Technical Note

A Method of Approximating Variation in Interdeparture Times for GI/G/1 Multiqueue Models with Exhaustive Services

HIDEKI SAKAMOTO†

In this paper we present a method for calculating the coefficients of variation for the times between departures from the individual queues in exhaustive multiqueue models with renewal input for general service-time distributions. This method enables performance evaluation for computer networks and telecommunication systems that use token passing or poling.

1. Introduction

A very powerful method called QNA has been proposed for performance evaluation of computer networks and telecommunication systems [1)]. It is the approximate analysis of traditional general queueing-network models in which customers are probabilistically transferred between a number of $GI/G/m$ queues. The QNA method is based on a parametric-decomposition technique: each queue in queueing-network models is decomposed one by one using the formulas of approximating the average waiting time and the interdeparture-time distribution of $GI/G/m$ queues.

To handle systems that use token passing or poling, however, the scope of the analysis has to be expanded into a general queuing-network model that includes multiqueue models as a specific case. This, in turn, entails a requirement for knowing the average waiting times and the interdeparture-time distributions of $GI/G/1$ multiqueue models. Since average waiting times have already been treated in Ref. 2), in this paper we will focus on deriving a formula to approximate the coefficients of variation of the interdeparture times of the queues in exhaustive multiqueue models with renewal input for unspecified service-time distributions.

In Ref. 3), departure processes from a vacation model are analyzed. However it is not suitable for extending the QNA method, because the arrival distribution is limited to be only exponential. The other approximation method that has been proposed in Ref. 4) is not sufficient, either. Its queueing model really depends on the system architecture to be evaluated.

2. Multiqueue Queuing Models

First let us define the following conditions, based on a multiqueue model such as that shown in **Fig. 1**, in which there are N independent queues, all connected to the same server.

(1) The server cycles through the queues, and if there are any customers waiting at the current queue, the server processes that queue until it is empty of customers (i.e. an "exhaustive" model).

(2) The interarrival times between the customers arriving at the individual queues i ($i = 1, \cdots, N$) are independent of one another, and follow a general distribution having a mean of γ_i^{-1} and a coefficient of distribution of c_{ai}.

(3) The service times spent on the customers arriving at the individual queues i ($i = 1, \cdots, N$) are independent of one another, and follow a general distribution having a mean of s_i and a coefficient of distribution of c_{si}. The utilization rate ρ_i is given by $\gamma_i s_i$.

(4) The times spent by the server to transit (walking time) from queue i to queue $i(\mathrm{mod}N) + 1$ are independent of one another, and follow a general distribution having a mean of u_i and a coefficient of distribution of c_{ui}.

(5) The customer capacity of each queue i is considered to be limitless, and call loss is assumed not to occur.

3. The Coefficient of Variation of Interdeparture Times

In a multiqueue model of this type, during the period in which the server is serving a given queue i, that queue can receive uninterrupted service, being given priority over the other queues, irrespective of the arrival of

† NTT Multimedia Promotion Office

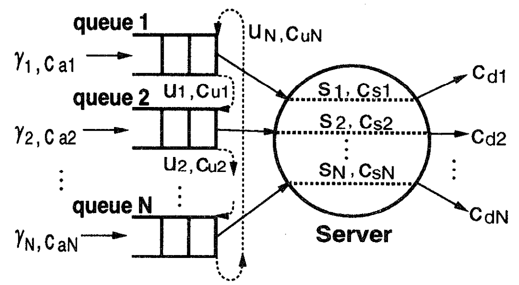


Fig. 1 Multiqueue model.

customers at those other queues. When there are no longer any customers to be serviced in queue i, and the server has entered the period during which it is serving other queues and walking between queues, however, any customers arriving at queue i will have to wait to be served. Let us designate the period during which service to queue i is suspended as the intervisit time I_i, and think of this period I_i as representing service time spent on customers belonging to a hypothetical separate class. This allows us to consider the multiqueue system as a non-preemptive (head-of-the-line) queueing model with the customers arriving at queue i being priority-class (high) customers, and the hypothetical customers serviced during the period I_i being non-priority class (low). Also note that, for multiqueue models, since the service time for queue i and the intervisit time I_i are mutually exclusive, the activity ratio can be determined as $(\rho_{low} \to (1-\rho_{high}))$, if we are considering the saturation state, in which all openings left by the priority class are occupied by the non-priority class. Here ρ_x means the utilization rate of each class.

