Singapore Management University
Institutional Knowledge at Singapore Management University

Research Collection School Of Information Systems | School of Information Systems

7-1993

Performance of a token-passing system with batch arrivals and its application to file transfers

Robert H. DENG
Singapore Management University, robertdeng@smu.edu.sg

Xuanyu ZHANG
Illinois Institute of Technology

Kuan Tase HUANG
National University of Singapore

DOI: https://doi.org/10.1016/0140-3664(93)90103-Y

Follow this and additional works at: https://ink.library.smu.edu.sg/sis_research

Part of the Information Security Commons

Citation

DENG, Robert H.; ZHANG, Xuanyu; and HUANG, Kuan Tase. Performance of a token-passing system with batch arrivals and its application to file transfers. (1993). *Computer Communications*. 16, (7), 422-431. Research Collection School Of Information Systems.
Available at: https://ink.library.smu.edu.sg/sis_research/196

This Journal Article is brought to you for free and open access by the School of Information Systems at Institutional Knowledge at Singapore Management University. It has been accepted for inclusion in Research Collection School Of Information Systems by an authorized administrator of Institutional Knowledge at Singapore Management University. For more information, please email libIR@smu.edu.sg.

Published in Computer Communications, Volume 16, Issue 7, July 1993, Pages 422-431.
https://doi.org/10.1016/0140-3664(93)90103-Y

Performance of a token-passing system with batch arrivals and its application to file transfers

Robert H Deng*, Xianyu Zhang† and Kuan-Tase Huang‡

This paper investigates the performance of token-passing systems with limited service and Poisson arrivals. For pure Poisson arrivals, the Laplace–Stieltjes Transform (LST) of an approximate customer/packet waiting time distribution is derived and expressed as a functional equation, from which the approximate mean and variance of waiting time are obtained; for batch Poisson arrivals, an approximate mean of waiting time is derived. Mean waiting time approximations are compared against both simulation results and other results presented in the literature; the variance approximation is evaluated by comparing it to simulation results only, since no other results on variance have been reported so far. As application examples of our analytical results, multimedia file transfers over token-passing LANs are modelled and studied.

Keywords: file transfer, LAN, token-passing system, performance evaluation

Token-passing is a widely used LAN access method in ring and bus networks[1]. Two types of token-passing operations are in common use today, multiple- and single-token operation[2]. In multiple-token operation, the transmitting station generates a new free token and places it on the network immediately after the last bit of transmitted data. As its name suggests, this type of operation permits several busy tokens and one free token on the network at one time. In contrast, single-token operation requires that a transmitting station waits until it has erased its own busy token before generating a new free token. The single-token operation ensures that there is only one token on the network at any given time. When the packet length to network latency ratio is greater than one, the single- and multiple-token operations are essentially the same; however, as this ratio becomes less than one (e.g. due to an increase of network transmission capacity), multiple-token operation becomes superior. Note that the ANSI/IEEE 802.5 token-ring LAN[3] uses single-token operation, while ANSI's FDDI[4] implements the multiple-token operation to take advantage of its 100 Mbits/s transmission rate. Token-passing systems can also be classified according to their service schemes: exhaustive, gated and limited services. In the exhaustive service scheme, the queue at a station must be empty before the token is passed to the next station. In the gated service scheme, only those messages in the queue at the time of the token's arrival are served. Messages that arrive while the token is already at the station will be served in the next cycle. Finally, in the limited service scheme, a specified maximum number of messages/packets in the queue, denoted k, will be served upon the arrival of the token.

The nature of the traffic offered by stations to a network is a major factor in determining system performance. Many researchers have noted that the pure Poisson traffic assumption may result in a quite dramatic error in performance estimation for some real systems. For example, in file transfer systems it is more appropriate to assume batch traffic arrivals than pure Poisson arrivals. In this paper, we focus on the token-passing protocol with multi-token operation, limited service (limited-to-one), and either pure Poisson or Poisson batch arrivals.

The queueing model of the token-passing system or polling system is a single-server multi-queue system

*Department of Electrical Engineering, National University of Singapore, Kent Ridge, Singapore 0511
†Department of Computer Science, Illinois Institute of Technology, Chicago, IL 60616, USA
‡Institute of Systems Science, National University of Singapore, Kent Ridge, Singapore 0511

with a cyclic service discipline. In queueing terminology, an active station is a queue, the token is the server, and the overhead associated with sending a token from one station to the next is the polling/walking time. Throughout the paper, we use the generic terms *customer*, *batch* (of customers), and *service time* to represent packet, message (a batch of packets) and packet transmission time, respectively. Multi-queue systems with cyclic service have been studied by many authors. Most of the key results concerning M/G/1 multi-queue systems, both approximate and exact, can be found in References 5–12. The excellent survey by Takagi[13] contains several further references.

The $M^{[x]}/G/1$ batch arrival multi-queue system has been discussed by Kuehn[14], who obtained mean waiting time approximations using the imbedded Markov chain technique; however, the approximation error of the Kuehn model becomes unacceptably large (over 30%) at medium-to-high offered loads[12]. Using diffusion approximations, Fischer[15] and Kimura and Takahashi[16] have studied mean waiting time approximations for the $M^{[x]}/G/1$ and $GI^{[x]}/G/1$ batch arrival models, respectively. Recently, Ibe and Chen[12] have obtained a very good approximation to the mean waiting time for the M/G/1 multi-queue system by utilizing the concept of sample cycle times. In this paper, we generalize the result of Ibe and Chen[12] to the study of $M^{[x]}/G/1$ models. As for their results, our result also clearly shows the dependence of the mean waiting time on the order in which the stations are located in the network. Neither the Kuehn model nor the Fischer, Kimura and Takahashi models are able to show this dependence. Furthermore, our approximations are more accurate than the previous results, as will be seen in our numerical examples.

