DISPERSITY ROUTING

N. F. Maxemchuk
RCA Laboratories
Princeton, N.J.

In this work a novel routing mechanism for store-and-forward date communications networks will be presented. Unlike conventional directory routing procedures, which route a message along a particular path between the source and destination, this routing mechanism sub-divides the message and disperses it through the maze of paths comprising the network. Therefore, the mechanism is referred to as dispersity routing. In a recent analysis[**?**], conditions have been found under which dispersity routing systems provide the following advantages over conventional directory procedures:

a significantly smaller average and variance of delay,

less sensitivity to both incremental and large increases in link utilization,

an ability to continue to operate, without adapting the routing rule, when complete link failures occur,

an ability to sustain a larger number of transmission errors before requiring the message to be retransmitted, and,

smaller nodal buffer requirements for the same probability of losing a message due to buffer overflow.

The reason for many of these advantages will be apparent when the routing mechanism is defined. To demonstrate the conditions under which the average and variance of delay are decreased, the reason for the decrease, and the amount of improvement which may be obtained, elementary network configurations will be analyzed. The comparisons are conducted between non-adaptive directory routing procedures and non-adaptive dispersity routing procedures. However, many of the adaptive routing procedures which are applicable in directory procedures are also applicable in dispersity routed system, and dispersity routing makes it possible to implement additional adaptive routing rules.

Dispersity routed systems are classified as redundant and non-redundant systems. In a non-redundant system, a message is divided into a number of equal length sub-messages, equal to the number of paths between the source and destination which are to be used. Each sub-message is directed along a different path and the message is reconstructed when the last sub-message arrives at the destination. In a redundant system the number of message sub-divisions is less than the number of paths which are to be used. Additional sub-messages are formed as a linear combination of the bits in the message sub-divisions, and each of the redundant and original sub-messages is transmitted along a different path. The link utilization is increased by the additional sub-messages. However, by the appropriate choice of linear combinations and by using techniques associated with erasure correcting codes, the message can be reconstructed without receiving all of the sub-messages. Thereby, the paths with the longest delays do not effect the system delay, and the effect of sub-messages which are lost due to transmission errors or buffer overflows is reduced.

To illustrate the concept of dispersity routing, consider a system with three paths between the source and destination. A conventional directory routing procedure would route the message along one of the three paths. A non-redundant dispersity routed system would divide the message into three equal length sub-messages and route each along a different path. A redundant system would operate in one of the two odes. The entire message can be transmitted on each path. This triples the link utilization but enables the destination to decode the message when the first segment is received. This technique has been referred to as selective flooding[**?**]. Alternatively, the message can be divided into two equal length sub-messages. A message with N bits can be dispersed on the three paths as:

Path 1	Path 2	Path 3
I_1	$I_{N/2+1}$	P_1
I_2	$I_{N/2+2}$	P_2
:	:	:
:	:	:
$I_{N/2}$	I_N	$P_{N/2}$

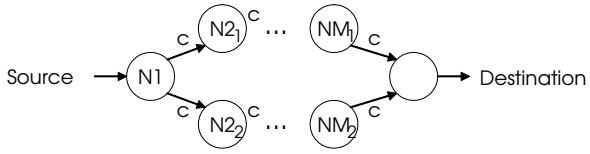
where I_1 to I_N are the N bits in the message and $P_1 = I_1 + I_{N/2+1}$. As a slightly more complex example of a redundant system, consider a system with seven paths between the source and destination. Divide the message into four equal length sub-messages, and transmit on the seven paths as:

Path 1	Path 2	Path 3	Path 4	Path 5	Path 6	Path 7
I_1	$I_{N/4+1}$	$I_{N/2+1}$	$I_{3N/4+1}$	$P_{5,1}$	$P_{6,1}$	$P_{7,1}$
I_1	$I_{N/4+2}$	$I_{N/2+2}$	$I_{3N/4+2}$	$P_{5,2}$	$P_{6,2}$	$P_{7,2}$
:	:	:	:	:	:	:
:	:	:	:	:	:	:
$I_{N/4}$	$I_{N/2}$	$I_{3N/4}$	I_N	$P_{5,N/4}$	$P_{6,N/4}$	$P_{7,N/4}$

The columns of this array represent the sub-message transmitted on the path whose number is at the top of the column. Considering each row of the array to be a codeword in a $(7, 4)$ Hamming code, the last three elements are determined by the first four. The Hamming code has a minimum distance between codewords of three, and therefore, any two erasures can be corrected. This implies that the transmitted message can be reconstructed while any two sub-messages are outstanding. In addition, the message may also be reconstructed while the last three sub-messages are outstanding.

