

ON THE SOJOURN TIME DISTRIBUTION IN CYCLIC QUEUEING SYSTEMS WITH A LiPS STATION

Takeshi Kawashima
The National Defense Academy

(Received April 10, 1986; Final March 10, 1987)

Abstract We consider a cyclic queueing network with two service stations, nodes 1 and 2. Node 1 has s servers and the FCFS discipline. Node 2 processes jobs under the m -limited processor sharing (m -LiPS) discipline, which is a generalization of the processor sharing (PS) discipline. We consider the joint distribution of successive sojourn times of a job at the two nodes, and propose a simple method to describe the Laplace-Stieltjes transform (LST) and the moments of these variables by their conditional marginal distributions. Further, we deduce linear equations giving the moments and the LST for an m -LiPS node. For some cases, the correlation coefficients and the coefficients of variation of the total sojourn times are calculated numerically.

1. Introduction

The processor sharing discipline (PS) is one of the interesting disciplines in computer engineering. In a service station with this discipline, the server's capacity is shared equally by all jobs present in the station. It may be impossible, however, to be shared by too many jobs because of the limitation of memory size and so forth. Yamazaki and Sakasegawa [16] proposed the m -limited processor sharing (m -LiPS) discipline, where m is a positive integer. In an m -LiPS station, the server's capacity is shared equally by at most m jobs present in the station, and the rest of the jobs, if any, have to wait for service in the queue in the arrival order. Clearly, the FCFS discipline at a single server station is described as the 1-LiPS discipline. If m is not less than the attainable maximum queue length, the m -LiPS discipline coincides with the PS discipline. In a computer system with m pseudo processors called "init" or "initiator", its job processing mechanism is modeled as an m -LiPS station (cf. Ishiguro [7]).

In this paper, we consider a cyclic queueing network with two service stations, nodes 1 and 2 (see Fig. 1). Node 1 has s servers and the FCFS discipline. Node 2 processes jobs under the m -LiPS discipline. The total number of jobs in the system is N . Service requirements of each job at nodes 1 and 2 are independent exponentially distributed random variables with rate μ_1 and μ_2 , respectively. Our purpose is to analyze the joint distribution of successive sojourn times at nodes 1 and 2 experienced by one job. Generally, computer users are interested in the characteristics of sojourn times or their own response times. It is well known that the equilibrium state probabilities for this system are expressed as product form solutions. The equilibrium state probabilities, however, do not give sufficient information for the sojourn times by themselves.

Here, we show that the joint distribution can be described by combining the marginal distributions, invoking the reversible Markov property. Further, we derive linear equations giving the Laplace-Stieltjes transform (LST) and the moments of the sojourn time at the m -LiPS node. Numerical results can be obtained easily from our result. For some cases, the coefficient correlations and the coefficients of variation of the total sojourn times are calculated numerically.

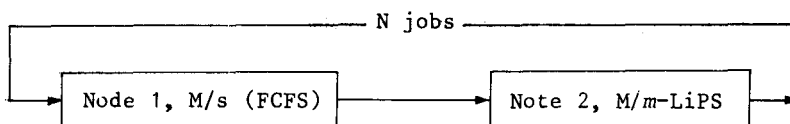


Fig. 1. Cyclic Queueing System

2. Preliminary and Related Results

Let $q_i(t)$ ($-\infty < t < \infty$) be the number of jobs present at node i ($i=1,2$) at time t . Since $q_1(t) + q_2(t) = N$ holds always, the system's behaviour can be described by $q_2(t)$, which is used mainly. We denote $q_2(t)$ by $q(t)$ simply. According to the description of our system, $q(t)$ is an ergodic Markov process. Let P be a probability measure under which $q(t)$ is in equilibrium. Then, we have the following for any $t \in (-\infty, \infty)$.

$$(2.1) \quad P(q(t)=q) = G(N)^{-1} \mu_2^{-q} \prod_{j=1}^{N-q} (\min(j,s)\mu_1)^{-1} \quad (q=0,1,\dots,N),$$

where $G(N)$ is a normalizing constant. We denote the right hand side of (2.1) by $f_N(q)$ for convenience.