TOKEN-PASSING MODEL

The multi-queue model of the token-passing system considered in this paper is depicted in *Figure 1*. The server (token) serves N queues (with infinite buffer capacities) in a cyclic manner. The service discipline is limited service (limited-to-one). When the server visits

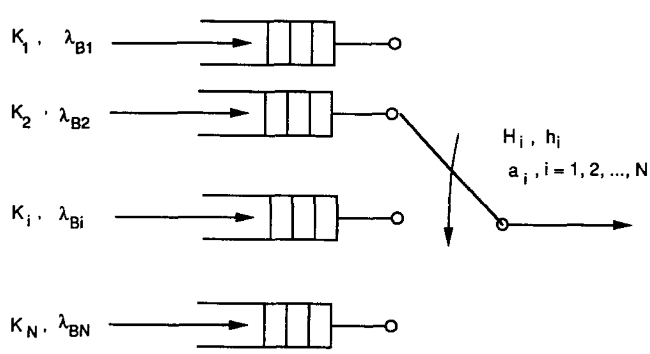

Figure 1 Queueing model of token passing system

a queue, say the ith queue, it only serves one customer if any is present; otherwise, it moves to the subsequent $(i + 1)$th queue (when the server is in the Nth queue, it moves to the first queue). The walking/polling times of the server between the ith and the $(i + 1)$th queue are independent, identically distributed (i.i.d.) random variables A_i with distribution $F_{Ai}(\bullet)$, mean a_i and nth moment $a_i^{(n)}$. The mean of the total polling time during a cycle of the server t_0 is given by:

$$t_0 = \sum_{i=1}^{N} a_i \tag{1}$$

where t_0 is also called system latency of the token-passing system.

Batches of customers arrive at all queues according to independent Poisson processes with rates λ_{B1}, $\lambda_{B2}, \ldots, \lambda_{BN}$. Batch sizes are i.i.d. random variables K_i with distribution:

$$b_{ik} = \Pr\{K_i = k\}, \quad k = 0, 1, 2, \ldots \tag{2}$$

and first and second moments denoted by $E[K_i]$ and $E[K_i^2]$. Customers which arrive at the ith queue are called type-i customers. The mean arrival rate of type-i customers is given by $\lambda_i = \lambda_{Bi}E[K_i]$. Service times of type-$i$ customers are i.i.d. random variables H_i with distribution $F_{Hi}(\bullet)$, mean h_i and the nth moment $h_i^{(n)}$; the service process is also independent of the arrival process and of the polling process.

The utilization at the ith queue ρ_i is defined as:

$$\rho_i = \lambda_{Bi}E[K_i]h_i, \quad i = 1, 2, \ldots, N \tag{3}$$

The total utilization of the server ρ is defined as:

$$\rho = \sum_{i=1}^{N} \rho_i \tag{4}$$

This paper deals only with the steady state of the multi-queue system. It was shown by Kuehn[14] that the following conditions are needed for stability of the system:

$$\rho < 1 \quad \text{and} \quad \max(\lambda_i) \sum_{i=1}^{N} a_i < 1 - \rho \tag{5}$$

CYCLE TIME ANALYSIS

The scanning epoch at a given queue is defined as the instant the server arrives at the queue. For any fixed queue, say the ith queue, let C_i denote the random cycle time between two consecutive scanning epochs at the ith queue. We refer to such a cycle as an i-cycle. During an i-cycle, all queues receive a service opportunity; however, not all queues may have customers waiting to be served. We observe that C_i is independent of the queue index i[6] and we will use C to denote the cycle time random variable for any queue i, $i = 1, 2, \ldots, N$. The mean of the cycle time C, $E[C]$, is given by[6, 14]:

$$E[C] = t_0/(1 - \rho) \tag{6}$$

The conditional cycle time D_i, first introduced by Kuehn[14], is defined as the i-cycle time assuming that the ith queue has a customer service time contribution to the i-cycle. The mean of D_i, $E[D_i]$, is given by[14]:

$$E[D_i] = (t_0 + h_i)/(1 - \rho + \rho_i) \tag{7}$$

Let α_{dik} be the probability for the service of a type-k customer ($k \neq i$) during the conditional cycle time D_i. It can be shown that[14]:

$$\alpha_{dik} = \lambda_k E[D_i] \tag{8}$$

The exact solution of the conditional cycle time distribution $F_{Di}(\bullet)$ is still unknown except for its mean $E[D_i]$ in equation (7). Using the independence assumption, Kuehn[14] gave the following approximation for the LST of the conditional cycle time distribution:

$$\Phi_{Di}(s) = \prod_{k=1}^{N} \Phi_{Ak}(s) \prod_{k \neq i}^{N} (\alpha_{dik}\Phi_{Hk}(s) + (1 - \alpha_{dik}))\Phi_{Hi}(s) \tag{9}$$

where $\Phi_{Ak}(s)$ and $\Phi_{Hk}(s)$ are the LST of the polling time distribution $F_{Ak}(\bullet)$ and the service time distribution $F_{Hk}(\bullet)$, respectively. From equation (9), the second moment of the conditional cycle time can be evaluated as:

$$E[D_i^2] = \sum_{k=1}^{N} \mathrm{Var}(A_k) + \sum_{k \neq i}^{N} (\alpha_{dik} h_k^{(2)} - \alpha_{dik}^2 h_k^2) + \mathrm{Var}(H_i) + E[D_i]^2 \tag{10}$$

Motivated by Ibe and Cheng[12], in the following we introduce the concept of conditional sample cycle times U_j and V_j, $j = 1, 2, \ldots, N$, where U_j is defined as the cycle time seen by an outside observer assuming that the server is serving the jth queue at the observation epoch, and V_j is the cycle time seen by an outside observer assuming that the server is polling the jth queue (walking from the $(j-1)$th queue to the jth queue) at the observation epoch. We remark that U_j and V_j are not arbitrary cycle times, but cycle times 'sampled' by an outside observer.