In general, when there are N paths in the system and the message is divided into K sub-messages, the row of the array describes an (N, K) code, a code with N transmitted bits and K information bits. The maximum minimum distance between codewords in this class of block codes is $d - N - K + 1$, and the erasure correcting ability of a code with this minimum distance is $N - K$. A code with this minimum distance is appropriately referred to as a maximum-distance-separable code[**?**]. This code enables the message to be decoded after receiving any K sub-messages, and this is the smallest number of sub-messages needed to reconstruct an arbitrary message. The only binary maximum-distance-separable codes are the trivial codes, such as the single parity check code and the repetition code shown in the three path example. To obtain a code with this minimum distance characteristic in most instances requires using more complex, non-binary codes. For instance, the Reed-Solomon Codes are a well known class of non-binary maximum-distance-separable codes. A code in this class can be used in the seven path example to enable decoding after receiving any four sub-messages. The symbols of the codeword in this code are in Galois field with eight elements. Therefore, instead of encoding each row of the array separately, three rows would be encoded simultaneously and the three binary bits in the same column would define a single element in the Galois field GF(8). Whether or not the additional complexity of a non-binary code is warranted to obtain additional erasure correcting ability depends on the system to be implemented. In the remainder of this paper dispersity routed systems will be referred to as (N, J, K) systems where N is the number of paths, J is the number of message sub-divisions and K is the number of receptions required before decoding.

To understand the reason for the decreased average delay in a dispersity routed system, consider a store-and-forward network with two paths between each source and destination.



If the message interarrivals and service times are exponentially distributed, the average message length is $1/\mu$, and there are λ messages per second on each path, the delay at each node from Kleinrock's model[**?**] is:

$$E(D) = \frac{1}{\mu C} \frac{1}{1-\rho} \tag{1}$$

where the link utilization $\rho = \frac{\lambda}{\mu C}$. The average delay on a path in a directory routed system with M intermediate nodes is:

$$E(D) = \frac{M}{\mu C} \frac{1}{1-\rho} \tag{2}$$

If all messages are divided in half, and half the message is transmitted on each link, the number of messages per second is 2λ and the average message length is $\frac{1}{2\mu}$. Therefore, the link utilization is $\frac{2\lambda}{(2\mu)(C)} = \frac{\lambda}{\mu C}$, the same as that in the directory procedure, and the average delay on each path is:

$$E(D) = \frac{M\lambda}{2\mu C} \frac{2}{2-\rho} \tag{3}$$

which is half that in the directory procedure. The average message delay for the two path system using dispersity routing is not quite half that using conventional directory procedures because the message cannot be reconstructed at the receiver until the later of the two sub-messages is received. However, the reduction in the single path delay is the basis for expecting a decrease in the average message delay when a message is dispersed through the system. In addition, as the number of possible paths increases, the potential reduction in the average delay that can be obtained by dispersity routing increases. However, the time between the average sub-message arrival and the last sub-message arrival in a non-redundant system also increases and eliminates an increasingly larger part of the reduction that is obtained. Introducing redundant sub-message length, thereby increasing the delay on any one path, but eliminates the need to wait for the last sub-message. When a large number of paths are used, redundant systems decrease the average delay below that in non-redundant systems.

To determine the effects of waiting for the last sub-message in a non-redundant system and inserting extra sub-messages in a redundant system, an elementary system will be analyzed. This system consists of N error-free paths with one queue on each path, with infinite buffers and independent, identically distributed waiting times in each queue, and no additional bits transmitted in the dispersity routed system to identify the sub-messages. The effects of transmission error, finite buffers, more than one node per path, unequal delay distributions on the paths, and the additional bits that must be transmitted to identify sub-messages in a dispersity routed system, have been incorporated into a more complete analysis[**?**]. However, these effects obscure the basic result and will not be considered at this time.

