Finally, we prove (4) by noting that by (11) we have
\[ \lim_{t \to \infty} P(W_n \geq t) = P(N = n) \]
and subsequently
\[ \lim_{t \to \infty} P(W_{n+1} \geq t) - \lim_{t \to \infty} P(W_n \geq t) = P(N = n). \] 

**REFERENCES**


**Communication Networks: Message Path Delays**

IZHAK RUBIN, MEMBER, IEEE

Abstract—A communication network is modeled by a weighted graph. The vertices of the graph represent stations with storage capabilities, while the edges of the graph represent communication channels (or other information processing media). Channel capacity weights are assigned to the edges of the network. The network is assumed to operate in a store-and-forward manner, so that when a channel is busy the messages directed into it are stored at the station, where it joins a queue that is governed by a first-come first-served service discipline.

Assuming that fixed-length messages arrive at random at the network, following the statistics of a Poisson point process, we solve for the steady-state distributions of the message overall delay time, for the average message waiting times at the individual stations, for the average memory size requirements at the stations, as well as for other statistical characteristics of the message flow along a communication path.

I. INTRODUCTION

A communication network is modeled by a weighted graph. The vertices of the graph represent stations with storage capabilities, while the edges of the graph represent communication channels (or other information processing media). Channel capacity weights are assigned to the edges of the network. The network is assumed to operate in a store-and-forward manner, so that when a channel is busy the messages directed into it are stored at the station, where it joins a queue that is governed by a first-come first-served service discipline.

Assuming that fixed-length messages arrive at random at the network, following the statistics of a Poisson point process, we solve for the steady-state distributions of the message overall delay time, for the average message waiting times at the individual stations, for the average memory size requirements at the stations, as well as for other statistical characteristics of the message flow along a communication path.

A large variety of information transmission (and processing) networks can be described by this model. In a satellite communication system, the stations (vertices) of the network (graph) represent satellites, ground stations, or airborne stations; these stations are interconnected by communication channels (edges). The weighting function associated with a satellite communication network assigns appropriate weights to the channels (like channel capacities, noise characteristics, etc.) as well as weights to the stations (information processing capabilities, power limitations, etc.). A similar model is utilized to describe a telephone, telegraph, space communication, or a computer communication network. In the latter network, the stations represent the users’ or the computers’ processing units, and communication channels interconnect the various users and computing facilities. In certain situations, one may want to associate the communication channels with the vertices of the graph and the stations with its edges (as in cases where time delays in the network are evaluated and the main time delay involved is associated with information processing in the station, like an encoding or decoding procedure, rather than along the channel). Similar models are of considerable importance in many other areas, such as transportation, biology, management, and operations research.

Communication networks are generally considered, as assumed here, to operate in a “store-and-forward” fashion. A message arriving at a station will be directed into the outgoing appropriate channel, following the system’s routing policy, and be transmitted over this channel if it is free for transmission. If the latter channel is busy, the message will be stored at the station and join a queue of messages.
that is assumed to be governed by a first-come first-served service discipline.

The two major considerations in analysis and synthesis of a communication network are that of congestion and time delays in the network and that of reliable transmission of information over the noisy channels. The latter consideration requires one to develop appropriate channel and source coding procedures to decrease the overall probability of error in message transmission through the network and to minimize a distortion measure, respectively. Such problems associated with simple communication networks have recently been considered by information theorists, and they are starting to draw considerable attention. See [5] for channel coding for broadcast channels (modeled by a network with one transmitting source and multiple receivers) and [6] for discussions concerning coding subject to a fidelity criterion for some simple networks under special noise characteristics.

When considering a communication network for which appropriate coding techniques have already been applied to combat noise interference (for example, by conventional coding methods, as is the case for the existing computer and satellite communication networks), so that one can view the resulting network to be noiseless, the only remaining major consideration is that of congestion and time delays. The latter is the subject of the present paper.

We assume that messages of constant length arrive at a station of a communication network at random times, governed by the statistics of a Poisson point process. Each channel (edge) in the network (graph) is assigned a capacity. Considering an arbitrary path in the network, which leads the messages from their origin to their destination through the network, we are interested in obtaining the time delays experienced by a message over this path. Using results and methods from queueing theory, we derive the (steady-state) distribution functions for the overall message waiting and delay times, and obtain, as well, the average storage requirements at the various stations. Using our results one can proceed, using time-delay criteria, with analysis and synthesis of communication networks when nonsimultaneous flows in paths are considered between the various stations of the communication network.