The mean $E[U_j]$ of the conditional sample cycle time U_j can be obtained by investigating the flow balance of the system. Let α_{ujk} be the probability for the service of a type-k customer ($k \neq j$) during the conditional sample cycle time U_j. Using a similar argument as in Kueln[14] and Kumura and Takahashi[16], we can show that:

$$\alpha_{ujk} = \lambda_k E[U_j], \quad k \neq j \tag{11}$$

At steady state, the mean number of arriving type-k customers per sample cycle is equal to the mean number of served type-k customers per sample cycle. From this it follows that:

$$E[U_j] = \sum_{k=1}^{N} a_k + 2h_{0j} + \sum_{k \neq j}^{N} \alpha_{ujk} h_k \tag{12}$$

where:

$$h_{0j} = \frac{h_j^{(2)}}{2h_j} \tag{13}$$

is the mean residual life of the type-j customer service time found in service by the outside observer. Inserting (11) into (12) we find:

$$E[U_j] = \frac{(2h_{0j} + t_0)}{(1 - \rho + \rho_j)} \tag{14}$$

and

$$\alpha_{ujk} = \frac{\lambda_k(2h_{0j} + t_0)}{(1 - \rho + \rho_j)}, \quad k \neq j \tag{15}$$

Let α_{vjk} be the probability for the service of a type-k customer during the conditional sample cycle time V_j. Using the same approach as above we have:

$$E[V_j] = \frac{(2a_{0j} + t_0 - a_j)}{(1 - \rho)} \tag{16}$$

and:

$$\alpha_{vjk} = \frac{\lambda_k(2a_{0j} + t_0 - a_j)}{(1 - \rho)} \tag{17}$$

where:

$$a_{0j} = \frac{a_j^{(2)}}{2a_j} \tag{18}$$

is the mean residual life of the jth queue's polling time A_j.

WAITING TIME ANALYSIS

In this section, we analyse the multi-queue model presented earlier. Although the following solution is formally exact, the analysis approach is approximate due to the independence assumption, i.e. we assume that the cycle times C, D_i, U_j and V_j are all i.i.d. random variables.

Residual life R_i analysis

Consider a tagged customer that arrives at the ith queue. The residual life R_i of an i-cycle time seen by the tagged customer is defined as the time interval from the instant the tagged customer arrives at the ith queue until the completion of the i-cycle (i.e. the completion of the first polling of the ith queue since the arrival of the tagged customer). Since customer batch arrivals follow Poisson distribution, the residual life R_i is statistically identical to the residual life of a sample i-cycle seen by an outside observer.

Let R_{uji} be the residual life R_i conditional on the server was serving a type-j customer when the tagged customer arrived at the ith queue. Also, let R_{vji} be the residual life R_i conditional on the server was polling the jth queue when the tagged customer arrived at the ith queue. We observe that the random variables R_{uji} and R_{vji} are statistically identical to the residual life of U_j and V_j, respectively.

Let L_{ukj} be a Bernoulli random variable defined on $\{0, 1\}$, with $L_{ukj} = 1$ if there is a type-k customer service during the conditional sample cycle time U_j; and $L_{ukj} = 0$ otherwise, i.e.:

$$L_{ukj} = \begin{cases} 1, & \Pr\{L_{ukj} = 1\} = \alpha_{ukj} \\ 0, & \Pr\{L_{ukj} = 1\} = 1 - \alpha_{ukj} \end{cases} \tag{19}$$

where α_{ukj} is given in equation (15). Then the conditional residual life R_{uji} can be expressed as:

$$R_{uji} = \sum_{k=j+1}^{i} A_k + \sum_{k=j+1}^{i-1} L_{ukj} H_k + H_{0j} \tag{20}$$

where the random variable H_{0j} is the residual life of type-j customer service time. In the above summations we have used the following conventions: when $k > N$, $k = k(\mathrm{mod}\, N)$; when $i < j + 1$, the first summation will be taken from $j + 1$ to N, and then from 1 to i; and when $i - 1 < j + 1$, the second summation will be taken from $j + 1$ to N, and then from 1 to $i - 1$. Taking the LST of equation (20) and using the independence assumption, we obtain:

$$\Phi_{Ruji}(s) = \prod_{k=j+1}^{i} \Phi_{Ak}(s) \prod_{k=j+1}^{i-1} (\alpha_{ukj}\Phi_{Hk}(s) + (1 - \alpha_{ukj}))\frac{1 - \Phi_{Hj}(s)}{sh_j} \tag{21}$$

where the last term is the LST of the residual life distribution of type-j customer service times. Similarly, the LST of the distribution of R_{vji}, $\Phi_{RVji}(s)$, can be obtained as:

$$\Phi_{Rvji}(s) = \prod_{k=j+1}^{i} \Phi_{Ak}(s) \prod_{k=j}^{i-1} (\alpha_{vkj}\Phi_{Hk}(s) + (1 - \alpha_{vkj}))\frac{1 - \Phi_{Aj}(s)}{sa_j}, \quad j \neq i \tag{22a}$$

and:

$$\Phi_{Rvji}(s) = \frac{1 - \Phi_{Aj}(s)}{sa_j}, \quad j = i \tag{22b}$$

where the last term is the LST of the residual life distribution of the jth queue's polling time, and where α_{vkj} is given in equation (17).

Define p_{uj} and p_{vj} as the probabilities that the server is serving a type-j customer and that the server is polling the jth queue, respectively, when the tagged customer arrives at the ith queue. Following the approach of Ibe and Chang[12], and using the fact that customer batch arrivals are Poisson distributed, it can be shown that:

$$p_{uj} = \lambda_j h_j \tag{23}$$

and:

$$p_{vj} = \frac{a_j}{E[C]} \tag{24}$$

Note that $\Sigma(p_{uj} + p_{vj}) = 1$. Then the residual life of the sample cycle time seen by the tagged customer is:

$$R_i = \sum_{j=1}^{N} \{p_{uj}R_{uji} + p_{vj}R_{vji}\} \tag{25}$$

and the LST of its distribution is:

$$\Phi_{Ri}(s) = \sum_{j=1}^{N} \{p_{uj}\Phi_{Ruji}(s) + p_{vj}\Phi_{Rvji}(s)\} \tag{26}$$

Let $\mathrm{Var}(H_{0j})$ denote the variance of the residual life of type-j customer service time. Let $\mathrm{Var}(A_{0j})$ denote the variance of the residual life of the jth queue's polling time. From renewal theory it is known that[18]:

$$\mathrm{Var}(H_{0j}) = \frac{h_j^{(3)}}{3h_j} - \left(\frac{h_j^{(2)}}{2h_j}\right)^2 \tag{27}$$

$$\mathrm{Var}(A_{0j}) = \frac{a_j^{(3)}}{3a_j} - \left(\frac{a_j^{(2)}}{2a_j}\right)^2 \tag{28}$$

