A5 in Figs. 3–5 are reproduced and approximated with normal distribution functions with variance ≈ 2.7 . Note that A_3 and A_4 represent the empirical curves of the lowest and highest probabilities for given R , for the graphs shown in Fig. 5.

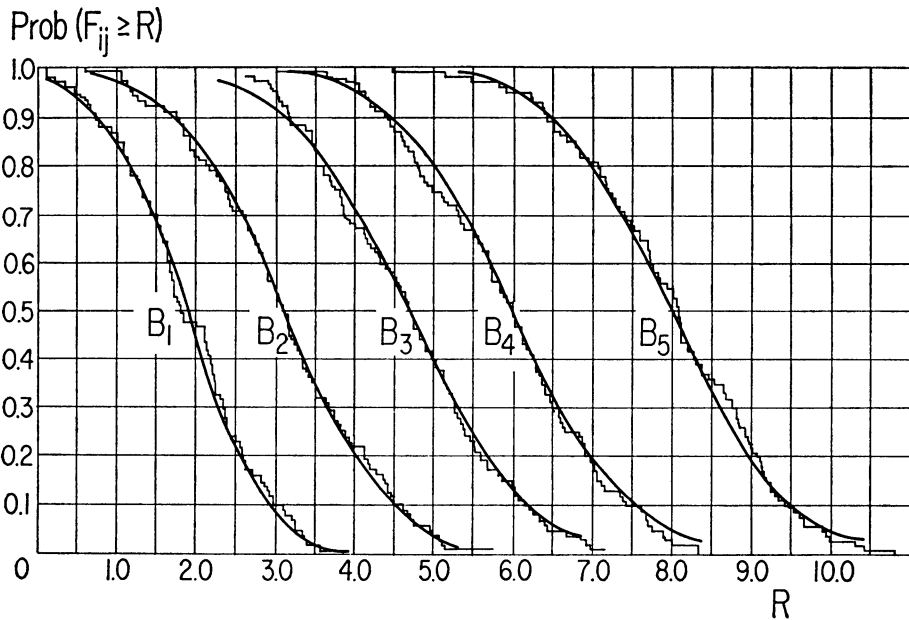


FIG. 10. The empirical curves labeled B1–B5 in Figs. 6–8 are reproduced and approximated with normal distribution functions. The variances of the approximating curves are: $B_1 - .64$; $B_2 - 1.22$; $B_3 - 1.4$; $B_4 - 1.4$; $B_5 - 1.4$.

IV. Mean and variance of maximum flow rate. The probability distribution of the maximum flow between a pair of vertices is difficult to compute, even if we make the simplifying (but usually accurate) assumption that the branch flows are normally distributed. However, on the basis of the results reported in the last section, it seems that the maximum flow is itself normally distributed. This observation is supported by the conclusions reached in a paper by C. E. Clark [10].

If we are allowed to assume that the maximum flow rate probability distribution may be adequately approximated by a normal distribution, the problem of finding $\text{Prob} \{F_{ij} \geq R\}$ is still far from solved. We must now find EF_{ij} and $\text{Var} F_{ij}$. From the integral form of $\text{Prob} \{F_{ij} \geq R\}$ given in Eqs. (12) and (14), we can see that this is a formidable problem. However, if we use the approach given by Clark [10] (which seems to be the best analytic approach available), the problem becomes manageable.

Let $|A_1|, \dots, |A_n|$ be random variables with arbitrary means, variances, and correlations. Clark suggests the following approach for computing the moments of $\min(|A_1|, \dots, |A_n|)$ (actually, Clark discusses $\max(|A_1|, \dots, |A_n|)$ but the treatment of these two problems is identical). The random variable $\min(|A_1|, \dots, |A_{i+1}|)$ can be written as

$$\min(|A_1|, \dots, |A_{i+1}|) = \min[\min(|A_1|, \dots, |A_i|), |A_{i+1}|]. \tag{28}$$

Suppose we know the density of $\min(|A_1|, \dots, |A_i|)$ and the correlation between $\min(|A_1|, \dots, |A_i|)$ and $|A_{i+1}|$. Then, if we know how to find the density of the minimum of *two* random variables, we can find the density of $\min(|A_1|, \dots, |A_{i+1}|)$. We can give exact expressions for the moments of $\min(|A_1|, |A_2|)$ and if $r[\cdot, \cdot]$ denotes the coefficient of linear correlation, we can also give an exact expression for

$$r[\min(|A_1|, |A_2|), |A_i|] \quad (i \geq 3).$$

Let ν_i be the *i*th absolute moment of $\min(|A_1|, |A_2|)$, let δ_i be the mean of $|A_i|$ and β_i^2 be the variance of $|A_i|$. Then

$$\nu_1 = \delta_1\Phi(\alpha) + \delta_2\Phi(-\alpha) - \alpha\phi(\alpha), \tag{29a}$$

$$\nu_2 = (\delta_1^2 + \beta_1^2)\Phi(\alpha) + (\delta_2^2 + \beta_2^2)\Phi(-\alpha) - (\delta_1 + \delta_2)\alpha\phi(\alpha), \tag{29b}$$

