**Centrum voor Wiskunde en Informatica**
Centre for Mathematics and Computer Science

O.J. Boxma, B. Meister

# Waiting-time approximations for cyclic-service systems with switch-over times

Department of Operations Research and System Theory Report OS-R8510 September

The Centre for Mathematics and Computer Science is a research institute of the Stichting Mathematisch Centrum, which was founded on February 11, 1946, as a nonprofit institution aiming at the promotion of mathematics, computer science, and their applications. It is sponsored by the Dutch Government through the Netherlands Organization for the Advancement of Pure Research (Z.W.O.).

# Waiting-Time Approximations for Cyclic-Service Systems with Switch-Over Times

O.J. Boxma
*Centre for Mathematics and Computer Science*
*P.O. Box 4079, 1009 AB Amsterdam, The Netherlands*

B. Meister
*IBM Zurich Research Laboratory*
*CH-8803 Rüschlikon, Switzerland*

Mean waiting-time approximations are derived for a single-server multi-queue system with nonexhaustive cyclic service. Non-zero switch-over times of the server between consecutive queues are assumed. The main tool used in the derivation is a pseudo-conservation law recently found by Watson. The approximation is simpler and, as extensive simulations show, more accurate than existing approximations. Moreover, it gives very good insight into the qualitative behavior of cyclic-service queueing systems.

*1980 Mathematics Subject Classification:* 60K25, 60K30, 68M20. 69C40
*Key Words & Phrases:* queueing system, cyclic service, switch-over times, mean waiting time.
*Note:* This report will be submitted for publication elsewhere.

## 1. Introduction

In local area networks with ring or bus topology, medium access control protocols based on token passing have become increasingly popular. Such networks can be modeled as single-server multi-queue systems with a cyclic-service discipline, e.g., exhaustive, gated or nonexhaustive service. When token rings or buses become longer and/or faster such that propagation delays are noticeable, it becomes important to model the times for passing the token, subsequently called switch-over times. In real-time applications a number of stations, e.g., measurement devices, is often scanned in a fixed order. Again, switch-over times of the server may have an impact on system performance, especially when task switching takes place between the service of two (consecutive) stations.

A basic queueing model for the performance evaluation of such cyclic-service systems with switch-over times is investigated in this paper; it will now be described in detail.

### *Model description*

A single service facility serves $N$ queues $Q_1, \ldots, Q_N$ (with infinite buffer capacities) in a cyclic manner. The service discipline considered is ordinary cyclic or nonexhaustive service (sometimes also called chaining, polling, or alternating service): When the server visits a queue, it only serves one customer (if any is present). The switch-over times of the server between the i-th and (i+1)-th queue are independent, identically distributed stochastic variables $S_i$ with first moment $s_i$ and variance $\Psi_i^2$. The mean of the total switch-over time during a cycle of the server, $s$, is given by

Report OS-R8510
Centre for Mathematics and Computer Science
P.O. Box 4079, 1009 AB Amsterdam, The Netherlands

$$s = \sum_{1}^{N} s_i. \tag{1.1}$$

Customers arrive at all queues according to independent Poisson processes with rates $\lambda_1, \ldots, \lambda_N$; the total arrival rate is $\Lambda$. Customers which arrive at $Q_i$ are called type-i customers. The service times of type-i customers are independent, identically distributed stochastic variables with distribution $B_i(.)$, with first and second moments $\beta_i$ and $\beta_i^{(2)}$; the service process is also independent of the arrival process and of the switch-over process. The utilization at $Q_i$ , $\rho_i$ , is defined as

$$\rho_i = \lambda_i \beta_i, \qquad i=1,...,N. \tag{1.2}$$

The total utilization of the server, $\rho$ , is defined as

$$\rho = \sum_{1}^{N} \rho_i. \tag{1.3}$$

It was shown by Kuehn [7] that the following conditions are necessary and sufficient for stability of the system:

$$\rho < 1 \quad \text{and} \quad \max(\lambda_i)\, s < 1-\rho. \tag{1.4}$$

In fact, it is easily shown (cf. Kuehn [7]) in the stationary situation that the mean cycle time for $Q_i$, i.e., the mean interarrival time of the server at $Q_i$ , is independent of i, and is given by

$$Ec = \frac{s}{1-\rho}, \tag{1.5}$$

which immediately implies the necessity of the above stability conditions.