From equation (26), the mean $E[R_i]$ and the second moment $E[R_i^2]$ of the residual life R_i can be evaluated. After some calculations, we obtain from equation (26) that:

$$E[R_i] = \sum_{j=1}^{N} \{p_{uj}E[R_{uji}] + p_{vj}E[R_{vji}]\} \tag{29}$$

where:

$$E[R_{uji}] = \sum_{k=j+1}^{i} a_k + \sum_{k=j+1}^{i-1} \alpha_{ukj}h_k + h_{0j} \tag{30}$$

$$E[R_{vji}] = \begin{cases} \sum_{k=j+1}^{i} a_k + \sum_{k=j}^{i-1} \alpha_{vkj}h_k + a_{0j} & j \neq i \\ a_{0j}, & j = i \end{cases} \tag{31}$$

and that:

$$E[R_i^2] = \sum_{j=1}^{N} \{p_{uji}E[R_{uji}^2] + p_{vji}E[R_{vji}^2]\} \tag{32}$$

where:

$$E[R_{uji}^2] = \sum_{k=j+1}^{i} \mathrm{Var}(A_k) + \sum_{k=j+1}^{i-1} (\alpha_{ukj}h_k^{(2)} - \alpha_{ukj}^2 h_k^2) + \mathrm{Var}(H_{0j}) + E[R_{uji}]^2 \tag{33}$$

$$E[R_{vji}^2] = \begin{cases} \sum_{k=j+1}^{i} \mathrm{Var}(A_k) + \sum_{k=j}^{i-1} (\alpha_{vkj}h_k^{(2)} - \alpha_{vkj}^2 h_k^2) \\ \quad + \mathrm{Var}(H_{0j}) + E[R_{uji}]^2, & j \neq i \\ \mathrm{Var}(A_{0j}) + E[R_{vji}]^2, & j = i \end{cases} \tag{34}$$

Functional equation on customer waiting time for systems with pure Poisson arrivals

Consider a tagged customer that is the kth arrival at the ith queue, and denote:

- $W_i(k)$: the waiting time in the ith queue of the tagged customer.
- $R_i(k)$: the residual life of the sample cycle seen by the tagged customer.
- $D_i(l)$: the conditional i-cycle time due to the service of the lth customer at the ith queue (if any is present), $l < k$.
- $N_i(k)$: the number of customers found waiting in the ith queue (not including the one in service) by the tagged customer upon arrival.
- $p_{ij}(k)$: the probability that $N_i(k) = j$.

We have:

$$W_i(k) = R_i(k) + \sum_{j=0}^{\infty} p_{ij}(k) \sum_{l=k-j}^{k-1} D_i(l) \tag{35}$$

Assuming that the ith queue is in steady state (i.e. $k \to \infty$), then $W_i(k)$, $R_i(k)$, and $p_{ij}(k)$ are independent of the index k, and $D_i(l)$ is independent of the index l, and they can be written as W_i, R_i, p_{ij} and D_i, respectively. Taking the LST of the above equation and using the independence assumption, it follows that:

$$\Phi_{Wi}(s) = \Phi_{Ri}(s) \sum_{j=0}^{\infty} p_{ij}(\Phi_{Di}(s))^j \tag{36}$$

In equation (36), $\Phi_{Wi}(s)$ is the LST of $F_{Wi}(\bullet)$, the customer waiting time distribution at the ith queue, and $\Phi_{Ri}(s)$ and $\Phi_{Di}(s)$ are given in equations (26) and (9), respectively.

Introducing the probability generating function for the state distribution p_{ij} at the ith queue, $j = 0, 1, 2, \ldots$:

$$P_i(z) = \sum_{j=0}^{\infty} p_{ij}z^j \tag{37}$$

Equation (36) can be rewritten as:

$$\Phi_{Wi}(s) = \Phi_{Ri}(s)P_i(\Phi_{Di}(s)) \tag{38}$$

To proceed with our analysis, we now derive a relationship between the generating function $P_i(z)$ and the LST $\Phi_{Wi}(s)$. Regarding the ith queue as an M/G/1 system, in the steady state, the queue appears statistically identical to an arriving and a departing customer. Also, the number of customers in the ith queue left behind by the tagged departing customer can alternatively be considered as the number of arriving customers during the waiting time of the tagged customer. Therefore, the distribution p_{ij} is the same as the distribution of the number of arriving type-i customers during the waiting time of the tagged customer. With this observation, it is easy to show that:

$$P_i(z) = \Phi_{Wi}[\lambda_i(1 - z)] \tag{39}$$

Substituting equation (39) into (38), it follows that:

$$\Phi_{Wi}(s) = \Phi_{Ri}(s)\Phi_{Wi}(\lambda_i[1 - \Phi_{Di}(s)]) \tag{40}$$

This result gives the LST for the customer waiting time distribution expressed as a functional equation. The functional equation (40) is usually impossible to invert. However, the various moments of the customer waiting time can easily be obtained. In particular, the mean and the variance of the customer waiting times are given by:

$$E[W_i] = \frac{E[R_i]}{(1 - \lambda_i E[D_i])} \tag{41}$$

$$\mathrm{Var}[W_i] = \frac{E[R_i^2] + \lambda_i E[W_i](E[D_i^2] + 2E[R_i]E[D_i])}{1 - \lambda_i^2 E[D_i]^2} - E[W_i]^2 \tag{42}$$

respectively. Equation (41) is identical to the result obtained by Ibe and Cheng[12].

Mean waiting time for systems with batch Poisson arrivals

The functional equation approach presented above does not apply to systems with batch arrivals, since the queue with batch arrivals no longer appears statistically identical to an arriving and departing customer. In the following, we derive the mean customer waiting time for systems with batch Poisson arrivals using a mean value approach.

