ZASTOSOWANIA MATEMATYKI
APPLICATIONES MATHEMATICAE
XV, 4 (1977)

D. STOYAN (Freiberg, G.D.R.)

INEQUALITIES FOR MULTISERVER QUEUES AND A TANDEM QUEUE

1. Introduction. In this paper an exponential bound for the stationary waiting time distribution in a wide range of $GI/G/s$ queues is presented. It is based on the results of Brumelle [3] and exponential bounds for single-server queues. Additionally, for a tandem queue $GI/G/1 \to \ldots \to G/1$, upper and lower bounds for the mean stationary waiting times at the service stations are given. They follow from bounds of Suzuki and Maruta [11] and an extremality property of the regular input in single-server queues.

2. Exponential bound for the waiting time distribution in $GI/G/s$. Exponential bounds for the waiting time distribution function (d. f.) in single-server queues have been known since some years; they were established by Kingman [4] (see also [2] and [7]). In this paper, a similar bound for multiserver queues is proved, using the fundamental results of Brumelle [3] and the monotonicity properties of $GI/G/1$. Let $\overset{(2)}{\leqslant}$ denote the following semi-ordering relation for d. f.'s:

$$F_1 \overset{(2)}{\leqslant} F_2 \Leftrightarrow \int_t^\infty [1-F_1(x)]\,dx \leqslant \int_t^\infty [1-F_2(x)]\,dx \quad \text{for all real } t.$$

In the following, a d. f. F with mean m will be called a *d. f. of type* E_2 if

$$F \overset{(2)}{\leqslant} 1-\exp\left[-\frac{1}{m}\right].$$

Every d. f. of type NBUE is of type E_2 (see [1]). Criteria for E_2-distributions are given by Rolski [6].

THEOREM 1. *Let A and B be the interarrival and service time d. f.'s with means λ^{-1} and μ^{-1}, respectively, in a $GI/G/s$ queue. If A and B are both of type E_2, then for the stationary waiting time d. f. W the inequality*

$$(1) \qquad 1-W(t) = \overline{W}(t) \leqslant \exp\left[-\left(\mu-\frac{\lambda}{s}\right)t\right] \quad (0 \leqslant t < \infty)$$

is true.

Proof. Let $W^*(t)$ be the stationary waiting time d. f. in the $GI/G/1$ queue with the interarrival time d. f. A and the service time d. f. B^*,

$$B^*(t) = \frac{1}{s} B(t) + \frac{s-1}{s} \qquad (0 \leqslant t < \infty).$$

Brumelle [3] proved the inequality

$$W^*(t) \leqslant W(t) \qquad (0 \leqslant t < \infty). \tag{2}$$

We use the exponential bounds for single-server queues from [2] to bound W^* and, consequently, W. Let K^* be the d. f. such that

$$K^*(t) = \int_{\max\{0,-t\}}^{\infty} B^*(t+x)\, dA(x) \qquad (-\infty < t < \infty).$$

For W^* the exponential bound (see [4])

$$1 - W^*(t) = \overline{W}^*(t) \leqslant \exp[-\theta^* t] \tag{3}$$

is true, where θ^* is the solution of the equation

$$\int_{-\infty}^{\infty} \exp[\theta^* x]\, dK^*(x) = 1. \tag{4}$$

(Since A and B are of type E_2, there exists a positive number θ such that $\int_{-\infty}^{\infty} \exp[\theta x]\, dK^*(x) < \infty$. Consequently, under the assumptions of Theorem 1, equation (4) has a positive solution.)