Important performance measures in multi-queue systems are the mean waiting times $Ew_i$ at the individual queues $Q_i$ , i=1,...,N. In the case of nonexhaustive service, considered here, the determination of exact values of the mean waiting times is an extremely complicated mathematical problem which could not be solved so far except for a few special cases. A complete exact analysis of the case of N=2 queues without switch-over times and of the case of two queues with identical characteristics with switch-over times has been presented in [5] and [4] and in [1], respectively (also leading to waiting-time and queue-length distributions). The solution method transforms the problem into a Riemann-Hilbert boundary value problem, and it is not yet clear how it can be generalized to solve the model with more than two queues. Using a different method, Nomura and Tsukamoto [9] and Takagi (cf. [12]) have obtained the exact mean waiting times for a system with an arbitrary number of queues which all have identical characteristics. The excellent survey of Takagi and Kleinrock [12] contains several further references.
*Note:* This case, in which switch-over times, arrival rates and service time distributions are the same for each queue, will be denoted in the sequel as the completely symmetric case.

The intractability of the general model has led several authors to the development of mean waiting-time approximations. An important approximation is due to Kuehn [7], who obtains, a.o., mean waiting-time approximations for nonexhaustive cyclic-service systems with and without switch-over times and with batch Poisson input. Earlier references for mean waiting-time approximations can also be found in [7]. An approximation for systems with multiple cyclic servers is given in [8]; the case of cyclic systems with finite-capacity queues has been considered in [13].

In the present paper, the method used in [2] for cyclic service systems *without* switch-over times is generalized to obtain simple yet accurate mean waiting-time approximations for the model *with* switch-over times. This generalization is made possible by means of a pseudo-conservation law, recently obtained for this model by Watson [14]. The approximation is derived in Section 2. In Section 3 the accuracy of the approximation is assessed. Some conclusions are presented in Section 4.

## 2. The approximation

We first need some definitions:
$x_i$ denotes the queue length at $Q_i$ just before the arrival of a type-i customer;
$c_i$ denotes the length of a cycle of the server which starts with a service at $Q_i$ and ends when the server returns to $Q_i$ (an "i-cycle");
$rc_i$ denotes a residual i-cycle, i.e., the time from the arrival of a type-i customer until the server returns to $Q_i$. An arriving type-i customer first has to wait until the server returns to $Q_i$ and subsequently he has to wait until all customers in front of him have been served. Therefore, the mean waiting time of this customer consists of two parts: a residual cycle $rc_i$ and just as many i-cycles as there are type-i customers waiting; hence

$$Ew_i = Erc_i + Ex_i\, Ec_i\,. \tag{2.1}$$

Owing to the fact that Poisson arrivals see time averages (cf. Wolff [15]) $Ex_i$ equals the mean number of waiting customers at $Q_i$ at an arbitrary instant of time. This permits the use of Little's formula, yielding

$$Ew_i = \frac{Erc_i}{1 - \lambda_i\, Ec_i}\,. \tag{2.2}$$

Similarly to [2], we introduce two approximation assumptions to estimate the two unknowns $Ec_i$ and $Erc_i$.

*Assumption 1:*

$$Ec_i = \frac{\beta_i + s}{1 - \rho + \rho_i}\,, \qquad i = 1,...,N. \tag{2.3}$$

This approximation, which is due to Kuehn [7], can be motivated as follows. An i-cycle consists of a type-i service and, possibly, services of customers of other types, plus the sum of $N$ switch-over times. Define

$$\begin{aligned} \alpha_{ij} &= Pr(i-cycle\ contains\ a\ type-j\ service) \qquad (2.4)\\ &= E[number\ of\ type-j\ services\ in\ an\ i-cycle] \approx \lambda_j\, Ec_i\,, \qquad j{\neq}i; \end{aligned}$$

the second equality holds because an i-cycle contains at most one type-i service.
Hence,

$$Ec_i = \beta_i + s + \sum_{j\neq i} \alpha_{ij}\, \beta_j\,, \tag{2.5}$$

which together with (2.4) yields our assumption (2.3). Equation (2.3) is trivially exact for N=1. The approximation in (2.4) is based on a balance-of-flow argument. It should be very accurate in the completely symmetric case; it should also be very accurate for light traffic, but not for heavy traffic with highly asymmetric arrival rates and/or service demands.

*Assumption 2:*

$Erc_i$ is independent of i.