To analyze the elementary system, the system delay is divided into two components, the system service time, and the system waiting time. The system service time is the time spent transmitting the message through the system, and the system waiting is the time spent waiting for the transmission facility. If the messages are exponentially distributed with average length $1/\mu_0$ and are transmitted on a channel with capacity C, the mean and variance of the service time are $\frac{1}{\mu_0 C}$ and $(\frac{1}{\mu_0 C})^2$. In an (N, J, K) dispersity routed system, the mean and variance of the service time at a single node are $\frac{1}{J\mu_0 C}$ and $(\frac{1}{J\mu_0 C})^2$. In the elementary system, this is the service time on each path and the system service time. The system waiting time in an (N, J, K) system is the K^{th} longest of the N signal path waiting times. In the elementary system, in which each path has an independent, identical waiting time distribution,

the distribution of the system waiting time, $F_W(t)$, is equal to the distribution of the K^{th} of N order statistics from a parent population equal to the distribution of the single path waiting time, $F(t)$. Therefore,

$$dF_W(t) = K\binom{N}{K}F^{K-1}(t)[1-F(t)]^{N-K}DF(t). \qquad (4)$$

Conducting the analysis with independent waiting time distributions on each of the paths is valid if different sets of messages used each of the queues. In the elementary system described, the same messages insert sub-messages in each queue. Therefore, the instantaneous waiting time in each of the queues would be the same, and the system delay would equal the single path delay. However, the elementary system is not a practical system, in that it is unlikely that N separate paths would directly connect a single source and destination. Instead, the result is meant to be indicative of a system with N paths between a source and destination, each path having a number of different intermediate nodes. According to Kleinrock's independence argument[**?**], successive nodes in a store-and-forward network can be analyzed independently. Since the correlation between waiting times in nodes on different paths should be less than that between

In a system with exponentially distributed message lengths and inter-arrival times, the waiting time distribution at a node is:

$$F(t) = \begin{cases} 1-\rho e^{-\mu C(1-\rho)t} & t \geq 0 \\ 0 & \text{elsewhere.} \end{cases}$$

Substituting F(t) into equation 4, the mean and variance of the system waiting time are

$$\mu_W = \frac{K}{\mu C(1-\rho)}\binom{N}{K}\sum_{j=0}^{K-1}(-1)^j\binom{K-1}{j}\frac{\rho^{N-K+1+j}}{N-K+1+j}, \qquad (5)$$

and,

$$\sigma_W^2 = \frac{2K}{[\mu C(1-\rho)]^2}\binom{N}{K}\sum_{j=0}^{K-1}(-1)^j\binom{K-1}{j}\frac{\rho^{N-K+1+j}}{[N-K+1+j]^2}. \qquad (6)$$

In an (N, J, K) dispersity routed network, the relationship between the link utilization ρ and the message utilization, ρ_0, is:

$$\rho = \frac{N}{J}\rho_0, \qquad (7)$$

and relationship between the average sub-message length $\frac{1}{\mu}$, and the original message length, $\frac{1}{\mu_0}$, is:

$$\frac{1}{\mu} = \frac{1}{J}\frac{1}{\mu_0}.$$

The average system delay equals the sum of the average waiting time and average service time. And, since the service time and waiting time are independent, the variance of the system delay equals the sum of the variances of these two quantities.

The average and variance of the system delays in a number of systems have been plotted versus message utilization in Figures 1-4. To limit the number of parameters, the average delay is normalized as $\frac{E(D)}{1/\mu_0 C}$ and the average variance as $\frac{VAR(D)}{(1/\mu_0 C)^2}$. The curves labeled $(1,1,1)$ represent the conventional directory procedures. The curves labeled (N, J, K) represent dispersity routed systems with independent waiting times on each of the N paths. And, the curves labeled $2C$ and $5C$ represent the single path delay in the $(2,2,2)$ and $(5,5,5)$ systems and also the delay in these systems if the waiting time on each path is identical instead of independent. The Figures 1 and 2, the distance between the $(1,1,1)$ curve and $(2,2,2)$ curve equals the decrease in the mean and variance of delay obtained by routing half of each message along each of two paths rather than half of the messages along each path. This distance between the curves representing the $(1,1,1)$ and $(5,5,5)$ systems equals the decrease obtained by using a five path non-redundant dispersity routed system instead of a conventional directory procedure.