Let $E[W_i(1)]$ be the mean waiting time of the first customer in an arbitrary message at the ith queue. Let

N_i be the number of customers found waiting at the ith queue by a batch upon arrival. We have:

$$E[W_i(1)] = E[R_i] + E[N_i]E[D_i] = E[R_i] + \lambda_i E[D_i]E[W_i] \quad (43)$$

where in the last step we have used Little's formula. From Kuehn[14], for batch arrivals we also have:

$$E[W_i] = E[W_i(1)] + \frac{E[D_i]}{2}\left(\frac{E[K_i^2]}{E[K_i]} - 1\right) \quad (44)$$

Substituting equation (44) into (43) and solving for $E[W_i]$ we arrive at:

$$E[W_i] = \frac{E[R_i] + E[D_i](E[K_i^2]/E[K_i] - 1)/2}{1 - \lambda_i E[D_i]} \quad (45)$$

NUMERICAL RESULTS

In this section, we present numerical examples in order to discuss some general characteristics of the multi-queue model, and to show the degree of accuracy of our approximate solution. The mean waiting time approximation is compared against simulation results generated by the IBM RESQ2 package[19], and other results presented in the literature; the variance approximation is evaluated by comparing it to simulation results only, since no other results on this have been reported so far. The numerical results will be given for two types of models: (1) M/G/1 multi-queue models (*Tables 1* and *2*); and (2) $M^{[x]}/G/1$ multi-queue models (*Tables 3-6*). Throughout the following examples, it is assumed that the number of queues $N = 10$; the mean customer service time $h_i = 1$, and the queue walking/polling time is exponentially distributed with mean $a_i = 0.05$, $i = 1, 2, \ldots, N$.

A more detailed discussion of the results follows. *Table 1* shows results of our approximate mean $E[W_i]$ and standard derivation σ_{Wi} of customer waiting time with respect to ρ for the M/M/1 multi-queue. The table also includes simulation results on $E[W_i]$ and σ_{Wi}, the exact and approximate mean waiting time obtained from Takagi[10] and Boxma and Meister[11], and from Kuehn[14], respectively. From this example, as well as from the rest of the examples, we observe that the proposed approximation on σ_{Wi} gives an underestimation, and that it is quite close to the simulation result at low to moderate values of ρ. *Table 2* shows results from the nonsymmetric M/M/1 multi-queue ($\lambda_1 = 6\lambda, \lambda_i = \lambda, i \geq 2$). Both the simulation results and our approximations indicate that identical queues do not necessarily have the same mean and variance of waiting time; they depend on a queue's position relative to the queue with the highest customer arrival rate. In this example, queue 1 has the highest arrival rate and all other queues have the same arrival rate. The mean and standard derivation of customer waiting time decrease almost uniformly from queue 2 to queue 10 (for compactness, only queues 2, 4, 6, 8 and 10 are listed). For an explanation of this phenomenon, the reader is directed to Box and Truong[8] and to Ibe and Cheng[12].

Tables 3-6 show results for $M^{[x]}/G/1$ multi-queues. The tables give mean customer waiting time obtained from simulation, our approximations, and the results of Kimura and Takahashi[16] and Kuehn[14]. *Tables 3* and *4* show results for symmetrical $M^{[x]}/M/1$ multi-queues with constant and geometrically distributed batch sizes, respectively. All three approximations on $E[W_i]$ are quite accurate; however, our approximation is the closest to the simulation.

The results on nonsymmetrical $M^{[x]}/M/1$ multi-queues with constant and geometrically distributed batch size are given in *Tables 5* and *6*, respectively. From both our approximation results and simulation results we have again observed the waiting time dependence on the queue position for statistically identical queues; however, the dependence is far less weak compared to pure Poisson arrivals. Therefore, mean customer waiting times at consecutive queues with identical characteristics (i.e. queues from 2 to 10) are only represented by their average in the tables.

FILE TRANSFER OVER TOKEN-PASSING SYSTEMS

We now apply the analytical results developed above to evaluate the performance of multimedia file transfer

Table 1 Symmetrical M/M/1 multi-queue

ρ	Exact			Our model		Kuehn's model[14]
	$E[W_i]^a$	$E[W_i]^b$	$\sigma_{Wi}{}^b$	$E[W_i]^c$	σ_{Wi}	$E[W_i]$
0.1	0.41899	0.41364 (4.7%)	0.58118	0.41902	0.58391	0.40775
0.3	0.83942	0.82440 (5.0%)	1.27926	0.83971	1.16764	0.72901
0.5	1.63158	1.60263 (6.1%)	2.21910	1.63276	1.93800	1.21926
0.7	3.67925	3.62540 (3.8%)	4.73152	3.68341	3.60808	2.35258
0.9	21.36372	21.11024 (5.6%)	23.94623	21.39689	17.75597	13.16488

[a]Based on Boxma–Meister formula
[b]Based on simulation
[c]Based on Ibe–Chen formula
Note: A confidence level of 95% was used in the simulation: the percentage within parentheses denotes the relative width of the confidence interval.

Table 2 Asymmetrical M/M/1 multi-queue

ρ	Simulation		Our model		Kimura[16]	Kuehn[14]
	$E[W_i]$	σ_{Wi}	$E[W_i]^a$	σ_{Wi}	$E[W_i]$	$E[W_i]$
			Queue 1			
0.2	0.67249 (2.2%)	1.05357	0.65994	0.95605	0.661654	0.62641
0.4	1.47354 (7.7%)	2.29467	1.46073	1.83008	1.47229	1.26901
0.6	4.13450 (4.2%)	5.38628	3.84863	3.95958	3.91209	2.99000
0.8	44.93304 (8.9%)	44.58387	39.20802	32.41350	39.54176	25.71818
			Queue 2			
0.2	0.58972 (3.4%)	0.92633	0.58968	0.84592	0.58583	0.54675
0.4	1.10762 (2.1%)	1.60931	1.10271	1.41082	1.07604	0.89456
0.6	2.22369 (2.6%)	2.48569	2.13769	2.28341	1.98178	1.46153
0.8	6.17162 (3.0%)	6.53270	5.56162	4.84830	4.44515	3.19629
			Queue 4			
0.2	0.57800 (4.6%)	0.87145	0.58849	0.84461		
0.4	1.10090 (3.2%)	1.52315	1.09490	1.40592		
0.6	2.19492 (2.0%)	2.46494	2.09590	2.27110		
0.8	6.01273 (1.9%)	6.49710	5.30526	4.83403		
			Queue 6			
0.2	0.57666 (4.8%)	0.86720	0.58730	0.84350		
0.4	1.08827 (2.8%)	1.51794	1.08710	1.39861		
0.6	2.14972 (4.0%)	2.42195	2.05411	2.24614		
0.8	5.89917 (2.0%)	6.27625	5.04889	4.74622		
			Queue 8			
0.2	0.57094 (3.5%)	0.86256	0.58610	0.84258		
0.4	1.08791 (1.5%)	1.51251	1.07929	1.39076		
0.6	2.08976 (2.7%)	2.36025	2.01232	2.20812		
0.8	5.71441 (2.8%)	6.08167	4.79253	4.58062		
			Queue 10			
0.2	0.56992 (5.7%)	0.85998	0.58491	0.84186		
0.4	1.06272 (2.3%)	1.50039	1.07149	1.38172		
0.6	2.02811 (1.8%)	2.30015	1.97053	2.15633		
0.8	5.47500 (3.0%)	5.87009	4.53617	4.32833		

[a]Based on Ibe-Chen formula
Note: confidence interval 95%.