Approximate (limiting) average-time-delay expressions for communication networks (with simultaneous flows) have been obtained in [7] and applied to computer communication networks (see [8] and the references therein). In the latter analysis, message lengths are assumed to be exponentially distributed, so that Burke's theorem for an $M/M/1$ queueing system can be invoked to conclude that the message departure process at the first channel is (at steadystate) a Poisson point process, as is that at the input. Then, to avoid statistical dependencies, an “independence assumption” is made (see [7, p. 50]), which amounts to choosing the message length at random at each of the channels in the network. For large and (topologically) complex networks, the latter results have been observed (by simulation) in [7] to be good approximations (for the ARPA computer communication network, see also the measurements in [14]). In our situation, when path delays are sought (and constant message lengths are assumed, as is the case, for example, in the present computer communication networks that utilize fixed-length packets of messages), the latter approximation cannot be made, and we need to perform an exact analysis as presented by this paper. Related time-delay problems have been considered in association with queueing networks [9], queues in tandem [10], and a variety of computer processing systems. However, all the latter studies make the aforementioned independence assumption (i.e., taking the service duration at any queueing channel to be statistically independent of that in any other channel), which cannot be made in our problem, since a message carrying a fixed content of information is transferred through the network. For an approach to the analysis of queueing networks using the diffusion approximation see [11]. For studies of networks with deterministic channel delays see [12, ch. 9]. See [13] for a discrete-time queueing analysis of satellite networks.

The mathematical model of the communication network is presented in Section II. Time delays over a single channel path directly follow by using results from the queueing theory, and they are given in Section III. Delay characteristics for an $N$-channel path are studied in Section IV. In Section V we derive the distribution of the overall path delay and develop a capacity invariance property. The latter property asserts that the overall path delay is independent of the order of the various channel capacities and allows one subsequently to obtain the average time delay at each channel and the path delay distribution. An illustrative example is given in Section VI. In [15] we have derived the waiting-time distributions at the channels for a communication path, as well as other statistical properties of the channels’ queueing processes.

II. PRELIMINARIES

Communication Path

A communication path, as shown in Fig. 1(a), is considered. The branch $(v_i, v_{i+1})$ between vertices $v_i$ and $v_{i+1}$, $i = 1, 2, \cdots, N$, represents the $i$th channel, whose capacity is equal to $C_i$ bits/s. Thus each bit is delayed $1/C_i$ s by the $i$th channel. The vertices $v_p$, $i = 1, 2, \cdots, N + 1$, represent

![Fig. 1. (a) Communication path. (b) Path in communication network.](image)
stations or buffer systems equipped with memory storage units or queueing facilities. No restrictions are imposed on the size of the storage unit. (It will be clear from the results of the analysis how to control the memory sizes.) Messages arrive at the input station $v_1$ randomly in time according to a Poisson point process with intensity $\lambda$ messages/s. Each message (or packet of information) is assumed to be of a constant length $\beta$ bits/message. If the arriving message at $v_1$ finds channel 1 free, it is immediately transmitted to station $v_2$. The message transmission time delay over channel 1 is clearly

$$a_1 = \frac{\beta}{C_1}, \quad \text{s/message}.$$  

The output of channel 1 is stored at $v_2$ until the remaining message arrives. The latter message is stored at $v_2$ until channel 2 is free and then it is transmitted to $v_3$, and so forth, until the message leaves the system at $v_{N+1}$. If the arriving message at $v_1$ finds channel 1 busy (i.e., another message is being transmitted over this channel at its time of arrival), it will be stored at $v_1$ and wait until channel 1 is free and then it will be transmitted to $v_2$, etc. It is assumed that a first-come first-served discipline is employed to serve the waiting messages at $v_i$, $i = 1, 2, \ldots, N$.

In this paper, we derive steady-state distributions for the time delay of a message over the $N$-channel path, as well as the average waiting times and storage requirements at the individual stations.

We also note that $\beta$ bits out of the storage of $v_i$, $i = 2, 3, \ldots, N$, can be associated with the facilities of channel $i - 1$ for the purpose of collecting the message departing from channel $i - 1$. The latter message is then immediately transferred to the remaining storage facility of $v_i$. A storage of $\beta$ bits is required at $v_{N+1}$ since a departure is declared only when all the message is transmitted.

**Mathematical Model**

Each channel $i$ with its storage facilities can be considered a queueing system. For that purpose, consider a message transmitted over channel $i$ to be a customer that requires service from server $i$. The service time required by the customer (message) from the server (channel) $i$ is equal to his transmission time over the channel and is clearly given by

$$a_i = \frac{\beta}{C_i}, \quad \text{s.} \quad (1)$$

Upon leaving the $i$th queueing system, a customer enters the $(i + 1)$st queueing system where he is served on a first-come first-served basis and requires service from server $i$. The service time required by the customer (message) from the server (channel) $i$ is equal to his transmission time over the channel and is clearly given by

$$a_i = \frac{\beta}{C_i}, \quad \text{s.} \quad (1)$$

We will employ a queueing-theoretical approach.

The following notations will be utilized throughout the paper (and are common in queueing theory, see for example [1]). We let $X_t^{(i)}$ denote the number of messages stored at $v_i$ or being transmitted from $v_i$ (through channel $i$). Thus $\{X_t^{(i)}, t \geq 0\}$ is the queueing process associated with channel $i$. Assume $X_0^{(i)} = 0$, $i = 1, \ldots, n$. The (tandem) instants of arrival of messages at $v_i$ are denoted by $\{t_n^{(i)}, n = 1, 2, 3, \ldots\}$, and the departure stochastic point process from channel $i$ is denoted as $\{r_n^{(i)}, n = 1, 2, \ldots\}$, where $r_n^{(i)}$ and $r_n^{(i)}$ denote, for channel $i$, the instants of the $n$th arrival and departure, respectively. Clearly, for a communication path $t_n^{(i+1)} = r_n^{(i)}$. The waiting time at $v_i$ of the $n$th arriving message is denoted as $W_n^{(i)}$.

