University of Waterloo

DISTRIBUTION OF END-TO-END DELAY IN MESSAGE-SWITCHED NETWORKS*

by

J. W. Wong

Department of Computer Science
and
Computer Communications Networks Group
University of Waterloo
Waterloo, Ontario, Canada N2L 3G1

CCNG E-Report E-65
September 1977

Computer Communications Networks Group

P
91
C655
W58
1977

JUL 14 1977

ISSN 0384-5702
Queen
P
91
C655
W58
1977

DISTRIBUTION OF END-TO-END DELAY IN MESSAGE-SWITCHED NETWORKS*

by

J. W. Wong

Department of Computer Science
and
Computer Communications Networks Group
University of Waterloo
Waterloo, Ontario, Canada N2L 3G1

CCNG E-Report E-65
September 1977

Industry Canada
Library Queen

JUL 23 1998

Industrie Canada
Bibliothèque Queen

*This work was supported by the Department of Communications and the National Research Council of Canada

© COPYRIGHT: The Computer Communications Networks Group
The University of Waterloo
Waterloo, Ontario, Canada N2L 3G1

COMMUNICATIONS CANADA
OCT 11 1979
LIBRARY - BIBLIOTHÈQUE

TK
5105.5
W66

ABSTRACT

An open queueing network model is used to derive the distribution of end-to-end delay in a message-switched network. It is shown that under fixed routing, the end-to-end delay of messages belonging to a particular source-destination node pair is given by a sum of independent and exponentially distributed random variables. The generalization of this basic result to random routing and to messages belonging to a group of source-destination pairs is also considered. Numerical examples based on a hypothetical network are presented.

Keywords: Message-Switched Networks, Queueing Model, Delay Distribution

1. Introduction

A message-switched network <1> is a collection of switching nodes connected together by a set of communication channels. It provides a message service to the users at the various nodes. Messages in this network are routed from one node to another in a store-and-forward manner until they reach their destinations. A key performance measure of this network is the end-to-end delay which is the elapsed time from the arrival of a message at its source to the successful delivery of this message at its destination. In 1964, Kleinrock <1> developed an open queueing network model for message-switched networks and derived an expression for the mean end-to-end delay. This expression has been used extensively for performance analysis <3> and network design <3>.

Kleinrock's result is a mean delay taken over all the messages delivered by the network; no distinction is made on the basis of source or destination. In this paper, we treat messages with the same source-destination pair as belonging to a particular message class, and derive the distribution of end-to-end delay for each class. Both fixed and random routing are considered. Our result is therefore a detailed characterization of end-to-end delay in a message-switched network. It allows us to determine statistics such as the mean, variance, and 90-percentile of end-to-end delay for a particular source-destination pair.

Our derivation is based on Kleinrock's model <1> with emphasis given to classes of messages. A description of this model is given in section 2. This model is also a special case of the general queueing network model studied by Baskett, et.al. <4>. Baskett's result will therefore be used as the point of departure for our analysis. In section 3, our basic result on the distribution of end-to-end delay for the case of fixed routing is derived. This basic result is generalized to random routing in section 4 and to messages belonging to a group of source-destination pairs in section 5. Finally, sections 6 and 7 are devoted to numerical examples and application of results.

2. Model Description

We first assume, as in <1>, that the delay experienced by a message in a message-switched network is approximated by the queueing time and the data transfer time in the channels. The processing time at the switching nodes and the propagation delays are assumed to be negligible. Let M be the number of channels and C_i be the capacity of channel i, i = 1,2,...,M. In our open queueing network model, each of the M channels is represented by an independent server. We assume that all channels are error-free, and the queueing discipline at each channel is first-come, first-served (FCFS).

Messages are classified according to source-destination pairs. In particular, a message is said to belong to class (s,d) if its source node is s and its destination node is d. Let R be the total number of message classes. In a network with N switching nodes, $R = N(N-1)$. For convenience, we assume that message classes are numbered from 1 to R, and we use r instead of (s,d) to denote a message class. The arrival process of class r messages from outside the network is assumed to be Poisson with mean rate $\gamma(r)$. Message lengths for all classes are assumed to have the same exponential distribution, and we use $1/\mu$ to denote the mean message length. It follows from this last assumption that the service time of all messages at channel i is exponential with mean $1/\mu C_i$. For the mathematical analysis to be tractable, Kleinrock's independence assumption <1> is used. This assumption states that each time a message enters a switching node, a new length is chosen from the exponential message length distribution.