If K is a further d. f. with

$$K^* \overset{(2)}{\leqslant} K,$$

then, for the solution θ of the equation

$$\int_{-\infty}^{\infty} \exp[\theta x]\, dK(x) = 1, \tag{5}$$

the inequality $\theta \geqslant \theta^*$ holds [2]. Using this result, we are able to obtain bounds on the solution of (4). Since in case of equal means the relation

$$A_1 \overset{(2)}{\leqslant} A_2,\ B_1 \overset{(2)}{\leqslant} B_2 \Rightarrow K_1 \overset{(2)}{\leqslant} K_2$$

holds (see [8]) and since A and B are of type E_2, we have

$$K^* \overset{(2)}{\leqslant} K,$$

$$K(t) = \begin{cases} \dfrac{1}{s}\left[1 - \dfrac{\lambda}{\lambda+\mu} \exp[-\mu t]\right] + \dfrac{s-1}{s} & (t \geqslant 0), \\[2ex] \dfrac{1}{s} \dfrac{\mu}{\lambda+\mu} \exp[\lambda t] + \dfrac{s-1}{s} \exp[\lambda t] & (t < 0). \end{cases}$$

The solution of equation (5), which can be written in the form

$$\frac{\lambda\mu}{s(\lambda+\mu)(\mu-\theta)}+\frac{\lambda}{\lambda+\theta}-\frac{\lambda^2}{s(\lambda+\mu)(\lambda+\theta)}=1,$$

is $\theta = \mu-\lambda/s$. Using $\theta \geqslant \theta^*$, (3), and (2), we obtain (1). Of course, the accuracy of (1) is not good; the reader should interpret it as the first approach to good exponential bounds. The author presumes, under the assumption of Theorem 1, the bound

$$\overline{W}(t) \leqslant \exp[-(s\mu-\lambda)t]$$

to be true.

3. Bounds for the mean waiting time in a stationary tandem queue. In [9] it was shown that in a single-server queue idle at time $t = 0$ with interarrival times $a_1, a_2, \ldots$ and service times $\beta_1, \beta_2, \ldots$ the mean waiting time $m_{W,n}$ of the n-th customer (the first arrives at $t = 0$) is greater than the mean waiting time of the n-th customer in the single-server queue with interarrival times $Ea_1, Ea_2, \ldots$ and service times $\beta_1, \beta_2, \ldots$ If the interarrival and service times form metrically transitive sequences, using the monotonicity properties of $G/G/1$ (see Loynes [5]), we obtain the inequality

$$m_W^c \leqslant m_W \tag{6}$$

for the stationary mean waiting time m_W in a $G/G/1$ queue with interarrival times a_n and service times β_n and for the mean stationary waiting time m_W^c in $D/G/1$ with constant interarrival times $\lambda^{-1} = Ea_n$ and service times β_n.

Inequality (6) can be used to obtain a lower bound for the mean stationary waiting time m_W in $G/G/1$ and, especially, in $G/G/1$ queues, where the service times are independent of the interarrival times and form a sequence of independent, identically distributed random variables. In this way, in [10], for $GI/G/1$ queues the lower bound

$$m_W \geqslant \frac{\lambda\sigma_B^2}{2(1-\varrho)}-\frac{m_B}{2} \tag{7}$$

was obtained, which is also true for $G/G/1$ queues (λ is the intensity of the input, m_B is the mean service time, σ_B^2 denotes the variance of service times, and $\varrho = \lambda m_B$).

The cited results are applicable to tandem queues $GI/G\to \ldots \to GI/1$ in which s service stations with one server are in series. For such queues, Suzuki and Maruta [11] obtained the upper bound for the stationary mean waiting time at the k-th service station $m_{W,k}$ $(k = 1, 2, \ldots, s)$,

$$m_{W,1} \leqslant \frac{\sigma_A^2 + \sigma_{B,1}^2}{2(m_A - m_{B,1})},$$

$$(8) \qquad m_{W,k} \leqslant \frac{\sigma_A^2 + \sigma_{B,k}^2 + 2\sum_{i=1}^{k-1} \sigma_{B,i}^2}{2m_A(1-\varrho_k)} - \frac{\sum_{i=1}^{k-1} (1-\varrho_i) m_{W,i}}{1-\varrho_k} \qquad (i = 2, 3, \ldots),$$

where m_A is the mean interarrival time of the system input, $\sigma_{B,k}^2$ is the variance of the service time at the k-th station, σ_A^2 denotes the variance of the interarrival times of the system input, and $\varrho_k = m_{B,k}/m_A$. This bound in its original form is not often used in practice, since the authors do not state how to obtain the exact values or lower bounds for $m_{W,i}$ $(i < k)$.