This assumption is trivially exact for N = 1 and in the completely symmetric case. In the limiting case $\rho = 0$ it is also true, as can be seen in the following way. Consider, e.g., $Erc_1$ :

$$Erc_1 = \sum_{j=1}^{N} \frac{s_j}{s}\,(\tilde{s}_j + s_{j+1} + \cdots + s_N),$$

where $\tilde{s}_j$ is the mean residual switch-over time between $Q_j$ and $Q_{j+1}$. Using

$$\tilde{s}_j = \frac{ES_j^2}{2\,ES_j} = \frac{\Psi_j^2 + s_j^2}{2s_j},$$

it follows easily that

$$Erc_1 = \sum_{j=1}^{N} \frac{\Psi_j^2}{2s} + \frac{s}{2}.$$

This last expression equals the mean residual lifetime of $S$. For symmetry reasons, hence also $Erc_2,...,Erc_N$ equal the same expression (which could also have been derived from a simple probabilistic argument).

For small values of $\rho$, the probability that a type-i customer finds other customers present upon his arrival (anywhere in the system) is $O(\rho)$. Furthermore, the mean contribution to $Erc_i$ of work of other customers, arriving between his arrival and the moment at which the server reaches $Q_i$, is also $O(\rho)$. Hence,

$$Erc_i = \sum_{j=1}^{N} \frac{\Psi_j^2}{2s} + \frac{s}{2} + O(\rho), \quad \rho \to 0.$$

Unlike the case of zero switch-over times [2], the $O(\rho)$ term is not completely independent of $i$, but its influence is negligible for small values of $\rho$, because of the domination of the $O(1)$ term. Therefore Assumption 2 should be accurate for low traffic.

The only unknown in the expression (2.2) for $Ew_i$ is $Erc \equiv Erc_i$, which will be determined by means of the following pseudo-conservation law, due to Watson [14]:

$$\sum_{i=1}^{N} \rho_i\,(1-\alpha_i)\,Ew_i = \frac{\rho}{2(1-\rho)} \sum_{j=1}^{N} \lambda_j\,\beta_j^{(2)} + \frac{\rho}{2s} \sum_{j=1}^{N} \Psi_j^2 + \frac{s}{2(1-\rho)} \sum_{j=1}^{N} \rho_j\,(1+\rho_j), \tag{2.6}$$

with $\alpha_i$ defined as

$$\alpha_i = \lambda_i \frac{s}{1-\rho}; \tag{2.7}$$

$\alpha_i$ is the probability that the server, upon arrival at $Q_i$, finds at least one customer present.

*Remark 2.1*

Watson has derived formula (2.6) by first writing down a set of N recurrence relations for the N generating functions of the joint stationary queue-length distributions at arrival instants of the server at the various queues, and subsequently differentiating these relations twice, after each differentiation taking all generating function parameters equal to one. Finally, he arrives at (2.6) by cleverly eliminating all but N unknowns, which are simply expressed in the $Ew_i$.

If all switch-over times are zero, (2.6) reduces to Kleinrock's [6] conservation law for M/G/1-type queues: the right-hand side of (2.6) in this case constitutes the mean waiting time in an M/G/1 queue with arrival rate $\Lambda$ and with service-time distribution being a weighted sum of the individual service-time distributions. In the present paper, Relation (2.6) is called a pseudo-conservation law because the expression in the left-hand side has no (known) relation to a single-queue system.

Note that the expression in the right-hand side of (2.6) only involves the first two moments of the service- and switch-over times, and is independent of the polling order of the queues.

An estimate for $Erc \equiv Erc_i$ will be obtained by demanding that the mean waiting-time approximation fulfills the pseudo-conservation law of Watson (note that this immediately implies that the approximation also has the desirable properties of being exact for N=1 and in the completely symmetric case). From (2.2) and the two above assumptions,

$$Ew_i = Erc \frac{1-\rho+\rho_i}{1-\rho-\lambda_i s}, \quad i = 1,...,N. \tag{2.8}$$

Substituting these $Ew_i$ into (2.6) yields:

$$Erc = \frac{1-\rho}{(1-\rho)\rho + \sum_{j=1}^{N} \rho_j^2} \left[ \frac{\rho}{2(1-\rho)} \sum_{j=1}^{N} \lambda_j \beta_j^{(2)} + \frac{\rho}{2s} \sum_{j=1}^{N} \Psi_j^2 + \frac{s}{2(1-\rho)} \sum_{j=1}^{N} \rho_j (1+\rho_j) \right]. \tag{2.9}$$