Observing the evolution of the queueing process $\{X_t^{(i)}, t \geq 0\}$, we find that the process passes successively through idle and busy periods. We denote, for channel $i$, the sequence of idle periods (see Fig. 2) by $\{I_n^{(i)}, n = 0, 1, 2, \ldots\}$, and the sequence of busy periods by $\{B_n^{(i)}, n = 1, 2, \ldots\}$. Thus for $\{X_t^{(i)}, t \geq 0\}$, $I_n^{(i)}$ is the duration of the $(n + 1)$st idle period (and during it $X_t^{(i)} = 0$), and $B_n^{(i)}$ is the duration of the $n$th busy period (and during it $X_t^{(i)} > 0$). The number of messages transmitted through channel $i$ during the $n$th busy period (whose length is $B_n^{(i)}$) is denoted as $N_n^{(i)}$. The delay of the $n$th message at channel $i$, $\gamma_n^{(i)}$, is defined as the total of its waiting time and transmission delay time at channel $i$. Thus we have

$$\gamma_n^{(i)} = W_n^{(i)} + a_i. \quad (2)$$

The overall time delay $\gamma_m^{(i)}$ of the $m$th arriving message through the $N$-channel communication path is thus given by

$$\gamma_m^{(i)} = \sum_{i=1}^{N} \gamma_m^{(i)}. \quad (3)$$
Thus we are interested in obtaining the steady-state average message waiting times at the individual stations $\lim_{n \to \infty} E\{W_n^{(i)}\},$ $i = 1, 2, \cdots, N.$ Using these expressions we would calculate the average overall message delay time $\gamma, \gamma = \lim_{n \to \infty} E\{\gamma_n\},$ as well as the average required storage capability at $v_i$ (which is related to $\lim_{n \to \infty} E\{X_t^{(i)}\}$). Moreover, we wish to derive the limiting distributions for the overall delay and the busy-period durations at the individual channels.

### III. Single-Channel Path

Consider the case $N = 1,$ so that the path includes only a single channel. In this case we are thus considering a queueing system into which customers (messages) arrive according to a Poisson process with intensity $\lambda$ and where each customer (message) requires a deterministic service of length $a_i.$ Such a system is denoted in queueing theory as an $M/D/1$ system. Results for this queueing system are readily available (see [3], [2, pp. 32-38], [1, ch. II.4]). In particular, assuming the traffic intensity $\rho_1 = \lambda a_1 < 1,$ so that $\{X_t^{(1)}, t \geq 0\}$ is ergodic (positive-recurrent), we have

$$\lim_{n \to \infty} P(W_n^{(1)} = 0) = 1 - \rho_1$$

so that the traffic intensity $\rho_1$ expresses the limiting probability of channel 1 being busy. The limiting waiting-time distribution is given by ((2, p. 335)

$$F_{W^{(1)}}(x) = \lim_{n \to \infty} P(W_n^{(1)} \leq x)$$

$$= (1 - \rho_1) \sum_{i=0}^{\infty} \exp [-\lambda(x - ia_1) - (i!)^{-1}]$$

$$\cdot [-\lambda(x - ia_1)]^i$$

for $x = na_1 + t,$ $n \geq 0,$ $0 \leq t < a_1.$ In particular, the limiting average waiting time is then given by

$$\lim_{n \to \infty} E(W_n^{(1)}) = \frac{1}{\frac{1}{2} - \rho_1} a_1$$

so that the delay time is given as

$$\gamma^{(1)} = \lim_{n \to \infty} E(\gamma_n^{(1)}) = \frac{1}{\frac{1}{2} - \rho_1} a_1$$

For the average queue size we have (see [1, p. 247], or just by Little's theorem [4]) the average queue size is equal to $\lambda^{(1)}$,

$$\lim_{t \to \infty} E\{X_t^{(1)}\} = \frac{1 - \rho_1/2}{1 - \rho_1} \rho_1.$$  

Since every message contains $\beta$ bits, the average memory size required at $v_i,$ to be denoted as $M^{(i)}$, is given by

$$M^{(1)} = \frac{1 - \rho_1/2}{1 - \rho_1} \beta,$$

so that we have

$$M^{(1)} = \frac{1}{1 - \rho_1} \beta$$

Of particular interest for the following analysis are the distributions of the idle and busy periods of $\{X_t^{(1)}, t \geq 0\}.$ Since the arrival process is Poisson with intensity $\lambda,$ the idle period is exponentially distributed with parameter $\lambda.$ Thus the distribution of any idle period $I^{(i)}$ is given by

$$P(I^{(i)} \leq x) = [1 - \exp (-\lambda x)] U(x)$$

where $U(x)$ is the unit step function $U(x) = 1,$ for $x > 0,$ and $U(x) = 0,$ for $x \leq 0.$ The distribution of the number of customers served during a busy period $N^{(i)}$ is obtained as (see [2, p. 36])

$$P(N^{(i)} = n) = (n!)^{-1} (\lambda a_1)^n \cdot \exp (-\lambda a_1),$$

and the first two moments are obtained as

$$E(N^{(i)}) = (1 - \rho_1)^{-1}$$

$$E[N^{(i)}] = (1 - \rho_1)^{-2}.$$  

We note that $\{I^{(i)}, i \geq 0\}$ and $\{N^{(i)}, n \geq 1\}$ are stochastically independent sequences of independently identically distributed random variables with distributions given by equations (10) and (11), respectively.