The message routing algorithm can be fixed or random. In fixed routing, a unique path is defined for each message class, and we use $a(r)$ to denote the path for class r. $a(r)$ is essentially an ordered set of channels over which class r messages are routed. In random routing, we allow the possibility of alternate paths, and the routing algorithm selects one of these paths according to a probability distribution. We will use k_r to denote the number of alternate paths for class r, $a_j(r)$ to represent the set of channels in the j-th path, and $q_j(r)$ the probability that the j-th path is selected, $j = 1,2,\ldots,k_r$.

3. Distribution of End-to-End Delay

We first consider the case of fixed routing. Let λ_{ir} ($i = 1,2,\ldots,M$; $r = 1,2,\ldots,R$) be the mean arrival rate of class r messages to channel i. With fixed routing, λ_{ir} is given by:

$$\lambda_{ir} = \begin{cases} \gamma(r) & \text{if channel } i \in a(r) \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

Let ρ_{ir} be the utilization of channel i by class r messages,

$$\rho_{ir} = \lambda_{ir}/\mu C_i \tag{2}$$

The total utilization of channel i (denoted by ρ_i) can then be written as:

$$\rho_i = \sum_{r=1}^{R} \rho_{ir} \tag{3}$$

We require that $\rho_i \leq 1$ for $i = 1,2,\ldots,M$. This is equivalent to the requirement that no channel is saturated, the condition for a stable network.

Let $t_r(x)$ be the probability density function (pdf) of the end-to-end delay of class r messages, and $T_r^*(s)$ be its Laplace Transform, i.e.,

$$T_r^*(s) = \int_0^\infty e^{-sx} t_r(x)\,dx \tag{4}$$

The main result of this paper can be stated as follows:

<u>Theorem</u>: For our model of a message-switched network with fixed routing,

$$T_r^*(s) = \prod_{i \in a(r)} \frac{\mu C_i(1-\rho_i)}{s + \mu C_i(1-\rho_i)} \tag{5}$$

A proof of this theorem is given in the Appendix.

Let $|a(r)|$ be the number of channels in $a(r)$. Our theorem indicates that the end-to-end delay of class r messages is given by the sum of $|a(r)|$ independent random variables. The i-th random variable in this sum is exponential with mean $(\mu C_i(1-\rho_i))^{-1}$; it can be interpreted as the delay at the i-th channel in the path of class r. It is of interest to note that the mean of this i-th random variable is a function of ρ_i and not ρ_{ir}, implying that all messages routed through a particular channel have the same delay distribution at this channel.

$T_r^*(s)$ can easily be inverted, by using partial fraction <2>, to give $t_r(x)$. The mean $\overline{T}_r$ and variance σ_r^2 of class r delay can also be obtained from $T_r^*(s)$. They are given by:

$$\overline{T}_r = \sum_{i \in a(r)} \frac{1}{\mu C_i(1-\rho_i)} \tag{6}$$

and

$$\sigma_r^2 = \sum_{i \in a(r)} \frac{1}{[\mu C_i(1-\rho_i)]^2} \tag{7}$$

4. Generalization to Random Routing

With random routing, a class r message can be routed through one of k_r alternate paths, and the j-th path is selected with probability $q_j(r)$. Our analysis in the last section is applicable if we treat each alternate path as a separate message class. We thus replace class r by k_r artificial classes. Let these classes be $r_1, r_2, \ldots, r_{k_r}$, then

$$\gamma(r_j) = \gamma(r)q_j(r) \tag{8}$$

$$\lambda_{ir_j} = \begin{cases} \gamma(r_j) & \text{if channel } i \in a_j(r) \\ 0 & \text{otherwise} \end{cases} \tag{9}$$

and

$$\rho_i = \sum_{r=1}^{R} \sum_{j=1}^{k_r} \lambda_{ir_j}/\mu C_i \tag{10}$$

Applying our theorem in the last section, we get:

$$T^*_{r_j}(s) = \prod_{i\in a_j(r)} \frac{\mu C_i(1-\rho_i)}{s + \mu C_i(1-\rho_i)} \tag{11}$$