Finally, this yields our main result:

$$Ew_i \approx \frac{1-\rho+\rho_i}{1-\rho-\lambda_i s} \frac{1-\rho}{(1-\rho)\rho + \sum_{j=1}^{N} \rho_j^2} \left[ \frac{\rho}{2(1-\rho)} \sum_{j=1}^{N} \lambda_j \beta_j^{(2)} + \frac{\rho}{2s} \sum_{j=1}^{N} \Psi_j^2 + \frac{s}{2(1-\rho)} \sum_{j=1}^{N} \rho_j (1+\rho_j) \right], \quad i = 1,...,N. \tag{2.10}$$

*Remark 2.2*
In the special case of zero switch-over times, approximation (2.10) reduces to the approximation given in [2]:

$$Ew_i \approx \frac{1-\rho+\rho_i}{(1-\rho)\rho + \sum_{j=1}^{N} \rho_j^2} \frac{\rho}{2(1-\rho)} \sum_{j=1}^{N} \lambda_j \beta_j^{(2)}. \tag{2.11}$$

In the case N=1 it reduces to the exact mean waiting time for an M/G/1 model with vacations (see Skinner [11]); in the completely symmetric case it reduces to the exact result which has been derived in [9] and [12].

*Remark 2.3*
According to (2.10),

$$\frac{Ew_i}{Ew_j} \approx \frac{1-\rho+\rho_i}{1-\rho+\rho_j} \frac{1-\rho-\lambda_j s}{1-\rho-\lambda_i s}. \tag{2.12}$$

Formulas (2.10) and (2.12) suggest that the mean waiting time at a queue is much more sensitive to a change of arrival rate than to a change of mean service time. In particular, two queues in heavy traffic with the same service-time distribution but with slightly different arrival rates may have quite different mean waiting times.
Formula (2.10) also suggests that the mean switch-over time, s, can have a strong influence on the mean waiting times, whereas the means and variances of the individual switch-over times are not very critical. These observations will be confirmed by the simulation results presented in Section 3.

*Remark 2.4*
It is interesting to compare approximation (2.10) with the mean waiting-time approximation of Bux and Truong [3] for the case of exhaustive service:

$$Ew_i \approx \frac{1-\rho_i}{\rho - \sum_{j=1}^{N} \rho_j^2} \left[ \frac{\rho}{2(1-\rho)} \sum_{j=1}^{N} \lambda_j \beta_j^{(2)} + \frac{s}{2(1-\rho)} \sum_{j=1}^{N} \rho_j (1-\rho_j) \right]. \tag{2.13}$$

This formula was derived for the case of constant switch-over times, and it turns out to satisfy Watson's pseudo-conservation law for the exhaustive service discipline [14]. If the term

$$\frac{\rho}{2s} \sum_{j=1}^{N} \Psi_j^2$$

is added to the expression within brackets in the right-hand side of (2.13), to take random switch-over times into account, then a mean waiting-time approximation will result which is very similar in structure to the approximation (2.10) for the nonexhaustive service discipline.

Formula (2.13) reflects the property of the exhaustive service discipline that customers in light-traffic queues usually experience a longer waiting time than customers in heavy-traffic queues: customers arriving at a heavy-traffic queue have a better chance that their queue is currently being served than those arriving at a light-traffic queue. The nonexhaustive service discipline without switch-over times, on the other hand, leads to relatively small waiting times at light-traffic queues, as can be seen from (2.11). For non-zero switch-over times such a general statement cannot be made, but in most cases the behavior is similar to that for zero switch-over times (cf. (2.10) and (2.12)).

The derivation of approximation (2.10) suggests that it will be least accurate in heavy, very asymmetric, traffic. Numerical experiments confirm this (cf. Section 3), disclosing the most sensitive heavy-traffic case: if one or more queues $Q_{i_1}, \ldots, Q_{i_j}$ have relatively large arrival rates, so that these queues become nearly unstable (cf. (1.4)), approximation (2.10) has difficulties predicting the mean waiting times at the *other* queues accurately.