The following equation has been presented[5] for approximating the coefficient of variation $c_{d\text{-}high}$ of the interdeparture times for the priority class in a non-preemptive queueing system $(GI_{high}, GI_{low}/G_{high}, G_{low}/1)$ with independent inputs from two classes:

$$c_{d\text{-}high}^2 = 1 + (1-\rho_{high}^2)(c_{a\text{-}high}^2 - 1) + \rho_{high}^2\left\{c_{s\text{-}high}^2 + \frac{\rho_{low}s_{low}}{\rho_{high}s_{high}}(1+c_{s\text{-}low}^2) - 1\right\}. \quad (1)$$

If we substitute the parameter for queue i as the high-priority class parameter and that for intervisit time I_i as the low-priority class parameter, and further insert the function for the activity ratio, Eq. (1) becomes:

$$c_{di}^2 = 1 + (1-\rho_i^2)(c_{ai}^2 - 1) + \rho_i^2\left\{c_{si}^2 + \frac{(1-\rho_i)E(I_i)}{\rho_i h_i}(1+c_{Ii}^2) - 1\right\}, \quad (2)$$

where c_{Ii} is set by definition as

$$c_{Ii}^2 = \frac{E(I_i^2)}{\{E(I_i)\}^2} - 1. \quad (3)$$

The primary moment $E(I_i)$ and secondary moment $E(I_i^2)$ of the intervisit time were considered in Ref. 2), where they were given as the equations below Eq. (5) is an approximation equation.

$$E(I_i) = \frac{1-\rho_i}{1-\rho_o}c_o \quad (4)$$

$$E(I_i^2) = \frac{1}{(1-\rho_o)}\sum_{j=1}^{N}(1-\rho_i)s_i^2c_{si}^2 + \frac{1}{(1-\rho_o)^2}\sum_{k=1}^{N}\sum_{j=1,}^{N}\rho_j s_j c_{aj}^2(1+c_{sj}^2)u_k + E(I_i)^2 \quad (5)$$

In the above, $\rho_o = \sum_{k=1}^{N}\rho_k$ and $c_o = \sum_{k=1}^{N}u_k$.

4. Results of Evaluation

In **Fig. 2** we show the coefficients of variation of the interdeparture times for queue i obtained using Eq. (2), and the corresponding average waiting times $E(W_i)$ given in Ref. 2), both under the assumption of symmetrical traffic (i.e. traffic in which γ_i, s_i, and u_i, as well as c_{ai}, c_{si}, c_{ui} are all the same for every queue i). We set the number (N) of queues connected to the multiqueue system at 2, and the server's mean walking time u_i at 20% of the average service time s_i. In both cases the calculations matched with our simulation results within a 15% margin of error. Besides these two cases, we also have confirmed the effectiveness of this approxmation in some other cases such as $c_{ai} = 0.5/c_{si} = 0.5$ and $c_{ai} = 1.5/c_{si} = 1.5$.