The cross-hatched regions in Figures 1 and 2 represent the additional delay incurred by waiting for the last of N sub-messages instead of the sub-messages on just one of the paths. As expected, this penalty increases when the number of paths is increased from two to five. When redundant sub-messages are incorporated in the system, the link utilization, and hence the single path delay, increases. However, the message can be decoded before the last sub-message is received. When there are a large number of paths in the system, redundant sub-messages decrease the mean and variance of the delay below that in the non-redundant systems. This is shown in Figures 3 an d4, where the means and variance of delay in redundant and non-redundant systems are plotted for a configuration with five paths. The $(5,4,4)$ system, a system in which the message is divided into four parts and a single parity sub-message is added, is found to significantly decrease the mean and variance of delay over a range of link utilizations. The analysis of these simple systems is indicative of the results obtained in more complex systems and demonstrates the effects of dispersity routing on the man and variance of the network delay.

REFERENCES

[1] N.F. Maxemchuk, *Dispersity Routing in Store-and-Forward Networks*, submitted as Ph.D. Dissertation, University of Pennsylvania.

[2] Paul Baran, "On Distributed Communications Networks," *IEEE Trans. on COMM. Sys.*, pg. 1-9, March 1964.

[3] W.W. Peterson and E.J. Weldon, *Error Correcting Codes*, The MIT Press, 1972.

[4] L. Kleinrock, *Communication Nets*, McGraw-Hill Book Company, 1964.

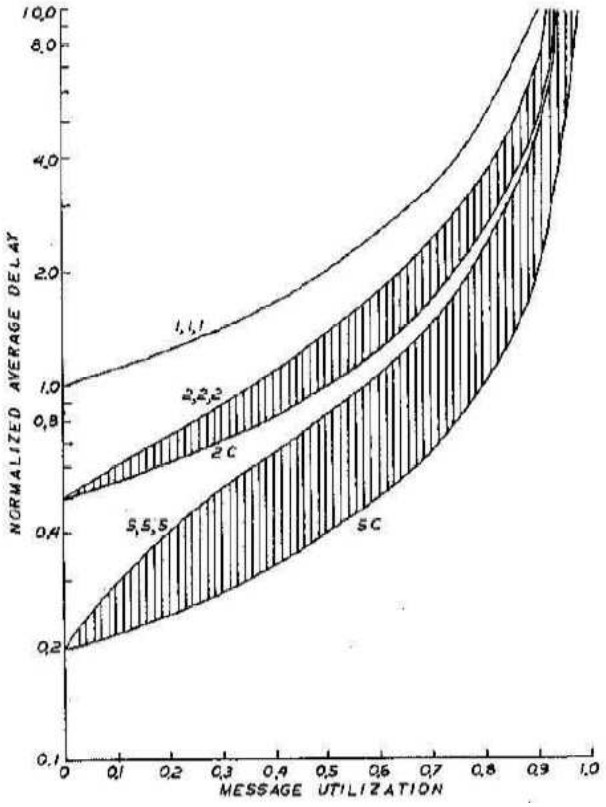


Fig. 1. Normalized average delay in (N, N, N) system with one queue per path.