Below a modification of the approximation for the latter queues is suggested. As a rule of thumb, we would like to recommend application of this modification in practical situations when the following three conditions are all fulfilled: (i) $\rho \geqslant 0.7$, (ii) the total mean switch-over time is not negligible, and (iii) the arrival rates at a small group of queues are at least three times as high as at any of the other queues.

The basic idea of the modification is the following. Remove the queues with a relatively high arrival rate from the system, and enlarge the switch-over times to compensate for the service times at the removed queues. The resulting system has a lower and more symmetric traffic load, and hence approximation (2.10) becomes much more accurate.

We now present the argument in some more detail. Suppose $Q_i$ is a queue with a relatively high arrival rate $\lambda_i$ (and hence relatively high $\alpha_i$). Consider a cyclic queueing system consisting of the $N-1$ queues $Q_1, \ldots, Q_{i-1}, Q_{i+1}, \ldots, Q_N$, with all queues having the same characteristics as in the original model, and with the switch-over time $SW_{i-1}$ from $Q_{i-1}$ to $Q_{i+1}$ being defined as:

$$SW_{i-1} = S_{i-1} + \tau_i + S_i, \tag{2.14}$$

where

$$Pr(\tau_i = 0) = 1 - \alpha_i,$$

$$Pr(\tau_i < t) = 1 - \alpha_i + \alpha_i B_i(t), \quad t > 0,$$

with

$$E\tau_i = \alpha_i \beta_i, \quad E\tau_i^2 = \alpha_i \beta_i^{(2)}.$$

So the switch-over time from $Q_{i-1}$ to $Q_{i+1}$ is composed of the switch-over times from $Q_{i-1}$ to $Q_i$ and from $Q_i$ to $Q_{i+1}$ in the original model, plus a stochastic variable $\tau_i$ which takes account of a possible service time in $Q_i$ in the original model. Clearly, when $\alpha_i$ is close to one, $Q_1, \ldots, Q_{i-1}, Q_{i+1}, \ldots, Q_N$ should behave very similarly in both models.
If another queue also has a relatively high arrival rate, the same reasoning should be applied once more, etc. In the finally resulting model, the total utilization will not be very high while the traffic is less asymmetric than in the original system; and Assumptions 1 and 2, in combination with the pseudo-conservation law (2.6) for this model, will lead to satisfactory mean waiting-time approximations for the queues of the modified model - and hence also for the queues with relatively low arrival rates of the original model.

Summarizing, in the modified approximation the mean waiting times in the queues with relatively low arrival rates are approximated by using (2.10) in the modified model with fewer queues and different total utilization and switch-over times.

## 3. Comparison with simulation

This section presents a comparison of the mean waiting-time approximation with simulation results, generated with the IBM RESQ2 package [10], and with the well-known approximation of Kuehn [7], together with some general observations. The numerical results are collected in six tables at the end of the paper. Representative examples have been chosen to estimate the accuracy of the approximation for different parameter combinations and service-time distributions. Watson's pseudo-conservation law permits a convenient additional validation of the accuracy of the simulation. The simulation results "fulfill" this law with an error of about 2% for $\rho$ up to 0.5 and an error of about 5% for $\rho = 0.8$.
The relative error of approximation (2.10), given in the tables, is defined as

$$100\% \, \frac{(\text{approximation result} - \text{simulation result})}{\text{simulation result}} .$$

A more detailed discussion of the results follows.
Tables 1 and 2 show results for $N=3$ queues. In Table 1 the arrival rates are equal but the service times different, whereas in Table 2 different arrival rates but equal service times have been chosen. The tables show that the effect of a higher arrival rate is much stronger than that of a higher mean service time. In Table 1 the mean waiting times at all queues are roughly the same, although the mean service times differ by a factor of three. In Table 2, where the arrival rates differ by a factor of three, this is no longer true in heavy traffic: the mean waiting times at the heavy-traffic queue are much larger than those at the other queues.
Comparing mean waiting times at the low-traffic queues in Tables 1 and 2 (which have the same utilization in both tables) it can be seen that, although the mean service times at $Q_2$ and $Q_3$ in Table 1 are *smaller* than those in Table 2, the mean waiting times at $Q_2$ and $Q_3$ in Table 1 are *larger* - due to the fact that the arrival rates are higher.
Tables 1 and 2 also reveal that the influence of random switch-over times in comparison to constant switch-over times is only marginal. All the above-mentioned phenomena are correctly predicted by the form of (2.10) (cf. also Remark 2.3).