With respect to P , two sequences of arrival times of jobs at nodes 1 and 2 are stationary point processes, from each of which we can define Palm probability measures P_1 and P_2 , respectively (cf. Franken, et al., p.23 [6]). $P_1(P_2)$ is a measure which is strictly stationary for the sequence of arrival times at node 1 (node 2). P_1 can be also considered as a conditionally probability measure of P under the condition that a job arrives at node 1 at time $t=0$, and P_2 be that under the condition that a job arrives at node 2 at $t=0$. We tag the job arriving at node 1 at $t=0$ under P_1 and the job arriving at node 2 under P_2 , respectively. Let W_1, W_2 be consecutive sojourn times of the tagged jobs at node 1 and 2. Then, we have the following (cf. Kawashima [9]), which is proved from the ergodicity of $q(t)$.

$$(2.2) \quad P_1(W_1 \leq x, W_2 \leq y) = P_2(W_1 \leq x, W_2 \leq y).$$

Let us consider the state of the other $N-1$ jobs at $t=0$. It is known that, with respect to both P_1 and P_2 , the state distribution is identical to the equilibrium distribution of the system with $N-1$ jobs circulating (cf. Kawashima [8] or Lavenberg and Reiser [11]). We have

$$(2.3) \quad P_1(q(0^+) = q) = P_2(q(0^+) = q+1) = f_{N-1}(q) \quad (q=0, 1, \dots, N-1).$$

In the following section, we deduce an expression of the joint distribution of W_1 and W_2 with respect to P_2 , which holds true with respect to P_1 .

For a system where nodes 1 and 2 are single-server nodes with the FCFS discipline, Chow [4] obtained the joint distribution of W_1 and W_2 with respect to P_1 , which has so called a product form. Boxma and Donk [1] also obtained it with respect to P_2 . For a system with many-server nodes, a similar expression is obtained by Kawashima and Torigoe [10]. Further, Schassberger and Daduna [15] considered a cyclic closed system consisting of a series of single-server nodes, a series of infinite-server nodes and two many-servers nodes. The two series are connected by the two many-servers nodes at each ends. They showed that the joint distribution of the successive sojourn times has a product form, considering the LST's. Their result covers Schassberger and Daduna [14], Boxma and Donk [1]. For systems with PS or m -LiPS nodes, there has been no result concerning the joint distribution of the sojourn times, in our knowledge. However, some results for the marginal sojourn time distribution in a closed system are known. Mitra [12] considered a system where node 1 has infinite servers and node 2 has the PS discipline. He obtained the linear differential equations giving the distribution of the sojourn time at node 2, and some numerical results. He also showed that the eigenvalues of the coefficient matrix are all real and nonpositive. Coffman, et. al. [5] obtained

the LST and the some moments of the sojourn time in a PS node with a Poisson arrival process.

3. Time Reversible Process

A stationary Markov process $X(t)$ ($-\infty < t < \infty$) is said to be time reversible if its time reversed process $X(-t)$ is stochastically identical to $X(t)$. An equivalent condition to be reversible for $X(t)$ is as follows:

$$(3.1) \quad P(s)R(s, s') = P(s')R(s', s) \text{ for all } s, s' \in S,$$

where S , $P(s)$ and $R(s, s')$ are the state space, the equilibrium distribution and the transition rate of $X(t)$, respectively (cf. Reich [13]). For the process $q(t)$, we have

$$R(q, q') = \begin{cases} \min(N-q, s)\mu_1 & \text{if } q' = q+1, q \geq 0, \\ \mu_2 & \text{if } q' = q-1, q \geq 1, \\ 0 & \text{otherwise.} \end{cases}$$

From these and (2.1), we can easily verify that (3.1) holds. Thus, we have

Lemma 1. $q(t)$ is a time reversible Markov process.