### IV. Delays in N-Channel Path

Consider an $N$-channel path, where $N \geq 2.$ Due to the series structure of the path, we clearly have

$$t_i^{(i)} = t_{i-1}^{(i-1)} + t_i^{(i-1)}, i \geq 2$$

so that the interarrival time at channel $i,$ $i \geq 2,$ is given by (see Fig. 2)

$$T^{(i)}_{n+1} = t_n^{(i)} - t_{n+1}^{(i)} = t^{(i-1)}_{n+1} - t^{(i-1)}_{n}$$

$$= (a_{i-1}, t^{(i-1)}_{n+1} - t^{(i-1)}_{n}, a_i)$$

where $I^{(i-1)}_{n+1}$ denotes the duration of the idle period at $v_{i-1}$ prior to $t^{(i-1)}_{n+1}.$ The waiting time at channel $i$ follows the relationship

$$W_{n+1}^{(i)} = [W_n^{(i)} + a_i - T^{(i)}_{n+1}]^+.$$  

Hence, by (14) and (15), we conclude that

$$W_n^{(i)} = 0, \quad a_i \leq a_{i-1}, \quad i \geq 2.$$  

However, these expressions can be simplified as follows. We define a sequence of ladder indices $\{k_j, j = 1, 2, \cdots, m\},$ so that

$$k_1 = 1$$

$$k_j = \min \{i: k_{j-1} < i \leq n, a_i > a_{k_{j-1}}\}, \quad for \ j > 2.$$  

Thus $k_2 = i,$ if $a_1 \geq a_2, 2 \leq j \leq i - 1,$ and $a_i < a_1,$ so that $k_2$ is the index of the first element of $\{a_n, 1 \leq i \leq N\},$ which is larger than $a_1.$ Similarly, $k_j$ is the index of the first element of $\{a_n, 1 \leq i \leq N\},$ following $a_{k_{j-1}},$ which is larger
than \(a_{k_j-1}\). For example, for the sequence \(\{a_i\} = \{6,8,2,5,10\}\), we have \(k_1 = 1, k_2 = 2, k_5 = 5, a_{k_5} = 6, a_{k_1} = 8, a_{k_2} = 10, a_3 = 3, a_4 = 9, m = 3\). The number of ladder indices \(m\) is clearly defined by \(\max_j k_j = k_m, 1 \leq m \leq N\). We observe that \(k_1 < k_2 < \cdots < k_m\) and \(a_{k_1} < a_{k_2} < \cdots < a_{k_m}\). We will now show that a message will have a zero waiting time at any channel \(j\) where \(j\) is not a ladder index of \(\{a_i\}\). Thus a message may have to wait only at channels whose index corresponds to a ladder index of \(\{a_i\}\). For that purpose, we first obtain the following property.

**Lemma 1.** For \(i \geq 2\), we have

\[
\frac{t_n^{(i)} - t_n^{(i-1)}}{t_n^{(i-2)}} = r_n^{(i)} - r_n^{(i-1)}
\]

where

\[
m_i = \max\{k_j: 1 \leq k_j \leq m, k_j < i\}.
\]

**Proof:** Consider an arbitrary channel \(i, i \geq 2\). The integer \(m_i\) represents the largest ladder index smaller than \(i\). Thus, if \(m_i = i - 1\), then (18) clearly holds, due to relation (13). If \(m_i = i - 2\), then \(i = 2\) is a ladder index and by (17) we observe that \(a_{i-2} \geq a_{i-1}\). Consequently, by (16), \(W^{(i-1)} = 0\) and \(t_n^{(i-1)} = t_n^{(i-2)} + a_{i-1}\), so that \(r_{n+1}^{(i-2)} - r_n^{(i-1)} = t_n^{(i-1)} = t_n^{(i-2)} - t_{n+1}^{(i-1)} = t_n^{(i-1)}\), and (18) holds. If \(m_i = i - 3\), then \(i = 3\) is a ladder index, and \(i = 2\) and \(i = 1\) are not ladder indices. Hence, by (17), \(a_{i-3} \geq a_{i-4}, a_{i-3} \geq a_{i-4}\). Subsequently, \(W^{(i-2)} = 0\) and \(\frac{t_{n+1}^{(i-2)} - a_{i-2}^{(i-2)}}{t_{n+1}^{(i-2)} - a_{i-2}^{(i-2)}} = \frac{t_{n+1}^{(i-2)} - t_{n+1}^{(i-3)}}{t_{n+1}^{(i-2)} - t_{n+1}^{(i-3)}}\), and (18) holds. Clearly, the proof extends to any arbitrary value \(m_i = i - k, 1 \leq k \leq i - 1\). Q.E.D.