The stability condition (1.4) indicates that $\gamma_i := \rho + \lambda_i s$ must be smaller than one. If $\gamma_i$ is nearly one, the mean waiting time at $Q_i$ becomes very large even if $\rho$ is considerably smaller than one. An example is the case $\rho=0.8$ and $s_i=0.1$ in Table 2, for which $\gamma_1=0.98$. The original approximation

(2.10) yields an error of about 50% for the low-traffic queues. In Table 2, for $\rho=0.8$, the modified approximation for the low-traffic queues has been used; this way good results have also been obtained for this extreme case.

As already mentioned in [2], the mean waiting times at different queues with identical characteristics need not be the same, as they depend slightly on the locations of these queues with respect to queues with other traffic patterns. Our approximation does not take this effect into account (and neither does Kuehn's approximation). In our simulations, these differences have been very small. Therefore, mean waiting times at consecutive queues with identical characteristics are only represented by their average in the tables.

The Tables 3-6 give results for $N=16$ queues. Only constant switch-over times are considered, as the choice of the switch-over time distributions has little bearing on the results. Table 3 is similar to Table 1, but now $Q_1$ and $Q_7$ have relatively long mean service times; (2.10) gives a good approximation. In Tables 4 and 5, as in Table 2, different arrival rates are considered. $\gamma_1 = \cdots = \gamma_4 = 0.928$ and $\gamma_1 = 0.896$ in these respective tables if $\rho=0.8$. The modified approximation has been used for the low-traffic queues in these cases, removing $Q_1, \ldots, Q_4$ and $Q_1$, respectively.
The combined effect of different service-time distributions and different arrival rates is shown in Table 6. This case is very asymmetric, $\rho_1$ and $\rho_7$ being 18 times as large as the other $\rho_i$. Here, too, in the heavy-traffic case (with $\gamma_1 = \gamma_7 = 0.92$) the modified approximation has been used, but here the improvement is negligible.

## 4. Conclusions

A simple mean waiting-time approximation for nonexhaustive cyclic-service systems with switch-over times has been derived and investigated. The results can be summarized as follows.
- Approximation (2.10) is constructed in such a way that it fulfills Watson's pseudo-conservation law, and hence it is in particular exact for the completely symmetric case.
- The approximation gives considerable insight into both the qualitative and quantitative behavior of the mean waiting times.
- The approximation yields generally better results than known approximations.
- The relative error of the approximation is a few percent for utilizations up to 0.5 in all of our examples. For traffic patterns which are not too asymmetric, the error is rather small for a utilization of 0.8 (cf. Table 3). The error in (2.10) becomes larger in cases of strong asymmetry, when some of the queues become nearly unstable. In such cases, the modified approximation described at the end of Section 2 usually leads to considerable improvements.
- The approximation accuracy generally improves with an increasing number of queues, a property which seems to hold for all approximations known. This might be explained by an "averaging out" effect which stabilizes systems with a larger number of queues.
- The error is very insensitive to changes, in mean or distribution, of the switch-over times.

## Acknowledgment

The authors would like to express their thanks to W. Bux for reviewing the manuscript.