Note that $q_1(t)$ is also a time reversible process. The following lemma is essential to deduce the joint distribution of W_1 and W_2 . The proof is given in Appendix 1. The lemma means that the sojourn time of a job leaving q jobs at node 1 is stochastically equal to that of a job finding q jobs at its arrival instant.

Lemma 2. $P_2(W_1 \leq x | q_1(0+) = q) = P_1(W_1 \leq x | q_1(0+) = q+1)$.

4. Joint Distribution of Sojourn Times

We consider the joint distribution of W_1 and W_2 with respect to P_2 . Since W_1 depends only on $q(-t)$ ($t \geq 0$) and W_2 on $q(t)$ ($t \geq 0$), they are conditionally independent under the condition that $q(0+) = q+1$ ($q=0, 1, \dots, N-1$). From (2.3), we have

$$\begin{aligned} & E_2(\exp(-z_1 W_1 - z_2 W_2)) \\ &= \sum_{q=0}^{N-1} f_{N-1}(q) E_2(\exp(-z_1 W_1 - z_2 W_2) | q(0+) = q+1) \\ &= \sum_{q=0}^{N-1} f_{N-1}(q) E_2(\exp(-z_1 W_1) | q(0+) = q+1) E_2(\exp(-z_2 W_2) | q(0+) = q+1), \end{aligned}$$

where E_i is the expectation with respect to P_i ($i=1,2$). Invoking Lemma 2, we have the next.

Theorem.

$$E_2(\exp(-z_1W_1 - z_2W_2)) = \sum_{q=0}^{N-1} f_{N-1}(q) E_1(\exp(-z_1W_1) | q_1(0-) = N-q-1) E_2(\exp(-z_2W_2) | q(0+) = q+1).$$

Now, observing that node 1 is FCFS, we have

$$(4.1) \quad E_1(\exp(-z_1W_1) | q_1(0-) = N-q-1) = \left(\frac{s\mu_1}{z_1 + s\mu_1} \right)^{\max(N-q-s, 0)} \cdot \frac{\mu_1}{z_1 + \mu_1}.$$

In order to determine $E_2(\exp(-z_2W_2) | q(0+) = q+1)$, consider the behaviour of the tagged job with respect to P_2 . If $q(0+) = q+1$ is not less than $m+1$, the tagged job has to wait for $q-m+1$ service completions to start its service because of m -LiPS discipline. If $q+1 \leq m$, then the tagged job's service starts at $t=0$. Further, the time interval between its arrival and the first service completion at either node 1 or node 2 is exponentially distributed with parameter $\mu(q+1)$ defined as follows:

$$\mu(q+1) = \min(N-q-1, s)\mu_1 + \mu_2.$$

Let $X_{q,i}$ be a random variable whose distribution is equal to that of the residual sojourn time experienced by the tagged job under the condition that there exist q jobs at node 2 (including the tagged job), and that the tagged job occupies the i -th position in the queue. If it is in service, we denote $i=0$. Let $g_{q,i}(z)$ ($z>0$) be LST of $X_{q,i}$ (the expectation of $\exp(-zX_{q,i})$). We can easily obtain the followings.

$$(4.2) \quad g_{q,0}(z) = \frac{1}{z + \mu(q)} \left\{ \min(N-q, s)\mu_1 g_{q+1,0}(z) + \frac{\min(q, m) - 1}{\min(q, m)} \mu_2 g_{q-1,0}(z) + \frac{1}{\min(q, m)} \mu_2 \right\} \quad (q=1, 2, \dots, N),$$

$$g_{q,i}(z) = \frac{1}{z + \mu(q)} \left\{ \min(N-q, s)\mu_1 g_{q+1,i}(z) + \mu_2 g_{q-1,i-1}(z) \right\} \quad (q=m+1, \dots, N; i=1, 2, \dots, q-m).$$