$T^*_r(s)$ can then be obtained by removing the artificial class from our model, i.e.,

$$T^*_r(s) = \sum_{j=1}^{k_r} q_j(r) T^*_{r_j}(s) \tag{12}$$

Similar to the case of fixed routing, this Laplace Transform can also be inverted to give $t_r(x)$. As to the mean and variance of class r delay, we have:

$$\overline{T}_r = \sum_{j=1}^{k_r} q_j(r) \sum_{i \in a_j(r)} \frac{1}{\mu C_i(1-\rho_i)} \tag{13}$$

and

$$\sigma_r^2 = \sum_{j=1}^{k_r} q_j(r) \left[\sum_{i \in a_j(r)} \frac{1}{[\mu C_i(1-\rho_i)]^2} + \left[\sum_{i \in a_j(r)} \frac{1}{\mu C_i(1-\rho_i)} \right]^2 \right] - \overline{T}_r \tag{14}$$

5. Generalization to Message Groups

It is often useful to consider the end-to-end delay of messages belonging to a group of source-destination pairs. For example, we can study the delay characteristics of (a) messages sent among a subset of the nodes, (b) messages sent from a particular source node, or (c) messages sent to a particular destination. We thus define a group G to contain a number of message classes, and a message is said to belong to group G if its class membership is in G. It is easy to see that our result for random routing is directly applicable to message groups. We thus have the following result for $T_G^*(s)$, the Laplace Transform of the pdf of group G delay:

$$T_G^*(s) = \sum_{r \in G} \frac{\gamma(r)}{\gamma_G} T_r^*(s) \tag{15}$$

where $T_G^*(s)$ is given by Eq.(5) or (12), and $\gamma_G = \sum_{r \in G} \gamma(r)$.

In the special case that all classes of messages belong to a single group, we have the Laplace Transform of the pdf of the overall end-to-end delay:

$$T^*(s) = \sum_{r=1}^{R} \frac{\gamma(r)}{\gamma} T_r^*(s) \tag{16}$$

where $\gamma = \sum_{r=1}^{R} \gamma(r)$.

6. Numerical Examples

Our numerical examples are based on the hypothetical network shown in Figure 1. This network has 5 nodes and 10 channels. The external arrival rate of messages belonging to each source-destination pair is given by the traffic matrix in Figure 2. All channels are assumed to have the same capacity, and the mean message length is chosen such that the mean service time at each channel (i.e., $1/\mu C_i$) has a value of 0.1.

We first consider the case of fixed routing and assume that the routing algorithm is based on the shortest path. In our example network, there is a unique shortest path between each pair of nodes. Suppose we are interested in the end-to-end delay from node 1 to node 2. Denoting this source-destination pair by class 1, we apply Eq.(5) and get:

$$T_1^*(s) = \left[\frac{3}{s+3}\right] \left[\frac{2}{s+2}\right] \left[\frac{4}{s+4}\right]$$

This Laplace Transform can be inverted to give:

$$t_1(x) = -24\, e^{-3x} + 12\, e^{-2x} + 12\, e^{-4x}$$

A plot of $t_1(x)$ is shown in Figure 3. The mean, variance, and 90-percentile of class 1 delay are also shown.

We next consider the case of random routing and assume that 25% of class 1 messages are shifted to the path {1,3,8,9}. This implies that the remaining 75% are sent over the shortest path {1,5,9}. Applying Eq.(12), and inverting the resulting $T_1^*(s)$, we get:

$$t_1(x) = -25.2\, e^{-3x} + 30\, e^{-2.5x} + 99\, e^{-4.5x} - 19.8\, e^{-5.5x} - 84\, e^{-4x}$$

A plot of this pdf, together with its mean, variance, and 90-percentile, are shown in Figure 4. A comparison between Figures 3 and 4 indicates that the mean class 1 delay under random routing is smaller. This is due to the fact that a fraction of traffic has been directed from a more heavily utilized channel (channel 5) to a couple of less heavily utilized channels (channels 3 and 8).

As a third example, we consider the end-to-end delay of all messages originated from node 1 under fixed, shortest-path routing. Applying our results for message groups (Eq.(15)), we get the plot shown in Figure 5.

Finally, in Figure 6, we show the pdf of the end-to-end delay over all messages under fixed, shortest-path routing.

7. Application of Results

The results of this paper provide a detailed characterization of end-to-end delay in a message-switched network. They are useful for performance analysis and network design. They also find application in the analysis or simulation of a user-resource network <3> (or a subscriber network <6>) where terminals communicate with remote computers via a message-switched network. The message-switched network can be treated as a "black-box" with delay distribution given by the inverted Laplace Transform of Eq.(5) or (12) depending on whether fixed or random routing is used. This would reduce (a) the complexity of analysis and (b) the cost of simulation.

It should be noted that the derivation of our results is based on a rather general open queueing network model. These results have a wider scope of application than simply to message-switched networks.