In particular, the following property follows readily from Lemma 1 and its proof by observing that, if \(m_i = i - k, 1 \leq k \leq i - 1\), then \(W_n^{(i)} = 0\), for \(i - k + 1 \leq j \leq i - 1\), for each \(n \geq 1\).

**Theorem 1:** In an \(N\)-channel path, if \(i\) is not a ladder index for \(\{a_i, 1 \leq i \leq N\}\) (i.e., \(i\) is not equal to any \(k_j, 1 \leq j \leq m\)), then

\[
W_n^{(i)} = 0
\]

for each \(n \geq 1, i \geq 2\).

Thus we have shown that for any channel \(i, i \geq 2\), we have for each \(n \geq 1\),

\[
a_i \leq a_{i-1} \text{ or } a_{i-1} < a_i \leq a_{m_i} = W_n^{(i)} = 0.
\]

From Lemma 1 and Theorem 1 we conclude that in order to calculate the distributions of the waiting times at the channels, we need to consider just a “reduced” modified path. The modified path is generated from the original path by “short-circuiting” all the channels whose index does not correspond to a ladder index (to be called nonladder channels in contrast to ladder channels). Thus the modified path consists of a series connection of the ladder channels in order of increasing ladder indices. The first channel of the modified path is thus channel \(k_1 = 1\), the second one is channel \(k_2, 1 \leq k_2 \leq N\). The corresponding transmission times for the ladder channels are \(a_1 < a_2 < a_3 < \cdots < a_m\). Since the waiting time \(W_n^{(i)}\) depends on the arrival process through the interarrival time \(T_{n+1}^{(i)}\), as seen by (14) and (15), Lemma 1 implies that the waiting time \(W_n^{(k_i-1)}\) will assume the same values in the original path and in the modified path.

Considering the \(m\)-channel modified path, we obtain that the \(n\)th message arriving at any channel can find it free only if it has found the preceding channel free. Thus, if \(\{W_{n^{(k_i-1)}} > 0\}\), then also \(\{W_{n^{(k_i)}} > 0\}\). This is indicated by the following proposition.

**Proposition 1:** For ladder channels \(i, k_i = 2, \cdots, m\),

\[
\text{event } \{W_n^{(k_i)} = 0\} \text{ can occur only if } \{W_n^{(k_i-1)} = 0\} \text{ has occurred.}
\]

**Proof:** By (14), (15), and Lemma 1, if \(W_n^{(k_i-1)} > 0\), we have \(W_{n+1}^{(k_i)} = [W_n^{(k_i)} + a_{k_i} - a_{k_i-1}]^+ > 0\), since \(W_{n^{(k_i)}} > 0\) and \(a_{k_i} > a_{k_i-1}\). Hence \(\{W_{n^{(k_i-1)}} > 0\} \Rightarrow \{W_{n^{(k_i)}} > 0\}\).

Subsequently, event \(\{W_n^{(k_i)} = 0\}\) can occur only if \(\{W_n^{(k_i-1)} = 0\}\). Q.E.D.

It is interesting to observe that the latter property, when used recursively for each \(k_i \geq 2\), implies that \(\{W_n^{(k_1)} = 0\} = \{W_n^{(1)} = 0\}\), so that the following result can be stated.

**Corollary 1:** For any ladder index \(k_i, 2 \leq k_i \leq m, \text{ and any } n \geq 1\), we have

\[
\{W_n^{(k_i)} = 0\} \text{ only if } \{W_n^{(1)} = 0\}.
\]

Thus Corollary 1 indicates that a message can have a zero waiting time at any ladder channel \(k_i \geq 2\), only if it had a zero waiting time at channel 1 while entering the path. Clearly, at nonladder channels all the messages have zero waiting times. Note that Proposition 1 states a necessary but not sufficient condition for \(W_n^{(k_i)} = 0\).

V. DISTRIBUTION OF OVERALL PATH DELAY

Let

\[
S^{(k)}_n = \sum_{i=1}^{k} W^{(i)}_n, \quad n \geq 1, \quad k \geq 1
\]

denote the overall waiting time of the \(n\)th message over channels \((1,2,\cdots,k)\). To obtain the distribution of the \(m\)th message delay \(\gamma_m\) where

\[
\gamma_m = S^{(N)}_m + \sum_{i=1}^{N} a_i
\]

we first obtain the distribution of \(S^{(N)}_m\). For that purpose, we will utilize the following lemma.\(^{2}\)

\(^{2}\) After this paper was written, we learned that tandem queues with constant service times had also been studied independently in [18] and [19].
Lemma 2: The random variable $S_n^{(k)}$ satisfies the relationship, $k \geq 1, n \geq 1$, 

$$S_{n+1}^{(k)} = [S_n^{(k)} + \max(a_1, a_2, \ldots, a_n) - T_{n+1}^{(k)}]_+.$$  \hspace{1cm} (24)

Proof: We will prove (24) by induction on $k \geq 1$.

First consider the case $k = 1$. Then $W_n^{(1)} = [W_n^{(1)} + a_1 - T_n^{(1)}]_+$. Assume now that (24) holds for $1 \leq k < m$. Then $W_n^{(m+1)} = 0$, for each $n \geq 1$. Hence $S_n^{(m+1)} = S_n^{(m)} + \max(a_1, a_2, \ldots, a_m) - T_n^{(m)}$.