Clearly, we have

$$(4.3) \quad E_2(\exp(-z_2W_2) | q(0+) = q+1) = g_{q+1, \max(q+1-m, 0)}(z_2).$$

Equations (4.2) are linear equations for $g_{q,i}(z)$'s. Solving this, we can have an expression of LST transforms for W_2 . It does not have, however, a simple expression even for small N and $m > 1$. Differentiating (4.2) or considering $X_{q,i}$'s directly, we can obtain linear equations giving the first and second moments. Let $x_{q,i}$ and $y_{q,i}$ be $E(X_{q,i})$ and $E(X_{q,i}^2)$, respectively, then they satisfy the next equations.

$$x_{q,0} = \frac{1}{\mu(q)} \left\{ \min(N-q,s)\mu_1 x_{q+1,0} + \frac{\min(q,m)-1}{\min(q,m)} \mu_2 x_{q-1,0} + 1 \right\} \quad (q=1,2,\dots,N),$$

$$x_{q,i} = \frac{1}{\mu(q)} \left\{ \min(N-q,s)\mu_1 x_{q+1,i} + \mu_2 x_{q-1,i-1} + 1 \right\} \quad (q=m+1,\dots,N; i=1,2,\dots,q-m),$$

$$y_{q,0} = \frac{2}{\mu(q)^2} + \frac{\min(N-q,s)\mu_1}{\mu(q)} \left(y_{q+1,0} + \frac{2}{\mu(q)} x_{q+1,0} \right) + \frac{(\min(q,m)-1)\mu_2}{\mu(q)\min(q,m)} \left(y_{q-1,0} + \frac{2}{\mu(q)} x_{q-1,0} \right) \quad (q=1,2,\dots,N),$$

$$y_{q,i} = \frac{2}{\mu(q)^2} + \frac{\min(N-q,s)\mu_1}{\mu(q)} \left(y_{q+1,i} + \frac{2}{\mu(q)} x_{q+1,i} \right) + \frac{\mu_2}{\mu(q)} \left(y_{q-1,i-1} + \frac{2}{\mu(q)} x_{q-1,i-1} \right) \quad (q=m+1,\dots,N; i=1,2,\dots,q-m).$$

Using (4.1), (4.3) and the above equations, we can calculate the correlation coefficients and the coefficients of variation for W_1+W_2 . After the calculations for many cases, it can be seen that the coefficient correlations seem to be always nonpositive. For a typical example, the results for the case $N=10$ and $s=2$ is shown in Fig. 2. Since (2.1) and (2.3) do not depend on m , the expectation of W_2 and the distribution of W_1 also do not depend on m . When μ_2/μ_1 is small, W_1 is close to zero and the variance of W_1+W_2 is almost equal to that of W_2 . From Fig. 2(b), we see that their variances increase with m . The PS discipline gives large variances.