8. Conclusion

We have used Kleinrock's model <1> to derive the distribution of end-to-end delay in a message-switched network. Both fixed and random routing have been considered. Our results find application in performance analysis of message-switched networks, and in the analysis or simulation of user-resource networks <3>.

References

1. L. Kleinrock, Communication Nets -- Stochastic Message Flow and Delays (McGraw-Hill, New York, 1964).
2. L. Kleinrock, Queueing Systems, Volume 1: Theory (Wiley-Interscience, New York, 1975).
3. L. Kleinrock, Queueing Systems, Volume 2: Computer Applications (Wiley-Interscience, New York, 1976).
4. F. Baskett, K. Chandy, R. Muntz, and F. Palacios, Open, Closed, and Mixed Network of Queues with Different Classes of Customers, Journal of the ACM 22 (1975) 248-260.
5. R. Cooper, Introduction to Queueing Theory (MacMillan, New York, 1972).
6. D. Davies and D. Barber, Communication Networks for Computers (John Wiley and Sons, New York, 1973).

Appendix: Proof of the Main Theorem

We first prove two lemmas.

Lemma 1:

Let $N_r(z)$ be the generating function of the total number of class r messages in the network at equilibrium,

$$N_r(z) = \prod_{i\in a(r)} \frac{1-\rho_i}{1-\rho_i+\rho_{ir}(1-z)} \qquad \text{(A1)}$$

Proof:

Let $(S_1, S_2, \ldots, S_M)$ be the state of our network model where $S_i = (n_{i1}, n_{i2}, \ldots, n_{iR})$ is the state of channel i and n_{ir} is the number of class r messages (in queue or in transmission) at channel i. Since our model is a special case of the general queueing network model analysed by Baskett, et.al. <4>, we apply Baskett's result and get the following expression for the equilibrium state probabilities:

$$P(S_1,S_2,\ldots,S_M) = P_1(S_1)P_2(S_2)\ldots P_M(S_M) \qquad \text{(A2)}$$

where

$$P_i(S_i) = (1-\rho_i)\left[\sum_{r=1}^{R} n_{ir}\right]! \prod_{r=1}^{R} \frac{1}{n_{ir}!}\rho_{ir}^{\,n_{ir}} \qquad \text{(A3)}$$

and ρ_{ir} and ρ_i are defined in Eqs.(2) and (3) respectively.

$P_i(S_i)$ is also the marginal probability that channel i is in state S_i. Let $N_{ir}(z)$ be the generating function of the number of class r messages at channel i. $N_{ir}(z)$ can be written as:

$$N_{ir}(z) = \sum_{\text{all states } S_i} P_i(S_i) z^{n_{ir}} \tag{A4}$$

Using Eq.(A3) in Eq.(A4), and after simplification, we get:

$$N_{ir}(z) = \frac{1-\rho_i}{1-\rho_i+\rho_{ir}(1-z)} \tag{A5}$$

From Eqs.(A2) and (A5), it is easy to see that $N_r(z)$ has the product form in Eq.(A1) because the equilibrium state probabilities are the same as if the state variable of the M channels are mutually independent.

Lemma 2:

Let $p_n(r)$ be the equilibrium probability that the number of class r messages in the network is n, and

$d_n(r)$ the probability that a class r departure left behind n class r messages.

Then

$$p_n(r) = d_n(r) \qquad n = 0,1,2,\ldots \tag{A6}$$

Proof:

A proof of this lemma for the single server queue is available in <2,5>. By treating the whole network as a single service facility, the same proof can be used for our network model.

We now prove our main theorem. Let $D_r(z)$ be the generating function of the number of class r messages left behind by a class r departure. Eq.(A6) implies that:

$$N_r(z) = D_r(z) \tag{A7}$$

Since we have assumed fixed routing and a FCFS discipline at each channel, the number of class r messages left behind by a class r departure must equal to the number of class r arrivals during the stay of the departing message in the network. Since we have also assumed a Poisson arrival process, $D_r(z)$ is given by <2>:

$$D_r(z) = T_r^*(\gamma(r) - \gamma(r)z) \tag{A8}$$

Substituting s for $\gamma(r)-\gamma(r)z$, Eq.(A8) is reduced to:

$$T_r^*(s) = D_r(1 - S/\gamma(r)) \tag{A9}$$

Finally, using Eq.(A1) and (A7) in Eq.(A9), we get:

$$T_r^*(s) = \prod_{i\epsilon a(r)} \frac{\mu C_i(1-\rho_i)}{s + \mu C_i(1-\rho_i)} \tag{A10}$$