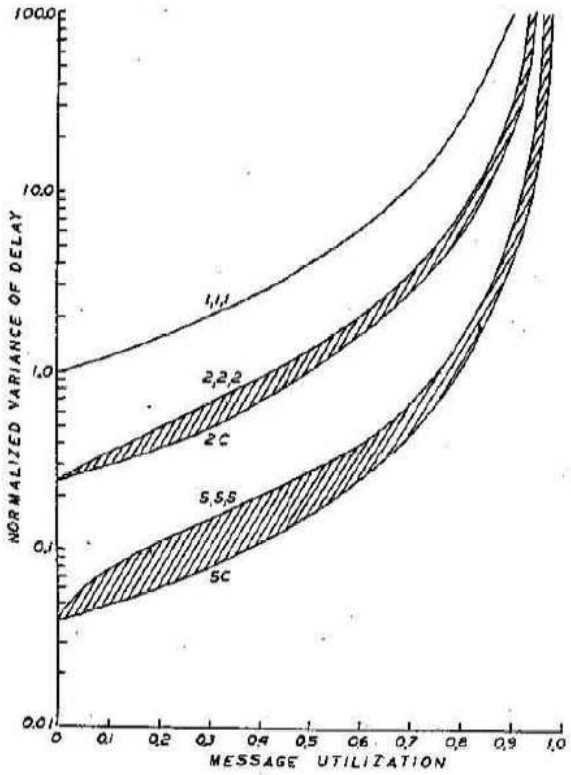


Fig. 2. Normalized variance of delay in (N, N, N) systems with one queue per path.

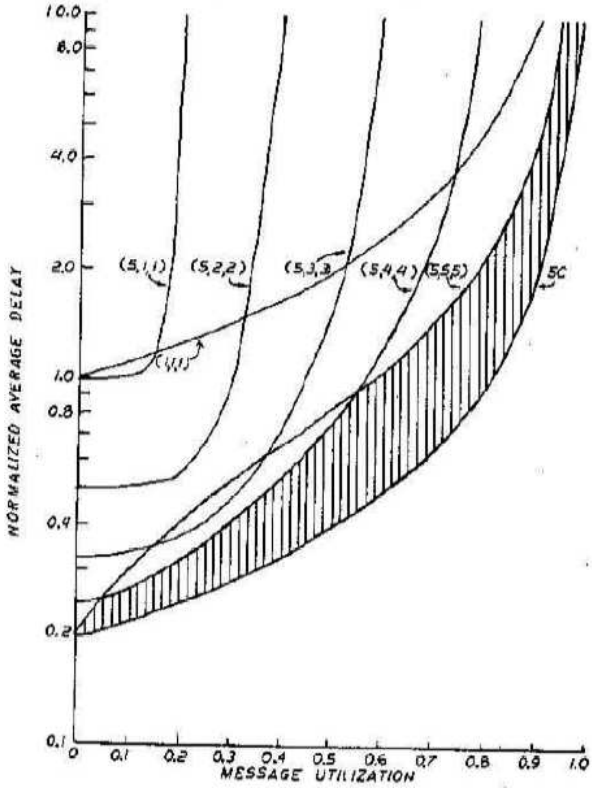


Fig. 3. Normalized average delay in systems with five paths and one queue per path.

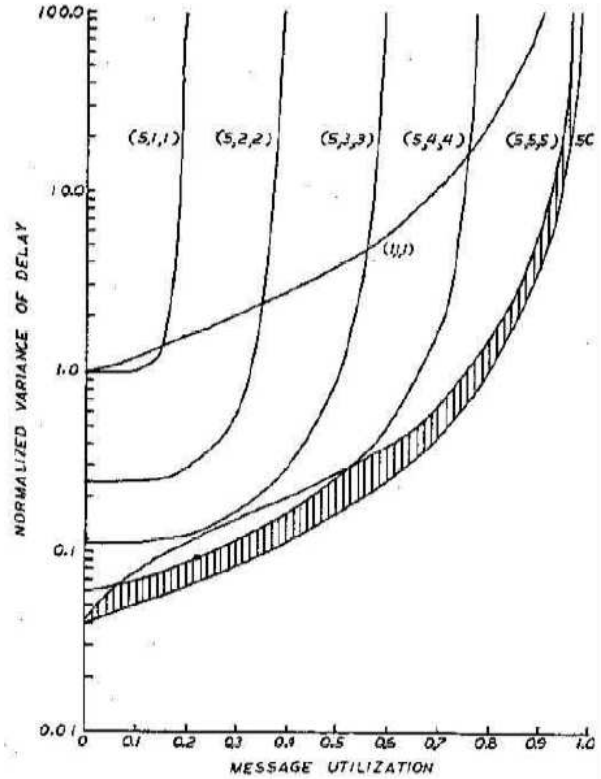


Fig. 4. Normalized variance of delay in systems with five paths and one queue per path.