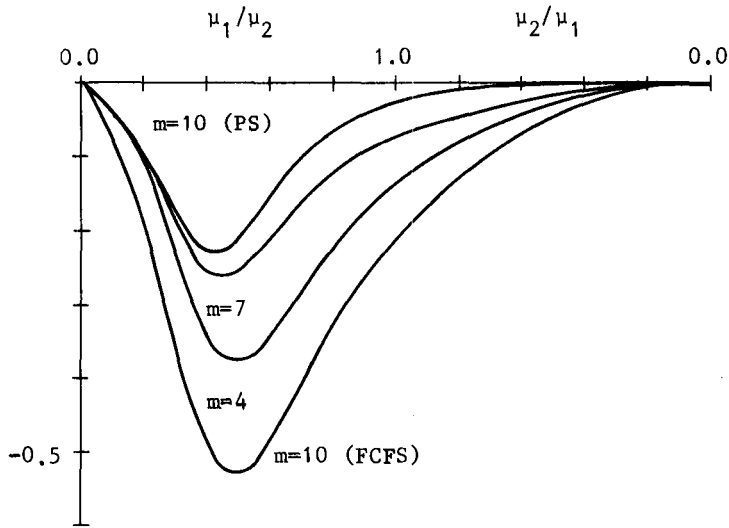


Fig. 2(a) Coefficient Correlation of W_1 and W_2 for $N=10, s=2$

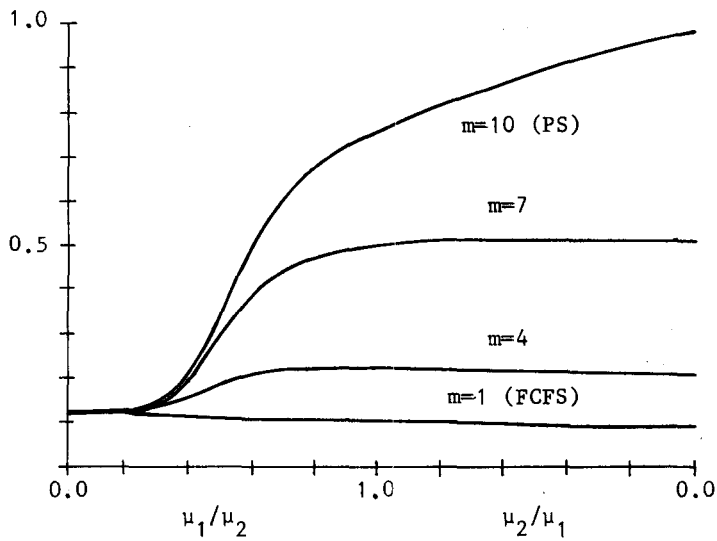


Fig. 2(b) Coefficient of Variation of $W_1 + W_2$ for $N=10, s=2$

5. Concluding Remark

We derived the linear equations for LST's of the sojourn times and some of their moments. It is also possible to obtain linear differential equations which the conditional distributions of W_2 satisfy as in Mitra [12]. From these equations, we can obtain the shape of the sojourn time distributions numerically. However, it requires more complicated computations.

Acknowledgement

The author would like to thank Professors G. Yamazaki, H. Sakasegawa, M. Miyazawa, and Mr. I. Kino for their valuable discussions and encouragements. He also thanks the referees for their valuable comments.

Appendix 1. Burke [3] proved Lemma 2 for $M/M/s$ system in two different methods. We apply one of his method to our system.

Proof Lemma 2. Since $q_1(t)$ is a Markov process, the behaviour of $q_1(t)$ ($t > 0$) under the condition $q_1(0+) = q+1$ is stochastically identical with respect to P and P_1 . We write this as follows;

$$(A.1) \quad P(q_1(t)(t > 0) | q_1(0+) = q+1) = P_1(q_1(t)(t > 0) | q_1(0+) = q+1) \\ (q=0, 1, \dots, N-1).$$

Further, we have from Lemma 1

$$(A.2) \quad P(q_1(t)(t > 0) | q_1(0+) = q+1) = P(q_1(-t)(t > 0) | q_1(0-) = q+1) \\ (q=0, 1, \dots, N-1).$$

Let A be the event that an arrival occurs at node 2 at $t=0$. From the definition of P_2 , we have

$$P_2(q_1(-t)(t > 0) | q_1(0-) = q+1) \\ = P(q_1(-t)(t > 0), q_1(0-) = q+1 | A) / P(q_1(0-) = q+1 | A) \\ = P(q_1(-t)(t > 0) | q_1(0-) = q+1) P(q_1(0-) = q+1 | A) / P(q_1(0-) = q+1 | A) \\ = P(q_1(-t)(t > 0) | q_1(0-) = q+1),$$

because $q_1(-t)$, the time reversed process of $q_1(t)$, is also a Markov process

with respect to P . Thus, we have

$$P_1(q_1(t)(t>0) | q_1(0+)=q+1) = P_2(q_1(-t)(t>0) | q_1(0-)=q+1).$$

Under P_2 , the event $\{q_1(0-)=q+1\}$ implies $\{q_1(0+)=q\}$ and vice versa, we obtain the following.