In regular queueing models, the coefficients of variation of the interdeparture times never exceed the coefficients of variation of the interarrival times or service times[1]. In multiqueue models, on the other hand, the servicing of the customer queues is done more and more in bursts as activity ratios increase, so these models are characterized by coefficients of

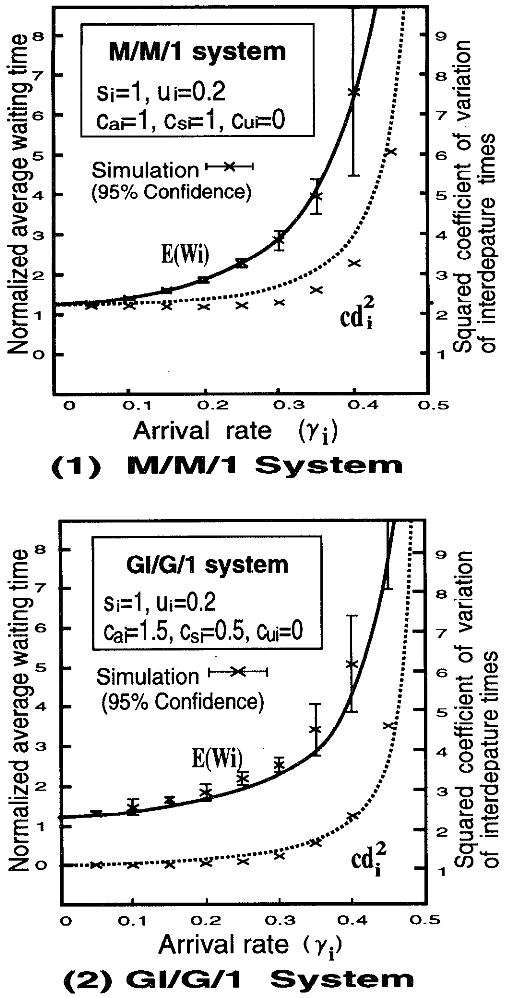


Fig. 2 Average waiting times and coefficients of interdeparture-time variation in a multiqueue model.

interdeparture-time variation that not only surpass the coefficients of variation of interarrival and service times, but actually become several times as large.

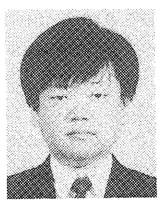
5. Conclusion

In this paper we have presented a method for calculating the coefficients of variation for the times between departures from the individual queues in exhaustive multiqueue models with renewal input for unspecified service-time distributions. Using numerical examples assuming symmetrical traffic, we showed that the calculation method presented matches well with results from a simulation. An approximate analysis of general queueing-network models including multiqueue systems can be conducted by incorporating the method proposed here and the average waiting time of Ref. 2) into the QNA method that was originally proposed for only traditional general queueing-network models.

References

1) Whitt, W.: The Queueing Network Analyzer, *Bell Syst. Tech. J.*, Vol.62, pp.2779–2815 (1983).
2) Kimura, G., and Takahashi, Y.: Traffic Analysis for Exhaustive-service Systems and Gate Token-ring Systems – The Group Renewal-input Case, *Trans. IEICE*, Vol.J71-B, No.2, pp.129–137 (1988).
3) Shimogawa, S. and Takahashi, Y.: Departure Process from a Vacation Model, Abstracts of Fall Meeting of OR Society of Japan, pp.110–111 (1989).
4) Takine, T., Takahashi, Y. and Hasegawa, T.: An Analysis for Interdeparture Process of a Polling System with Single Message Buffer at Each Station, *Trans. IEICE*, Vol.J70-B, No.9, pp.989–998 (1987).
5) Yoshino, H.: An Approximation Method for Queueing Network with Nonpreemptive Priority and its Performance, *Trans. IEICE*, Vol.E73, No.3, pp.386–394 (1990).

(Received November 7, 1996)
(Accepted February 5, 1997)



Hideki Sakamoto was born in 1961. He received his B.E. and M.E. degrees in communication engineering from Osaka University in 1984 and 1986, respectively. He joined NTT Laboratories in 1986 and was engaged in research and development of HDTV high-speed videotex systems and video-on-demand systems. He is a Manager in the NTT Multimedia Promotion Office, a member of the IEEE Communications Society, the ACM, and the Institute of Electronics, Information and Communication Engineers of Japan.