Now if $a_{n+1} > a_n$, we consider two cases.

In the first case, we assume that $S_n^{(k)} > 0$. Then, by (20), we have $W_n^{(k)} > 0$ and subsequently $W_n^{(m+1)} > 0$. Hence, by (18) and (14), $T_n^{(m+1)} = a_j$. Consequently, using the induction hypothesis, we obtain

$$S_{n+1}^{(k+1)} = S_{n+1}^{(k)} + W_n^{(k)}$$

so that (24) holds for $k = m+1$.

Consider now the second case where $a_{n+1} = a_n$. Then, by (20), we have $W_n^{(k)} = 0$ and subsequently $W_n^{(m+1)} = 0$. Hence, by (18) and (14), $T_n^{(m+1)} = a_j$. Consequently, using the induction hypothesis, we obtain

$$T_{n+1}^{(m+1)} = T_{n+1}^{(m)} + W_n^{(k)}$$

which when used recursively yields

$$T_n^{(m+1)} = T_n^{(1)} + \sum_{k=1}^{m} (T_{n+1}^{(k)} - T_n^{(k)}) = T_n^{(m+1)} - S_n^{(m)}$$

We thus obtain

$$S_{n+1}^{(m+1)} = W_n^{(m+1)}$$

$$S_{n+1}^{(m+1)} = [W_n^{(m+1)} + a_{n+1} - T_{n+1}^{(m+1)}]_+$$

$$S_{n+1}^{(m+1)} = [W_n^{(m+1)} + a_{n+1} - T_{n+1}^{(m+1)}]_+$$

which yields (24) for $k = m+1$. Q.E.D.

One readily observes in (24) that $T_n^{(1)}$ is statistically independent of $S_n^{(k)}$. Also recall that $T_n^{(1)}$ is the inter-arrival duration of a Poisson process with rate $\lambda$ and is, therefore, exponentially distributed with mean $\lambda^{-1}$. Hence, comparing (24) with the waiting-time recurrence relationship $W_n + a - T_n^{(1)}$ for the waiting time $W_n$ in an $M/D/1$ queuing system with service time $a$ and arrival rate $\lambda$, we obtain the following result for the distribution of $S_n^{(k)}$.

**Theorem 2:** The overall waiting time for the $m$th message at an $N$-channel path, $S_m^{(N)} = \sum_{i=1}^{N} W_n^{(i)}$, has the same distribution as the waiting time $W_n$ in an $M/D/1$ queuing system, with Poisson arrivals with rate $\lambda$ and service time equal to $a_{km} = \max(a_1, a_2, \ldots, a_N)$. If $\rho_{km} < 1$, the limiting distribution of the overall waiting time exists and is given by

$$S(x) = \lim_{m \to \infty} P(S_m^{(N)} \leq x)$$

$$= \lim_{m \to \infty} \left( \sum_{i=1}^{N} W_i^{(i)} \right)$$

$$= (1 - \rho_{km}) \sum_{i=0}^{\infty} \exp \left[ \lambda(x - ia_{km}) \right] (i!)^{-1}$$

$$\left[ -\lambda(x - ia_{km}) \right]^i$$

where $x = na_{km} + t, n \geq 0, 0 \leq t \leq a_{km}$. If $\rho_{km} \geq 1$, $\lim_{m \to \infty} P(S_m^{(N)} \leq x) = 0$, for each $x$.

In particular, expressions for the limiting average overall waiting time $W$ and overall average delay $\gamma$ for an $N$-channel path follow from (25) and (6). We also note that

$$W_n^{(k)} = S_n^{(k)} - S_n^{(k-1)}.$$  \hspace{1cm} (26)

Hence $E(W_n^{(k)}) = E(S_n^{(k)}) - E(S_n^{(k-1)})$, and we subsequently obtain expressions for the average waiting times and delays at the individual channels. The average queue size and memory size at each channel then follow by Little's theorem. (The latter theorem, see [4], asserts that $\lim_{m \to \infty} E(T_n^{(i)}) = \lambda(n_{in})_0$, where $\lambda(n_{in})_0 = \lim_{m \to \infty} E(T_n^{(i)})$. We have shown in [15] that for each $i, i = 1, 2, \ldots, N$.)

$$W_n^{(k)} = \frac{1}{2} \frac{\rho_{km}}{1 - \rho_{km}} a_{km} - \frac{1}{2} \frac{\rho_{km}}{1 - \rho_{km}} a_{km-1}$$  \hspace{1cm} (27)

where $\gamma^{(k)} = W^{(k)} + a_{km}$. (28)

which when used recursively yields

$$W_n^{(k)} = \frac{1}{2} \frac{\rho_{km}}{1 - \rho_{km}} a_{km} - \frac{1}{2} \frac{\rho_{km}}{1 - \rho_{km}} a_{km-1}$$  \hspace{1cm} (27)

where $\gamma^{(k)}$ is given by (28). For a nonladder channel $i$, $W^{(i)} = 0$ and $\gamma^{(i)} = a_i$. The overall average waiting time for the $N$-channel path $W$ is given by

$$W = \sum_{i=1}^{N} \frac{a_{km}}{2} - \frac{1}{2} \frac{\rho_{km}}{1 - \rho_{km}} a_{km}$$  \hspace{1cm} (30)

where $a_{km} = \max(a_1, a_2, \ldots, a_N)$ and $\rho_{max} = \lambda a_{max}$. The overall average delay for the $N$-channel path is equal to

$$\gamma = \frac{1}{2} \frac{\rho_{km}}{1 - \rho_{max}} a_{max} + \sum_{i=1}^{N} a_i.$$  \hspace{1cm} (31)