$$(A.4) \quad P_1(q_1(t)(t>0) | q_1(0+)=q+1) = P_2(q_1(-t)(t>0) | q_1(0+)=q).$$

Let $\xi(t)(t>0)$ be an arbitrary sample path of $q_1(t)$ such that $\xi(0+)=q+1$. We will show the next.

$$(A.5) \quad P_2(W_1=x_i | q_1(-t)=\xi(t)(t>0)) = P_1(W_1=x_i | q_1(t)=\xi(t)(t>0)),$$

where $\{x_1, x_2, \dots\}$ and $\{y_1, y_2, \dots\}$ be the sequences of downward and upward jump instants of $\xi(t)$, respectively. Let m_i be the number of the jobs in service at x_i-0 . Since the discipline at node 1 is FCFS and the service times are exponentially distributed,

$$(A.6) \quad P_1(W_1=x_i | q_1(t)=\xi(t)(t>0)) = \begin{cases} 0, & i \leq \max(q+1-s, 0) \\ m_i^{-1} \prod_{j=\max(q+1-s, 0)+1}^{i-1} (m_j-1)/m_j, & i > \max(q+1-s, 0). \end{cases}$$

On the other hand, consider the case where $q_1(-t)=\xi(t)(t>0)$ with respect to P_2 . In this case $-x_i$'s become arrival instants and $-y_i$'s become service completion instants. Let $-z_i$ be the service starting instant of the job arriving at $-x_i$ for $i > \max(q+1-s, 0)$. Then, we have

$$-z_i = \begin{cases} -x_i & \text{if } q_1(-x_i+0) \leq s, \\ -y_{j(i)} & \text{if } q_1(-x_i+0) \geq s+1, \end{cases}$$

where $-y_{j(i)}$ is the $q_1(-x_i+0)-s$ th point of $-y_j$'s after $-x_i$. If $z_i \neq x_i$, then the numbers of the jobs at $-x_i+0$ and at $-z_i-0$ are greater than s , that is, the numbers of the jobs in service at $-x_i+0$ and at $-z_i+0$ are both s . Thus the number of the jobs in service at $-z_i+0$ is m_i for all i . Now, the joint distribution of elapsed service times of the jobs in service at arbitrary time t are symmetric given $q_1(u)(u>t)$ (see Appendix 2). So, the probability that the tagged job's service start at $-z_i$ under the condition that it is in service at $-z_i+0$ is $1/m_i$. The event $\{W_1=x_i\}$ implies {the tagged job's service starts at $-z_i$ } and vice versa. Thus, the left hand side of (A.5) is equal to the right hand side of (A.6). We obtain (A.5). Integrating this by the probability density (A.4), we obtain Lemma 2. Q.E.D

Appendix 2. Consider $G/M/s$ with FCFS discipline, where G is not necessary GI. Suppose that there are m jobs in service at an arbitrary time t . We label the numbers $1, 2, \dots, m$ to the m jobs at random. Let X_i, Y_i be the elapsed and residual service times of the i -th job, respectively. Then, it holds that

$$P(X_1 \leq x_1, X_2 \leq x_2, \dots, X_m \leq x_m | Y_1 = y_1, Y_2 = y_2, \dots, Y_m = y_m)$$

is symmetric for x_1, x_2, \dots, x_m , and does not depend on y_1, y_2, \dots, y_m , where P is an arbitrary probability measure which is well defined for our system.