QED

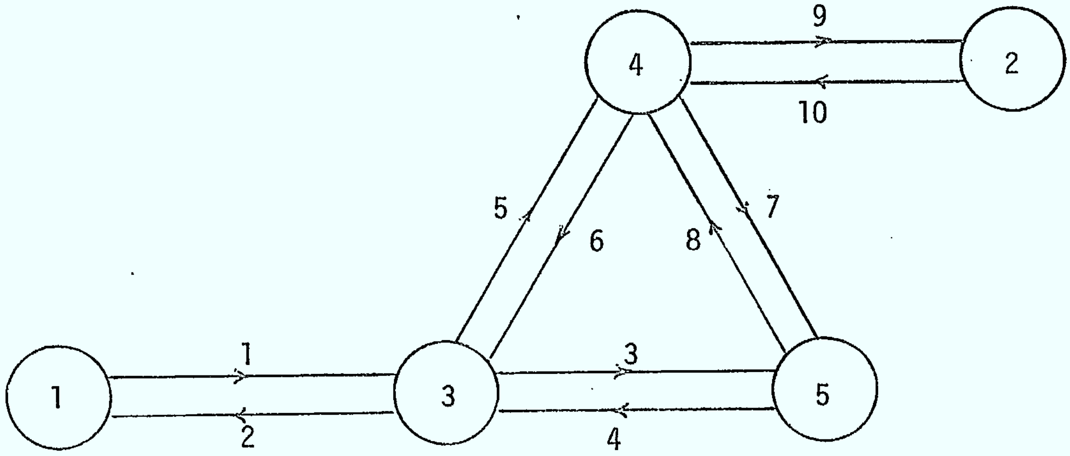

Figure 1. Hypothetical Network

DESTINATION

SOURCE	1	2	3	4	5
1	0	2	1	3	1
2	2	0	1	2	1
3	1	1	0	2	4
4	3	2	2	0	3
5	1	1	4	3	0

Figure 2. Traffic Matrix

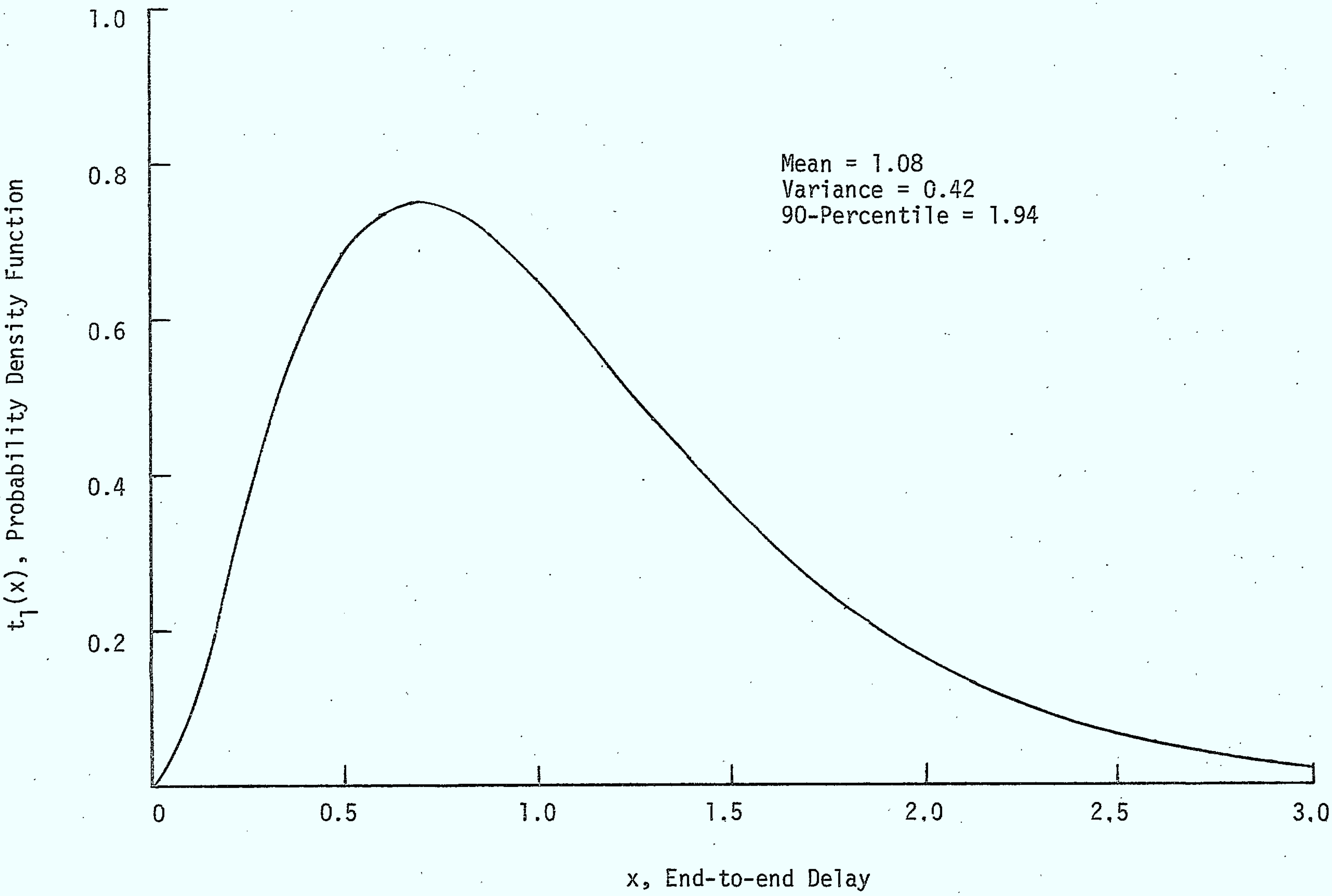

Figure 3. End-to-end Delay from Node 1 to Node 2 (Fixed Routing)