Table 3 Symmetrical $M^{[x]}/M/1$ multi-queue with constant batch size ($E[K] = 4$)

ρ	Simulation	Our model	Kimura[16]	Kuehn[14]
	$E[W_i]$	$E[W_i]$	$E[W_i]$	$E[W_i]$
0.1	2.98440 (4.4%)	2.93299	2.75884	2.92171
0.3	4.17897 (4.6%)	4.12438	3.92423	4.01368
0.5	6.47204 (7.7%)	6.36961	6.01890	5.95610
0.7	12.47658 (5.3%)	12.17397	11.25207	10.84315
0.9	62.17523 (4.8%)	62.30619	55.57403	54.07419

Note: confidence interval 90%.

over token-passing LANs, one of the most important applications in LAN communications. To simplify the description, we consider the transfer of text and image files over token-passing LANs. Generalization of the model to include other file types is straightforward, as can be seen from the following discussions.

A generic file transfer system configuration is depicted in *Figure 2*. We consider three system configurations:

1 *Single Text Server:* the single text (file) server configuration consists of a number of workstations sending text files to and retrieving text files from a single text server. The text requests are transmitted over the token-passing LAN from workstations to the text server, and are processed on a first-come-first-served basis at the text server. The text server may also issue text requests to workstations, i.e. text file transfers between the server and workstations are bidirectional.
2 *Multiple Text Servers:* utilization of the LAN may be

Table 4 Symmetrical $M^{[x]}/M/1$ multi-queue with geometrical batch size distribution ($b = p(1-p), k = 1, 2, \ldots, E[K] = 4, p = 0.25$)

ρ	Simulation	Our model	Kimura[16]	Kuehn[14]
	$E[W_i]$	$E[W_i]$	$E[W_i]$	$E[W_i]$
0.1	5.49609 (9.6%)	5.44695	5.27280	5.43568
0.3	7.23986 (8.1%)	7.40906	7.20890	7.29835
0.5	11.24821 (9.4%)	11.10645	10.75575	10.29694
0.7	19.79063 (8.6%)	20.66454	19.74264	19.3337
0.9	102.85097 (7.7%)	103.21549	96.48329	94.98350

Note: confidence interval 90%.

Table 5 Asymmetrical $M^{[x]}/M/1$ multi-queue with constant batch size ($E[K] = 4$)

ρ	Simulation	Our model	Kimura[16]	Kuehn[14]
	$E[W_i]$	$E[W_i]$	$E[W_i]$	$E[W_i]$
		Queue 1		
0.2	3.66399 (4.6%)	3.62046	3.22770	3.58694
0.4	5.97645 (4.1%)	5.78765	5.34006	5.59593
0.6	13.44173 (5.2%)	11.88435	11.09900	11.02572
0.8	114.57623 (9.8%)	95.45826	82.06984	81.96842
		Queue i ($i = 2, 3, \ldots, 10$)		
0.2	3.40843 (6.3%)	3.42343	3.25039	3.38288
0.4	4.76799 (5.3%)	4.92232	4.70113	4.71978
0.6	8.57251 (4.1%)	7.97516	7.57828	7.38259
0.8	20.27276 (4.0%)	18.02967	16.98395	16.17706

Note: confidence interval 90%.

Table 6 Asymmetrical $M^{[x]}/M/1$ multi-queue with geometrical batch size ($E[K] = 4$)

ρ	Simulation	Our model	Kimura[16]	Kuehn[14]
	$E[W_i]$	$E[W_i]$	$E[W_i]$	$E[W_i]$
		Queue 1		
0.2	7.01404 (11.4%)	6.58099	6.39340	6.54746
0.4	10.63600 (8.7%)	10.11457	9.81721	9.92286
0.6	21.68063 (7.4%)	19.92007	19.03410	19.06145
0.8	162.50624 (8.5%)	151.70850	138.31995	138.21866
		Queue i ($i = 2, 3, \ldots, 10$)		
0.2	6.43616 (10.1%)	6.25965	6.08652	6.219102
0.4	8.81094 (11%)	8.75755	8.53636	8.56501
0.6	15.31900 (13.1%)	13.896218	13.49933	13.30364
0.8	33.57623 (9.2%)	31.01045	29.96472	29.15784

Note: confidence interval 90%.

improved by having two or more text servers, each handling different texts in the system.

3 *Single Text Server and Single Image Server:* as in (1) and (2), we assume that text file transfers between workstations and the text server are bidirectional: a workstation may send text files to and retrieve text files from the text server. Unlike text file transfers, image file transfers are unidirectional, i.e. we assume that all image files are input from an external image input device and are already stored in the image (file) server. This configuration can be generalized to include multiple text and image servers, and to include servers of other file types.