Using (7), for $m_{W,k}$ we obtain the lower bound

$$(9) \qquad m_{W,k} \geqslant \frac{\sigma_{B,k}^2}{2m_A(1-\varrho_k)} - \frac{m_{B,k}}{2} \qquad (k = 1, 2, \ldots, s).$$

Combining (8) with (9), we obtain

$$(10) \qquad m_{W,k} \leqslant \frac{\sigma_A^2 + \sum_{i=1}^{k} \sigma_{B,i}^2}{2m_A(1-\varrho_k)} + \frac{\sum_{i=1}^{k-1} (1-\varrho_i) m_{B,i}}{2(1-\varrho_k)},$$

which does not have the above-mentioned shortcomings. As to the inaccuracy of (9) and (10), it can be pointed out that the order of inaccuracy of (9) is equal for all k, whereas it increases with increasing k for (10).

Added in proof. Using Theorem 3.3 in the paper by D. Stoyan, *Further stochastic order relations among GI/GI/1 queues with a common traffic intensity*, in Math. Operationsforsch. Statist., Series Optimization, 8 (1977), it is possible to prove the exponential bound

$$\overline{W}(t) \leqslant \frac{\lambda}{\mu s} \exp\left[-\left(\mu - \frac{\lambda}{s}\right) t\right] \qquad (0 \leqslant t < \infty),$$

which is better than (1).

References

[1] R. E. Barlow and F. Proschan, *Statistical theory of reliability and life testing*, Holt, Rinhart and Winston 1975.

[2] R. Bergmann and D. Stoyan, *On exponential bounds for the waiting time distribution function in GI/G/1*, J. Appl. Prob. 13 (1976), p. 411-417.

[3] S. L. Brumelle, *Some inequalities for parallel-server queues*, Opns. Res. 19 (1971), p. 402-413.

[4] J. F. C. Kingman, *Inequalities in the theory of queues*, J. Roy. Statist. Soc. B 32 (1970), p. 102-110.

[5] R. M. Loynes, *The stability of a queue with non-independent inter-arrival and service time*, Proc. Camb. Phil. Soc. 58 (1962), p. 494-520.

[6] T. Rolski, *On some classes of distribution functions determined by an order relation*, Proc. Symp. Math. Statist. J. Neyman, Warszawa 1976.

[7] H.-J. Rossberg and G. Siegel, *Die Bedeutung von Kingman's Integralungleichungen bei der Approximation der stationären Wartezeitverteilung im Modell GI/G/1 mit und ohne Verzögerung beim Beginn einer Beschäftigungsperiode*, Math. Operationsforschung Statist. 5 (1974), p. 687-699.

[8] D. Stoyan, *Monotonieeigenschaften stochastischer Modelle*, Z. angew. Math. Mech. 52 (1972), p. 23-32.

[9] — and H. Stoyan, *Bedienungstheoretische Anwendung von Halbordnungsrelationen für Verteilungsgesetze*, Wiss. Z. TU Dresden 21 (1972), p. 519-523.

[10] — *Inequalities for the mean waiting time in singule-server queues*, Izv. AN SSSR, Techn. Kib. No. 6 (1974), p. 104-106.

[11] T. Suzuki and M. Maruta, *Inequalities for two bulk queues and a tandem queue*, Mem. Defense Acad. 10 (1970), p. 81-102.

BERGAKADEMIE FREIBERG
SEKTION MATHEMATIK
DDR-92 FREIBERG

Received on 23. 2. 1976

D. STOYAN (Freiberg, NRD)

NIERÓWNOŚCI DLA WIELOKANAŁOWYCH I WIELOFAZOWYCH SYSTEMÓW OBSŁUGI MASOWEJ

STRESZCZENIE

W pracy podano ograniczenie wykładnicze dla stacjonarnych funkcji rozkładu czasu oczekiwania w szerokiej klasie systemów obsługi masowej typu $GI/GI/s$. Ponadto podane są dolne i górne ograniczenia średnich stacjonarnych czasów czekania w wielofazowych systemach obsługi masowej.