REFERENCES

1. O.J. BOXMA (1984). *Two symmetric queues with alternating service and switching times,* In: Performance '84, ed. E. Gelenbe (North-Holland, Amsterdam) pp. 409-431.
2. O.J. BOXMA, B. MEISTER (1985). *Waiting times in multi-queue systems with cyclic service,* IBM Research Report RZ 1348.
3. W. BUX, H.L. TRUONG (1983). *Mean-delay approximations for cyclic-service queueing systems,* Performance Evaluation 3, 187-196.
4. J.W. COHEN, O.J. BOXMA (1983). *Boundary Value Problems in Queueing Systems Analysis* (North-Holland, Amsterdam).
5. M. EISENBERG (1979). *Two queues with alternating service,* SIAM J. Appl. Math. 36, 287-303.
6. L. KLEINROCK (1976). *Queueing Systems Vol. II* (Wiley, New York).
7. P.J. KUEHN (1979). *Multi-queue systems with non-exhaustive cyclic service,* Bell Syst. Techn. J. 58, 671-698.
8. R.J.T. MORRIS, Y.T. WANG (1984). *Some results for multi-queue systems with multiple cyclic servers,* In: Performance of Computer-Communications Systems, eds. H. Rudin and W. Bux (North-Holland, Amsterdam), pp. 245-258.
9. M. NOMURA, K. TSUKAMOTO (1978). *Traffic analysis on polling systems,* The Transactions of the Institute of Electronics and Communication Engineers of Japan, Vol.J61-B, No. 7, pp. 600-607 (in Japanese).
10. C.H. SAUER, E.A. MACNAIR (1983). *Simulation of Computer Communication Systems,* (Prentice-Hall, Englewood Cliffs).
11. C.E. SKINNER (1967). *A priority queuing system with server-walking time,* Operations Res. 15, 278-285.
12. H. TAKAGI, L. KLEINROCK (1985). *A tutorial on the analysis of polling systems,* IBM Japan Science Institute Report TR87.
13. P. TRAN-GIA, T. RAITH (1984). *Multiqueue systems with finite capacity and nonexhaustive cyclic service,* Report Inst. of Communications Switching and Data Technics, Univ. of Stuttgart.
14. K.S. WATSON (1984). *Performance evaluation of cyclic service strategies - a survey,* In: Performance '84, ed. E. Gelenbe (North-Holland Publ. Cy., Amsterdam) pp. 521-533.
15. R.W. WOLFF (1982). *Poisson arrivals see time averages,* Operations Res. 30, 223-231.

TABLE 1

Comparison of the mean waiting-time approximation (2.10) with simulation and with Kuehn's approximation. $N=3$ queues, $\Lambda=1$, $\lambda_1=\lambda_2=\lambda_3=1/3$; all service-time distributions negative exponential with $\beta_2=\beta_3=(1/3)\beta_1$.

All switch-over times equal to 0.05.

| $\rho$ | 0.3 | 0.5 | 0.8 |
|---|---|---|---|
| $Ew_1$ simulation | 0.333 | 1.003 | 6.80 |
| $Ew_1$ approx. (2.10) | 0.331 | 1.041 | 7.77 |
| Error % | -0.6 | 3.8 | 14.3 |
| $Ew_1$ approx. Kuehn | 0.317 | 0.939 | 6.29 |
| $Ew_{2-3}$ simulation * | 0.289 | 0.830 | 5.38 |
| $Ew_2$ approx. (2.10) | 0.286 | 0.780 | 4.11 |
| Error % | -1.0 | -6.0 | -23.6 |
| $Ew_2$ approx. Kuehn | 0.263 | 0.645 | 3.00 |

All switch-over times equal to 0.10.

| $\rho$ | 0.3 | 0.5 | 0.8 |
|---|---|---|---|
| $Ew_1$ simulation | 0.506 | 1.381 | 10.72 |
| $Ew_1$ approx. (2.10) | 0.509 | 1.425 | 12.90 |
| Error % | 0.6 | 3.2 | 20.3 |
| $Ew_1$ approx. Kuehn | 0.493 | 1.309 | 10.64 |
| $Ew_{2-3}$ simulation * | 0.444 | 1.155 | 8.30 |
| $Ew_2$ approx. (2.10) | 0.439 | 1.069 | 6.83 |
| Error % | -1.1 | -7.4 | -17.7 |
| $Ew_2$ approx. Kuehn | 0.415 | 0.922 | 5.31 |

* The results represent mean waiting times averaged over the corresponding group of queues

TABLE 1 (Cont'd)

All switch-over times negative exponentially distributed with mean 0.05.

| $\rho$ | 0.3 | 0.5 | 0.8 |
|---|---|---|---|
| $Ew_1$ simulation | 0.356 | 1.056 | 6.90 |
| $Ew_1$ approx. (2.10) | 0.360 | 1.071 | 7.81 |
| Error % | 1.1 | 1.4 | 13.2 |
| $Ew_1$ approx. Kuehn | 0.341 | 0.961 | 6.31 |
| $Ew_{2-3}$ simulation * | 0.314 | 0.869 | 5.59 |
| $Ew_2$ approx. (2.10) | 0.311 | 0.804 | 4.13 |
| Error % | -1.0 | -7.5 | -26.1 |
| $Ew_2$ approx. Kuehn | 0.284 | 0.662 | 3.01 |

All switch-over times negative exponentially distributed with mean 0.10.