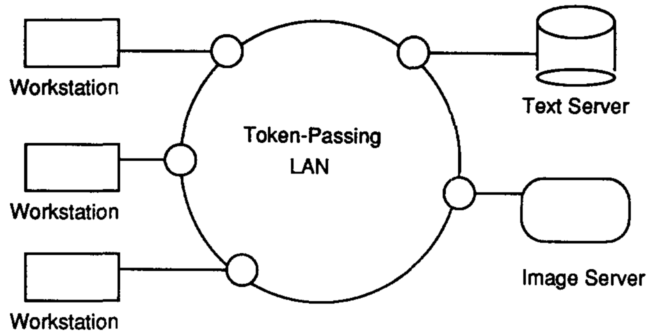

Figure 2 File transfer system configuration

In all system configurations, we assume that the issue of both text file and image file requests obey Poisson distributions. We also assume that a workstation is allowed to have more than one outstanding request at a time, so that the file arrival process at the servers is Poisson distributed. For text file length estimation we adopt the probability distribution given by Welzel[20] (see *Table 7*). Image file is assumed to have a constant length of 512 kbytes (which corresponds to 1024 × 1024 pixel image with a resolution level of 1 byte/pixel and a compression ratio of 2). Furthermore, since the traffic generated by the transmission of control data and query messages is very low compared to the traffic of text and image files, we will not take the traffic of data other than text and image files into consideration.

In the examples below, we are interested in the file sojourn time at each workstation and at the servers. The file sojourn time at a server is measured from the moment the request for a file from a workstation arrives at the server until the completion of the last packet transmission of the requested file. The file sojourn time at a workstation is defined similarly. Let the number of stations (workstations and servers) be N. The mean file sojourn time at the ith station for file of length $K_i = k$ is given by:

$$E[S_i/k] = E[W_i(1)] + (k-1)E[D_i] + h_i, \quad i = 1, 2, \ldots, N \tag{46}$$

Table 7 Probability distribution of text file length

Text file length (Kbyte)	Probability
2	0.51
8	0.20
15	0.07
24	0.06
39	0.05
69	0.05
100	0.06

and the mean file sojourn time at the ith station averaged over all file lengths is given by:

$$E[S_i] = \sum_{k=1}^{\infty} E[S_i/k] b_{ik} = E[W_i(1)] + (E[K_i] - 1)E[D_i] + h_i \quad (47)$$

Numerical results for the three system configurations are depicted in *Figures 3-5*, where the subscripts W, T and I denote workstation, text server and image server, respectively. In all the figures we assume that station walking time is exponentially distributed with mean $a_i = 0.05$, $i = 1, 2, \ldots, N$, that the number of workstations is 9, and that packet length is 1 kbyte with normalized packet transmission (customer service) time equal to 1. Text file lengths for both workstations and text servers follow the distribution provided in *Table 7*. Image length for the image server is 512 packets.

Figure 3 shows the results for the single text server configuration, where the packet arrival rate to the text server is nine times as large as the rates to each workstation, i.e. $\lambda_T = 9\lambda$, $\lambda_W = \lambda$. The results of the multiple (two in this case) text server configuration are shown in *Figure 4*, where text requests from workstations are addressed to the two servers with equal probability. In this case we assumed that $\lambda_T = 4.5\lambda, \lambda_W = \lambda$. Finally, results for the single text server and single image server configuration are given in *Figure 5*, where we have assumed that $\lambda_T = 9\lambda$, $\lambda_I = 9\lambda$ and $\lambda_W = 9\lambda$.