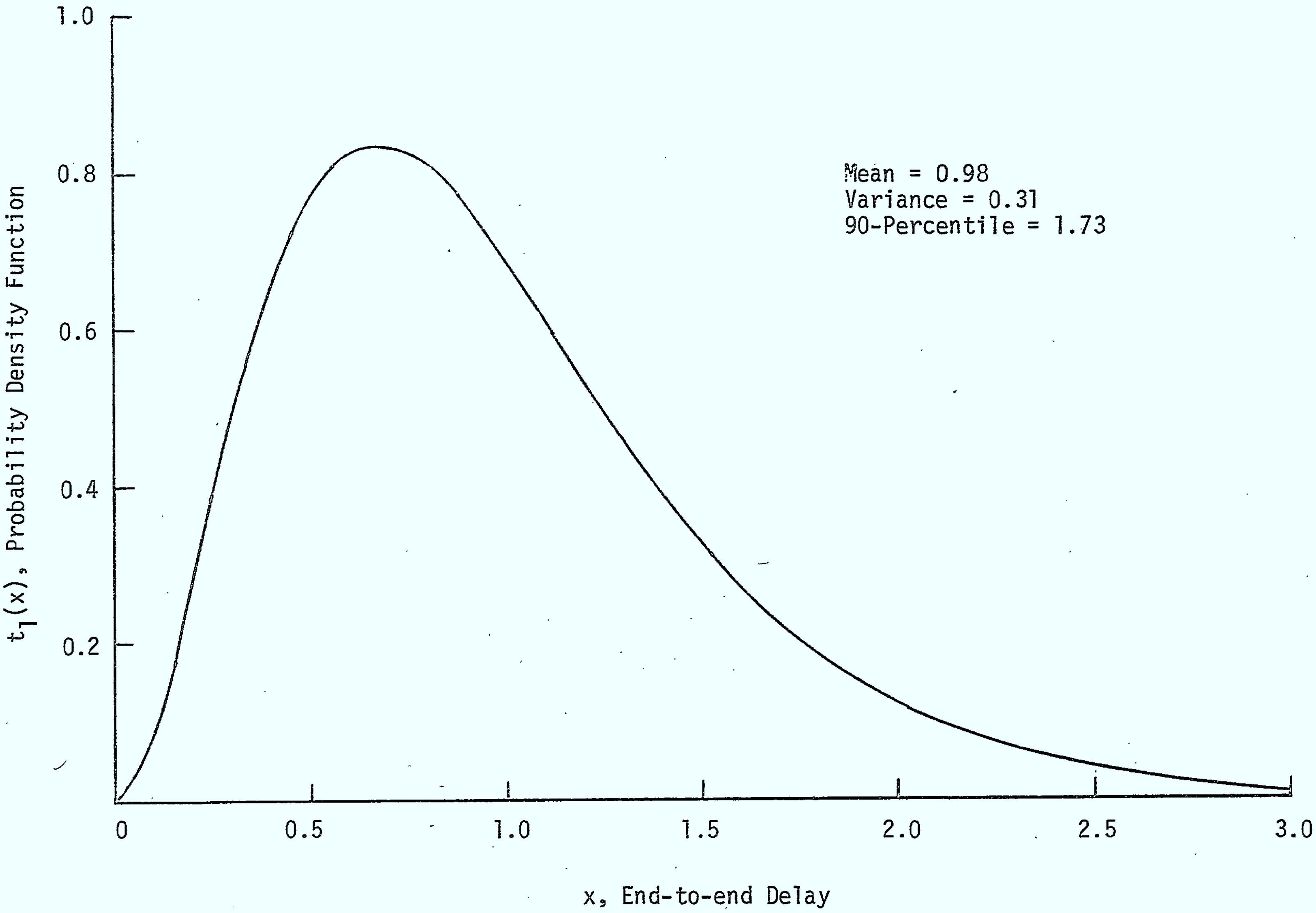

Figure 4: End-to-end Delay from Node 1 to Node 2 (Random Routing)