and

$$r[\min(|A_1|, |A_2|), |A_i|] = \{\beta_1 r(|A_1|, |A_i|)\Phi(\alpha) + \beta_2 r(|A_2|, |A_i|)\Phi(-\alpha)\} / (\nu_2 - \nu_1^2)^{1/2}, \quad i \geq 3, \tag{29c}$$

where

$$a^2 = \beta_1^2 + \beta_2^2 - 2\beta_1\beta_2r(|A_1|, |A_2|),$$

and

$$\alpha = (\delta_2 - \delta_1)/av.$$

Here, we have assumed that the special case $\beta_1 - \beta_2 = r(|A_1|, |A_2|) - 1 = 0$ does not happen. If so, $|A_1|$ and $|A_2|$ differ by a constant.

Now, if we assume that $\min(|A_1|, |A_2|)$ is itself normally distributed, we can use the above formulas to find the moments of $\min[|A_1|, |A_2|, |A_3|] = \min[\min(|A_1|, |A_2|), |A_3|]$. Then, if we assume that this variable is normally distributed we can find the

moments of $\min (|A_1|, |A_2|, |A_3|, |A_4|)$, and so on. Eventually, we will arrive at the moments of $\min (|A_1|, \dots, |A_q|)$. Naturally, each such calculation will be inaccurate since $\min (|A_1|, \dots, |A_q|)$ is not normally distributed. However, the errors introduced by this procedure appear to be insignificant. Clark discusses the errors of the approximation. A convincing argument concerning the adequacy of the approximation is given by the following table (adapted from Clark's paper) of the mean of the minimum of q independent standard normal variables.

q	Exact $E \min (A_1 , \dots, A_q)$	Approximation
2	-0.5642	-0.5642
3	-0.8463	-0.8476
4	-1.0294	-1.0310
5	-1.1630	-1.1643
6	-1.2672	-1.2679
7	-1.3522	-1.3522
8	-1.4236	-1.4230
9	-1.4850	-1.4837
10	-1.5388	-1.5367

Furthermore, the approximations to $1 - S_n(z)$ shown in Figs. 9 and 10 were drawn with means and variances found according to the above method. Therefore, we can find an excellent approximation to EF_{ij} and $\text{Var } F_{ij}$. This means that we can compute $\text{Prob } \{F_{ij} \leq R\}$ as

$$\text{Prob } \{F_{ij} \geq R\} = 1 - \Phi\left(\frac{R - EF_{ij}}{\text{Var } F_{ij}}\right). \tag{30}$$

V. Parametric statistical analysis. We can further pursue the observation that F_{ij} is approximately normally distributed. In [2], we investigated the case where the branch flows had unknown probability distributions but were observable. Again, let $\mathfrak{F}(k) = (f_1(k), \dots, f_b(k))'$ be a measurement of the flows (F_1, \dots, F_b) at time k ($k = 1, \dots, n$). As before, assume that $\mathfrak{F}(1), \dots, \mathfrak{F}(n)$ are identically and independently distributed. We want to test the hypothesis

$$H_1 : p = \text{Prob } \{F_{ij} \geq R\} \geq p_0 \tag{31a}$$

against the alternative

$$H_2 : p < p_0 \tag{31b}$$

at level α , where α is the probability of Type I error.

If we assume that F_{ij} is normally distributed with unknown mean ν_1 and unknown variance σ^2 , we have a parametric testing problem. Let

$$m_k \triangleq F_{ij}(k) = \min_{1 \leq i \leq a} \sum_{j=1}^b b_{ij}(c_j - f_j(k)), \quad k = 1, \dots, m \tag{32}$$

and let M_k be the random variable corresponding to the k th observation of maximum flow. M_1, M_2, \dots, M_n are identically and independently distributed $N(\nu_1, \sigma^2)$ random

variables. Furthermore, let T be the statistic defined by Eq. (33) and let \mathcal{G} be the group of transformations *multiplication by a positive constant*.

$$T = \frac{\left(\frac{n-1}{n}\right)^{1/2} \sum_{i=1}^n (m_i - R)}{\left\{ \sum_{i=1}^n \left[(m_i - R) - \frac{1}{n} \sum_{i=1}^n (m_i - R) \right]^2 \right\}^{1/2}}. \tag{33}$$

Then, T is a maximal invariant on the parameter space of the variables (ξ, σ) where $\xi = \nu_1 - R$ [11]. (If t is a particular value of T and g is a transformation on the parameter space of (ξ, σ) , then T is *invariant* under g if for every value of (ξ, σ) , $T_{(\xi, \sigma)}(t) = T_{g(\xi, \sigma)}(t)$. If \mathcal{G} is a group of transformations such that T is invariant for any $g \in \mathcal{G}$, T is a *maximal invariant* if $T_{(\xi_1, \sigma_1)}(t) = T_{(\xi_2, \sigma_2)}(t)$ for all t implies $(\xi_1, \sigma_1) = g(\xi_2, \sigma_2)$ for some $g \in \mathcal{G}$.) Suppose we consider the class of all tests which depend on M_1, \dots, M_n only through the maximal invariant T . Then, it can be shown [11] that the Uniformly Most Powerful Invariant Test (i.e. the U.M.P. test among this class) for testing H_1 against H_2 at level α , is