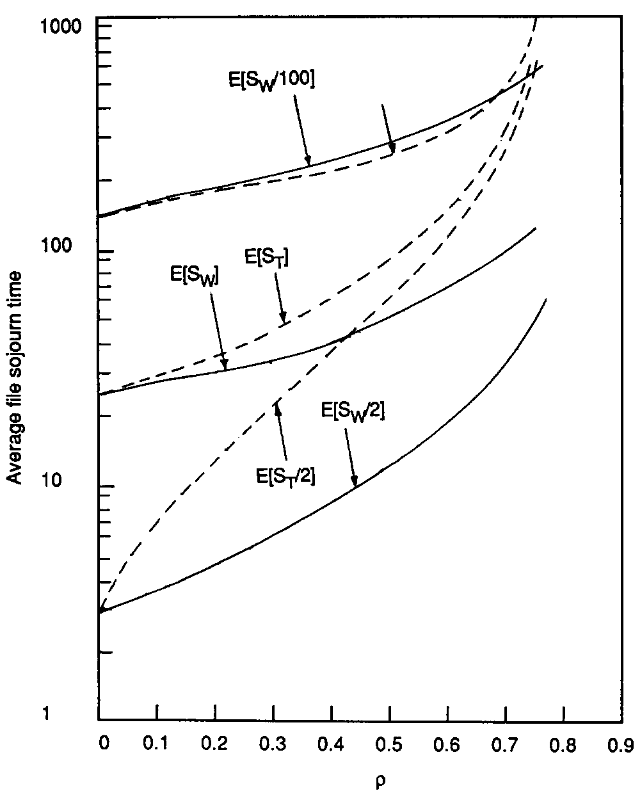

Figure 3 File transfer system configuration 1 — 1 text server

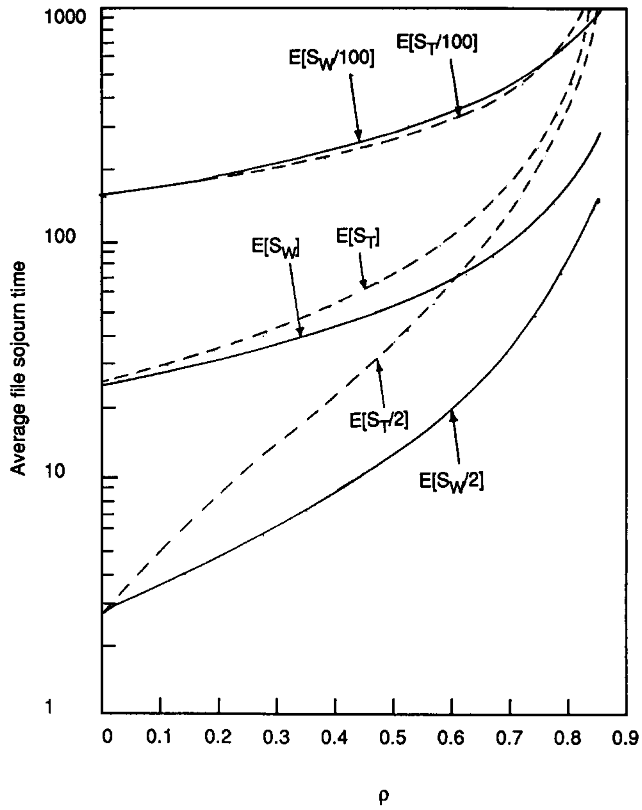

Figure 4 File transfer system configuration 1 — 2 text servers

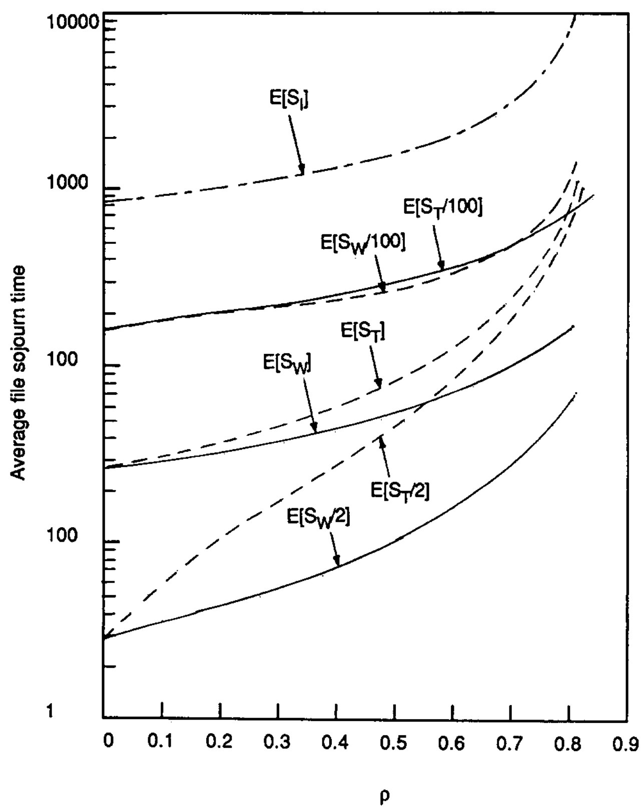

Figure 5 File transfer system configurations 3 — 1 text server, 1 image server

CONCLUSIONS

In this paper we have studied the performance of token-passing systems with either pure Poisson or Poisson batch arrivals. The LST of an approximate customer waiting time distribution has been derived for systems with pure Poisson arrivals, from which the approximate mean and variance of customer waiting time have been obtained. An approximate mean waiting time for systems with Poisson batch arrival was given. The approximations were verified by computer simulations. Both mean waiting time approximation and variance of waiting time approximation showed good agreement with simulation results, especially as low to medium traffic load.

We have also considered some application examples of the analytical technique developed in this paper. In particular, we have proposed three system configurations for file transfer over token-passing LANs. We evaluated the file sojourn time for each system configuration. Generalizations of the file transfer configurations would be (1) to include more file types, and (2) to include both file traffic and interactive traffic in the system model.

ACKNOWLEDGEMENTS

We wish to record our thanks to Dr H Takagi for pointing out an error in the earlier draft of this paper. We also like to thank the referees for their valuable comments.

At the time this research was undertaken, R H Deng and X Zhang were with the Institute of Systems Science, National University of Singapore; K-T Huang is an IBM assignee to the Institute of Systems Science from IBM Watson Research Center, Yorktown Heights, NY, USA

REFERENCES

1 **Bux, W** 'Token-ring local-area networks and their performance', *IEEE Proc.*, Vol 77 No 2 (February 1989) pp 238–256
2 **Hammond, J L and O'Reilly, P J P** *Performance Analysis of Local Computer Networks*, Addison-Wesley, Reading, MA (1986)
3 ANSI/IEEE Standard 802.5 *Local Area Networks: Token Ring Access Method* (1985)
4 ANSI X3.139 *FDDI Token Ring Media Access Control* (1987)
5 **Cooper, R B** 'Queues served in cyclic order: waiting times', *Bell Syst. Tech. J.*, Vol 49 (1970) pp 399–413
6 **Hashida, O and Ohara, K** 'Line accommodation capacity of a communication control unit', *Rev. Elec. Commun. Lab.*, NTT Vol 20 (1972) pp 231–239
7 **Eisenberg, M** 'Two queues with alternating service' *SIAM J. Appl. Math.*, Vol 36 (1979) pp 287–303
8 **Bux, W and Truong, H L** 'Mean delay approximation for cyclic-service queueing system', *Perf. Eval.*, Vol 3 (1983) pp 187–196
9 **Boxma, O J** 'Two symmetric queues with alternating service and switching times', in **Gelenbe, E (ed.)** *Performance '84*, North-Holland, Amsterdam (1984) pp 409–431
10 **Takagi, T** 'Mean message waiting times in symmetric multi-queue systems with cyclic service', *Perf. Eval.*, Vol 5 (November 1985) pp 271–277
11 **Boxma, O J and Meister, B** 'Waiting time approximations for cyclic service systems with switchover times', *Perf. Eval.*, Vol 7 (November 1987) pp 299–308
12 **Ibe, O C and Cheng, X** 'Approximate analysis of asymmetric single-service token-passing systems', *IEEE Trans. Commun.*, Vol 37 No 6 (June 1989) pp 572–577
13 **Takagi, H** *Analysis of Polling Systems*, MIT Press, Cambridge, MA (1986)
14 **Kuehn, P J** 'Multiqueue systems with non-exhaustive cyclic service', *Bell Syst. Tech. J.*, Vol 58 (1979) pp 671–698
15 **Fischer, M J** 'Analysis and design of loop service system via diffusion approximation', *Oper. Res.*, Vol 25 (1977) pp 269–278
16 **Kimura, G and Takahashi, Y** 'Diffusion approximation for a token ring system with nonexhaustive service', *IEEE J. Select. Areas in Commun.*, Vol 4 No 6 (September 1986) pp 794–801
17 **Feller, W** *An Introduction to Probability Theory and Its Applications Vol I (2nd Ed)*, Wiley, New York (1957)
18 **Ross, S M** *Stochastic Processes*, Wiley, New York (1983)
19 **Sauer, C H, MacNair, E A and Kurose, J F** *The Research Queueing Package Version 2, Introduction and Examples*, IBM TJ Watson Research Center, Yorktown Heights, New York (April 1982)
20 **Welzel, T** 'Simulation of a multiple token ring backbone', in **Spaniol, O and Danthine, A (eds)** *High Speed Local Area Networks*, North-Holland, Amsterdam (1987) pp 99–113