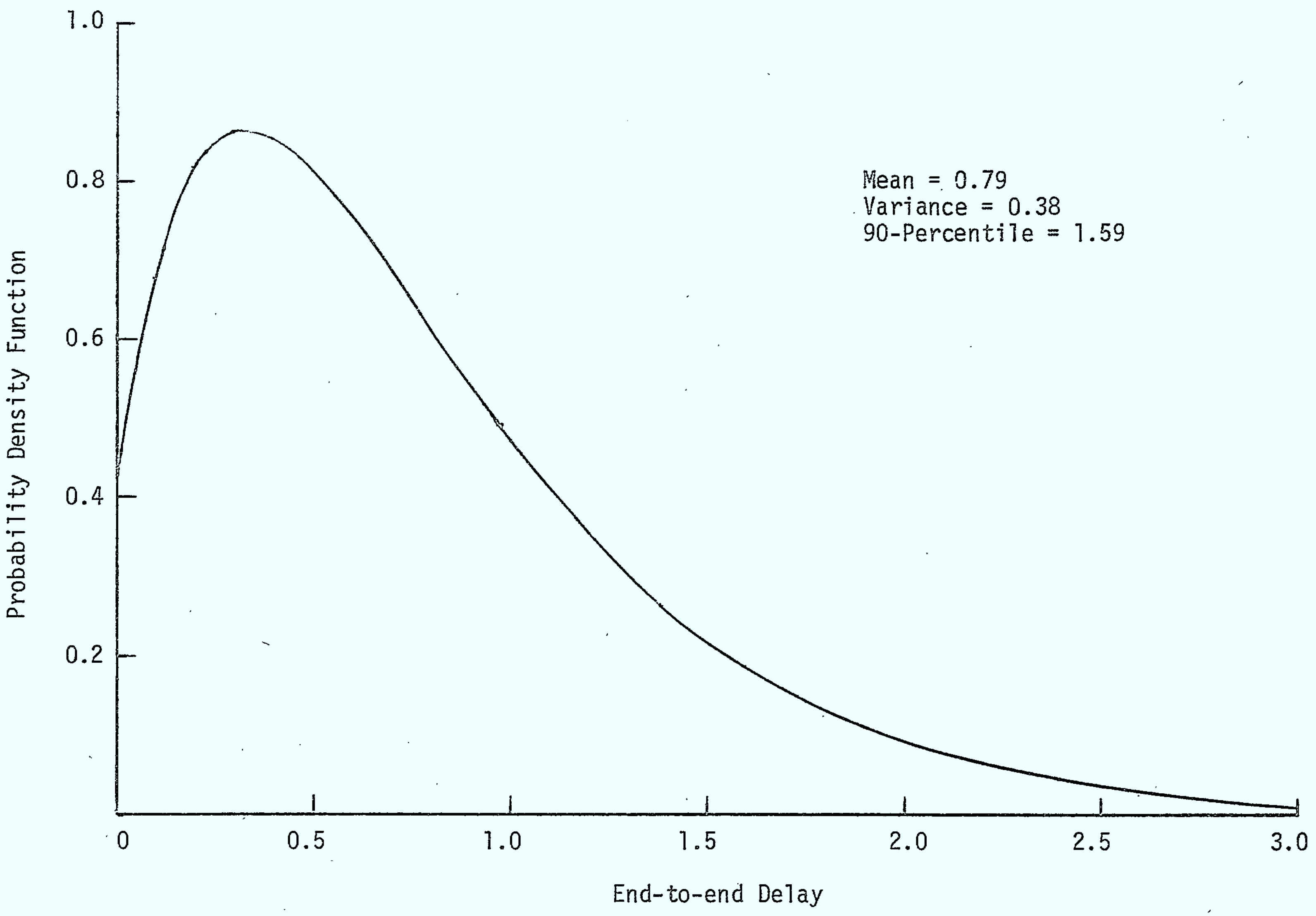

Figure 5. End-to-end Delay of Messages Originated from Node 1 (Fixed Routing)