Reject H_1 if

$$\frac{n^{1/2}(\bar{m} - R)}{\left(\sum_{k=1}^n (m_k - \bar{m})^2 / (n - 1) \right)^{1/2}} \leq K, \tag{34a}$$

and

Accept H_1 if

$$\frac{n^{1/2}(\bar{m} - R)}{\left(\sum_{k=1}^n (m_k - \bar{m})^2 / (n - 1) \right)^{1/2}} > K, \tag{34b}$$

where $\bar{m} = (1/n) \sum_{k=1}^n m_k$, and K is a constant determined by

$$\int_{-\infty}^K \int_0^\infty w^{(n-2)/2} \exp\left(-\frac{w}{2}\right) \exp\left\{-\frac{1}{2}\left(t\left(\frac{w}{n-1}\right)^{1/2} - n^{1/2}\theta_0\right)^2\right\} dw dt = \alpha 2^{n/2} \Gamma\left(\frac{n-1}{2}\right) (\pi(n-1))^{1/2}.$$

Thus, we have a statistical procedure for testing, on the basis of observations of branch flow, whether or not $\text{Prob} \{F_{ij} \geq R\} \geq p_0$. The procedure is optimum in the sense that the probability of rejecting H_1 ($p \geq p_0$) when it is true is no greater than α and the probability of accepting H_1 when it is false is minimum among all tests which depend on the sufficient statistic T . If we want stronger control over the probability of accepting H_1 when it is false, we should test $H'_1 : p \leq p_0$ against $H'_2 : p > p_0$. This is equivalent to testing $H'_1 : \text{Prob} \{F_{ij} < R\} \geq 1 - p_0 \triangleq p'_0$ against $H'_2 : \text{Prob} \{F_{ij} < R\} < p'_0$, which is the test discussed in detail by Lehmann. We further note that to perform our statistical test, we *do not* have to know the cut-set matrix B of G . Given the observed flow vector $\mathfrak{F}(k)$, we can form the graph G_k which has the same structure as G but has the branch capacity vector $\mathbf{c}(k) = \mathbf{c} - \mathfrak{F}(k)$. Then, m_k is equal to the maximum flow between v_i and v_j in the *deterministic* graph G_k . The maximum flow m_k can then be found without reference to the cut-sets of G_k , through application

of Ford and Fulkerson's vertex labeling algorithm [7]. This algorithm is extremely efficient, and its use in conjunction with the statistical test, makes the testing procedure quite practical.

VI. Conclusion. In this paper, we have given several procedures for computing $\text{Prob} \{F_{i,j} \geq R\}$. We have assumed normal branch flows, but most of the procedures we have given will yield reliable approximations to $\text{Prob} \{F_{i,j} \geq R\}$ even when branch flows are not normal variables. This is because the value of a cut-set is actually the sum of random variables and so will usually be approximately normal for a large graph. Clark's procedure seems to work quite well, even when the random variables being tested are not normally distributed. For example, in [10] he shows that the mean of the maximum of q uniformly distributed variables is close to the mean of the maximum q normally distributed variables. Finally, we gave a statistical procedure for testing the level of $\text{Prob} \{F_{i,j} \geq R\}$. This procedure appears to be extremely efficient. We can generalize the test to a sequential test but this generalization is complicated and is not discussed here. The application of our analysis techniques to the optimum synthesis of statistical nets will be the subject of another paper.

BIBLIOGRAPHY

- [1] H. Frank and S. L. Hakimi, *Probabilistic flows through a communication network*, IEEE Trans. on Circuit Theory, CT-12, No. 3, 413-414 (1965)
- [2] H. Frank and S. L. Hakimi, *On the optimum synthesis of statistical communications nets*, J. Franklin Institute 284, 407-416 (1967)
- [3] H. Frank and S. L. Hakimi, *Parametric synthesis of statistical communication nets*, Quart. Appl. Math. (to appear)
- [4] W. Feller, *An introduction to probability theory and its applications*, Vol. 1, Wiley, New York, 1957, p. 177
- [5] M. Fisz, *Probability theory and mathematical statistics*, Wiley, New York, 1963
- [6] H. Cramér, *Mathematical methods of statistics*, Princeton Univ. Press, Princeton, N. J., 1946
- [7] L. R. Ford, Jr., and D. R. Fulkerson, *Flows in networks*, Princeton Univ. Press, Princeton, N. J., 1962
- [8] F. R. Gantmacher, *The theory of matrices*, Vol. I, Chelsea, New York, 1960
- [9] S. Gupta, *Probability integrals of multivariate normal and multivariate t* , Ann. Math. Stat. 34, 792-828 (1963)
- [10] C. E. Clark, *The greatest of a finite set of random variables*, Operations Research, 145-162 (1961)
- [11] E. L. Lehmann, *Testing statistical hypotheses*, Wiley, New York, 1959