One observes that the distribution of $W_n^{(k)}$ cannot be directly calculated from (26) since $S_n^{(k)}$ and $S_n^{(k-1)}$ are not statistically independent. Using an appropriate embedded waiting-time sequence, the distribution of $W_n^{(k)}$ has been obtained by us in [15]. We also note that the number of messages transferred between two zeros of $\{S_n^{(k)}, n \geq 1\}$
are equal to the variable \( N_m^{(k_i)} \), for some \( m \geq 1 \), where \( N_m^{(k_i)} \) is the number of messages transferred through channel \( k_i \) during the \( m \)th busy period. Hence we conclude the following.

**Corollary 2:** For each ladder channel \( k_i, 1 \leq k_i \leq m, \rho_{k_i} < 1 \), the number of messages transferred through a busy period \( N_m^{(k_i)} \) follow distribution (11) with \( a_i \) replaced by \( a_{k_i} \). The duration of a busy period at channel \( k_i \) is given by \( B^{(k_i)} = N_m^{(k_i)}a_{k_i} \). For a nonladder channel \( i, N_m^{(i)} = 1, B_m^{(i)} = a_i \).

By the corollary and (12a), we further note that the average number of messages transferred during a busy period at ladder channel \( k_i \) is equal to \((1 - \rho_{k_i})^{-1}\), from which one readily concludes (see [15] for a detailed analysis) that for any channel \( i, \rho_i < 1 \), we have

\[
\lim_{n \to \infty} P(W_n^{(i)} = 0) = \lim_{i \to \infty} P(X_i^{(i)} = 0) = 1 - \rho_i. \tag{32}
\]

Hence, the \( i \)th channel in the communication path will be free with probability \( 1 - \rho_i \). We also notice (using Corollary 2 and (12a)) that the average number of busy periods at channel \( k_{i-1} \), which is included in a busy period at channel \( k_i \), is equal to \( E(N_m^{(k_{i-1})})/E(N_m^{(k_i)}) = \left(1 - \rho_{k_{i-1}}\right)/(1 - \rho_{k_i}) \).

It has been shown in [15] that the idle period at each channel is exponentially distributed with mean \( \lambda^{-1} \).

It is of particular interest to indicate that Lemma 2 implies the following interesting property, which follows by observing that in (24) only \( \max(a_1, \ldots, a_n) \) is utilized to evaluate the overall waiting time in the path.

**Theorem 4:** (Capacity ordering invariance property) The overall waiting time over an \( N \)-channel path with capacities \((C_1, C_2, \ldots, C_N)\) is the same as that over an \( N \)-channel path with capacities \((C_1', C_2', \ldots, C_N')\), where the latter sequence is an arbitrary ordering of \((C_1, \ldots, C_N)\). The overall waiting time depends only on the minimal capacity \( \min(C_1, \ldots, C_N) \).

We notice that Theorem 4 implies the result of Theorem 2, since one may order the given service sequence \((a_1, \ldots, a_N)\) so that the resulting sequence has \( a_{m_2} = \max(a_1, \ldots, a_N) \), as in the first service time. For the latter case, we have

\[
W_m^{(i)} = 0, \quad \text{for each } m \geq 1 \text{ and each } i, \quad 2 \leq i \leq N,
\]

with \( W_1^{(i)} \) being the waiting time for an \( M/D/1 \) system with service time \( a_{m_2} \). The overall waiting-time distribution given by Theorem 2 subsequently follows.

**VI. Example**

To illustrate the application of our results to communication networks, consider the network shown by Fig. 3 with channel capacities (in kbits/s) as indicated there. Let the message length be \( \beta = 1 \) kbits/message (i.e., for example, the approximate packet length in the ARPA computer network [8]). We wish to consider the overall average delays resulting when we transmit messages from \( v_1 \) to \( v_5 \) through a communication path and obtain the path that yields the minimal overall delay.

For the given network, there are four paths between \( v_1 \) and \( v_5 \), denoted by \( \pi_1 = v_1v_2v_3v_4v_5, \pi_2 = v_1v_2v_5, \pi_3 = v_1v_3v_5, \) and \( \pi_4 = v_1v_4v_5 \). The overall transmission times

\[
A = \sum d_i, \quad \text{and the minimal capacity } \min C_i \text{ are given for each path by}
\]

\[\begin{align*}
A(\pi_1) &= 0.11, \quad \text{s/message, min } C(\pi_1) = 25, \text{ kbits/s} \\
A(\pi_2) &= 0.08, \quad \text{s/message, min } C(\pi_2) = 20, \text{ kbits/s} \\
A(\pi_3) &= 0.07, \quad \text{s/message, min } C(\pi_3) = 50/3, \text{ kbit/s} \\
A(\pi_4) &= 0.10, \quad \text{s/message, min } C(\pi_4) = 20, \text{ kbits/s.}
\end{align*}\]