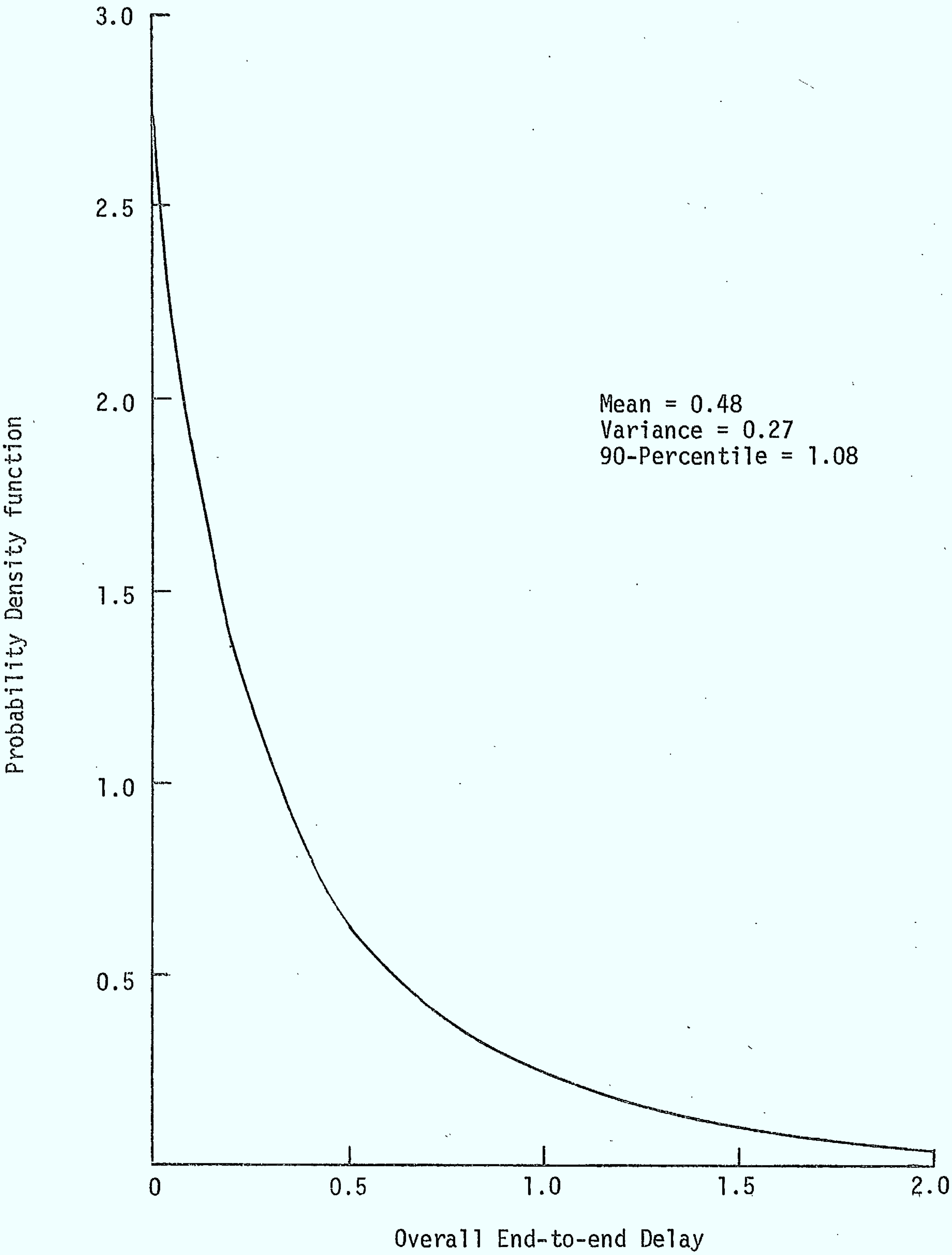

Figure 6. Overall End-to-end Delay

CACC / CCAC
87306

WONG, J.W.
--Distribution of end-to-end delay in message-switched networks.

P
91
C655
W58
1977

Date Due

FORM 109