Proof: Since service times are exponentially distributed and our labeling does not depend on the others, we have

$$P(Y_1 \leq y_1, \dots, Y_m \leq y_m | X_1, \dots, X_m) = P(Y_1 \leq y_1) \cdots P(Y_m \leq y_m),$$

$$P(X_1 \leq x_1, \dots, X_m \leq x_m) \text{ is symmetric for } x_1, \dots, x_m.$$

Then we have

$$\begin{aligned} & P((X_1, \dots, X_m) \in A, (Y_1, \dots, Y_m) \in B) \\ &= \int_A P((Y_1, \dots, Y_m) \in B | X_1, \dots, X_m) dP(X_1, \dots, X_m) \\ &= P((Y_1, \dots, Y_m) \in B) \int_A dP(X_1, \dots, X_m) \\ &= P((Y_1, \dots, Y_m) \in B) P((X_1, \dots, X_m) \in A), \end{aligned}$$

where A and B are arbitrary m -dimensional Borel sets. Thus, the random vector (X_1, \dots, X_m) is independent of Y_1, \dots, Y_m , and Y_i 's are independent each other.

Q.E.D.

The result in Appendix 2 means that the behaviour of these m jobs at the past are stochastically uniform and independent of the future. The next follows, which is used in the proof of Lemma 2.

$$P(X_i = 0 \mid \text{only one of } X_i \text{'s is } 0, Y_1 = y_1, Y_2 = y_2, \dots, Y_m = y_m) = 1/m.$$

References

- [1] Boxma, O.J., Donk, P. (1982): On Response-time and Cycle Time Distribution in cyclic queues, *Perf. Eval.*, Vol.2, 181-194.
- [2] Boxma, O.J., F.P. Kelly, A.G. Konheim (1984): The Product Form Distribution in Cyclic Exponential Queues, *J. ACM.*, 31, 128-133.

- [3] Burke, P.J. (1968): The Output Process of a Stationary M/M/s Queueing System, *A.M.S.*, 39, 1144-1152.
- [4] Chow, W.M. (1980): The Cycletime Distribution of Exponential Queues, *J. ACM.*, 27, 281-286.
- [5] Coffman, E.G., R.R. Muntz, H. Trotter (1970): Waiting Time Distributions for Processor-sharing Systems, *J. ACM.*, 17, 123-130.
- [6] Franken, P., D. König, U. Arndt, V. Schmitdt: Queues and Point Processes, J. Wiley, Chichester 1982.
- [7] Ishiguro, M. (1980): Running Time Delays in Processor-sharing System, *J. Inf. Processing*, 3, 38-44.
- [8] Kawashima, T. (1978): Turnaround time equations in queueing networks, *J. Opn. Res. Soc. Japan*, 21, 477-485.
- [9] Kawashima, T. (1982): A property of Two Palm Measures in Queueing Networks and its Applications, *J. Opn. Res. Soc. Japan.*, 25, 16-28.
- [10] Kawashima, T., N. Torigoe (1983): The Cycle Time Distribution in a Central Server Queueing System with Multi-server Stations, *Memo. Def. Acad.*, 23, 155-160.
- [11] Lavenberg, S.S., M. Reiser (1980): Stationary State Probabilities at Arrival Instants for Closed Queueing Networks with Multiple Types of Nodes, *J. Appl. Prob.*, 17, 1048-1061.
- [12] Mitra, D. (1981): Waiting time Distributions from Closed Queueing Network Models of Shared-processor Systems, *Performance* 81, 113-131, North-Holland, Amsterdam.
- [13] Reich, E. (1959): Waiting Times when Queues are in Tandem, *A.M.S.*, 28, 76-773.
- [14] Schassberger, R., H. Daduna (1983): The Time for a Round Trip in a Cycle of Exponential queues, *J. ACM.*, 30, 146-150.
- [15] Schassberger, R., H. Daduna (1984): Sojourn Times in Queueing Network with Multiserver Nodes, Preprint.
- [16] Yamazaki, G., H. Sakasegawa (1984): Limited Processor Sharing Queues (in Japanese), *Kokyuroku* 519, 99-113. Institute of Math, Kyoto Univ., Kyoto.

Takeshi KAWASHIMA: Defence Academy,
Hashirimizu, 1-10-20
Yokosuka, 239, Japan