The average delay \( \gamma(\pi) \) is now readily calculated for each path using (31), for any arrival rate \( \lambda \). To obtain the optimal (minimal-delay) path, we first observe that \( \gamma(\pi_1) < \gamma(\pi_2), \) for each \( \lambda < 20 \) message/s, and \( \gamma(\pi_2) = \gamma(\pi_4) = \infty, \) for \( \lambda \geq 20 \) message/s. We need thus only compare \( \pi_1, \pi_2, \) and \( \pi_3. \) By (31), we obtain

\[\begin{align*}
\gamma(\pi_1) &= 0.11 + \frac{1}{2 \beta - 0.04} \\
\gamma(\pi_2) &= 0.08 + \frac{1}{2 \beta - 0.05} \\
\gamma(\pi_3) &= 0.07 + \frac{1}{2 \beta - 0.06}
\end{align*}\]

from which one concludes that

\[\begin{align*}
\gamma(\pi_3) < \gamma(\pi_2) < \gamma(\pi_1), & \quad \text{if } \lambda < \lambda_0 \\
\gamma(\pi_2) < \gamma(\pi_3) < \gamma(\pi_1), & \quad \text{if } \lambda_0 < \lambda < \lambda_1 \\
\gamma(\pi_2) < \gamma(\pi_1) < \gamma(\pi_3), & \quad \text{if } \lambda_1 < \lambda < \lambda_2 \\
\gamma(\pi_1) < \gamma(\pi_2) < \gamma(\pi_3), & \quad \text{if } \lambda_2 < \lambda < 20 \\
\gamma(\pi_1) < \gamma(\pi_2) = \gamma(\pi_3) = \infty, & \quad \text{if } 20 \leq \lambda < 25
\end{align*}\]

where \( \lambda_0 = 7.6, \lambda_1 \approx 10.8, \lambda_2 \approx 13.7 \) message/s.

Thus for incoming message rates less than \( \lambda_0, \) path \( \pi_3 \) will be utilized since its overall transmission time \( A(\pi_3) \) is the shortest and the overall waiting time is not high. For high intensity rates \( \lambda_2 < \lambda < 25, \) path \( \pi_1 \) is optimal since at such rates the overall waiting times become the major delay factor and \( \pi_1 \) yields, for each \( \lambda, \) the lowest waiting time since its minimal capacity is the largest.

**Fig. 3.** Communication network for example. Channel capacities are in kbits/s.
VII. CONCLUSIONS

We have solved for the steady-state distributions of the message overall delays along paths in a communication network. The average memory storage requirements and waiting times at the stations have also been obtained. The following two points are readily observed.

1) Lemma 2, (24), holds for any incoming message point process. Hence, the distribution of the overall waiting time $S^{(n)}_w$ is deduced from (24), for any incoming message process (the message lengths are still assumed to be of fixed length). In particular, if the incoming messages follow the statistics of a renewal point process (i.e., \{T_{kn}^{(n)}; n \geq 1\} in (24) are independently identically distributed random variables), results from GI/D/1 queueing system theory are used to obtain the limiting waiting-time distribution and moments (see, for example, [1 ch. II.5 and II.6]).

2) The capacity assignment problem is readily solved. Thus assume that the total capacity over the $N$-channel path is given, $\sum_{i=1}^{N} C_i = C$ (so that $C > N\lambda \beta$ bits/s, to avoid infinite delays). We wish to find the values of the individual capacities $C_i$, $i = 1, 2, \ldots, N$, so that the overall average delay is minimized. The delay $\gamma$ is given by (31). Since the overall waiting time depends only on the value of the minimal capacity, we wish to choose the largest possible value of the minimal capacity. Hence, the overall waiting time is minimized by choosing equal capacities $C_i = C/N$, $i = 1, 2, \ldots, N$. The overall transmission time is $\sum_{i=1}^{N} a_i = \beta \sum_{i=1}^{N} C_i^{-1}$, which (by symmetry) is readily observed to be minimized as well by choosing equal channel capacities. Consequently, the average delay $\gamma$ is minimized by choosing equal channel capacities along the communication path.

Waiting-time distributions at the individual channels of the communication path, as well as other statistical characteristics of the queueing processes at the channels, have been derived by us in [15]. Also, assuming message lengths are governed by an arbitrary distribution, we have recently obtained overall message path delay distributions in packet-switching communication networks [16] extending the analysis in this paper, as well as waiting-time and busy-period characteristics at the individual channels [17] and studied multiterminal path flows in such networks [20]–[21].

Further time delay problems for communication networks are currently under investigation. While message flows in many realistic communication networks can be assumed to be nonnonsimultaneous, so that queueing delays are generated due to message random bursts in a specific source-sink flow (as for some satellite communication networks), in many other networks (such as computer communication networks) simultaneous flow considerations are of major importance. The techniques and results presented here and in [15] have proved powerful tools for further studies. In particular, using time-delay and memory storage considerations, analysis and synthesis problems are being studied for more general message flows in communication networks.

REFERENCES