| $\rho$ | 0.3 | 0.5 | 0.8 |
|---|---|---|---|
| $Ew_1$ simulation | 0.570 | 1.394 | 11.26 |
| $Ew_1$ approx. (2.10) | 0.570 | 1.494 | 13.02 |
| Error % | 0.0 | 7.2 | 15.6 |
| $Ew_1$ approx. Kuehn | 0.545 | 1.359 | 10.71 |
| $Ew_{2-3}$ simulation * | 0.502 | 1.196 | 8.60 |
| $Ew_2$ approx. (2.10) | 0.493 | 1.121 | 6.89 |
| Error % | -1.8 | -6.3 | -19.9 |
| $Ew_2$ approx. Kuehn | 0.459 | 0.960 | 5.34 |

* The results represent mean waiting times averaged over the corresponding group of queues

TABLE 2

Comparison of the mean waiting-time approximation (2.10) with simulation and with Kuehn's approximation. $N=3$ queues, $\Lambda=1$, $\lambda_1=0.6$, $\lambda_2=\lambda_3=0.2$; all service-time distributions negative exponential with identical means.

All switch-over times equal to 0.05.

| $\rho$ | 0.3 | 0.5 | 0.8 |
|---|---|---|---|
| $Ew_1$ simulation | 0.304 | 0.937 | 9.34 |
| $Ew_1$ approx. (2.10) | 0.303 | 0.925 | 8.30 |
| Error % | -0.3 | -1.3 | -11.1 |
| $Ew_1$ approx. Kuehn | 0.288 | 0.812 | 6.31 |
| $Ew_{2-3}$ simulation * | 0.236 | 0.581 | 1.89 |
| $Ew_2$ approx. (2.10) | 0.238 | 0.605 | 1.47 |
| Error % | 0.8 | 4.1 | -22.2 |
| $Ew_2$ approx. Kuehn | 0.225 | 0.535 | 2.47 |

All switch-over times equal to 0.10.

| $\rho$ | 0.3 | 0.5 | 0.8 |
|---|---|---|---|
| $Ew_1$ simulation | 0.525 | 1.510 | 55.70 |
| $Ew_1$ approx. (2.10) | 0.528 | 1.503 | 51.91 |
| Error % | 0.6 | -0.5 | -6.8 |
| $Ew_1$ approx. Kuehn | 0.510 | 1.356 | 40.77 |
| $Ew_{2-3}$ simulation * | 0.370 | 0.775 | 2.31 |
| $Ew_2$ approx. (2.10) | 0.371 | 0.820 | 2.22 |
| Error % | 0.3 | 5.8 | -3.9 |
| $Ew_2$ approx. Kuehn | 0.358 | 0.750 | 3.58 |

* The results represent mean waiting times averaged over the corresponding group of queues

TABLE 2 (Cont'd)

All switch-over times negative exponentially distributed with mean 0.05.

| $\rho$ | 0.3 | 0.5 | 0.8 |
|---|---|---|---|
| $Ew_1$ simulation | 0.333 | 0.976 | 9.09 |
| $Ew_1$ approx. (2.10) | 0.334 | 0.959 | 8.36 |
| Error % | 0.3 | -1.7 | -8.0 |
| $Ew_1$ approx. Kuehn | 0.313 | 0.836 | 6.34 |
| $Ew_{2-3}$ simulation * | 0.261 | 0.599 | 1.92 |
| $Ew_2$ approx. (2.10) | 0.262 | 0.628 | 1.48 |
| Error % | 0.4 | 4.8 | -22.9 |
| $Ew_2$ approx. Kuehn | 0.245 | 0.551 | 2.48 |

All switch-over times negative exponentially distributed with mean 0.10.

| $\rho$ | 0.3 | 0.5 | 0.8 |
|---|---|---|---|
| $Ew_1$ simulation | 0.600 | 1.625 | 53.87 |
| $Ew_1$ approx. (2.10) | 0.600 | 1.590 | 52.53 |
| Error % | 0.0 | -2.2 | -2.5 |
| $Ew_1$ approx. Kuehn | 0.569 | 1.419 | 41.19 |
| $Ew_{2-3}$ simulation * | 0.418 | 0.825 | 2.37 |
| $Ew_2$ approx. (2.10) | 0.421 | 0.867 | 2.25 |
| Error % | 0.7 | 5.1 | -5.1 |
| $Ew_2$ approx. Kuehn | 0.399 | 0.784 | 3.60 |

* The results represent mean waiting times averaged over the corresponding group of queues

TABLE 3

Comparison of the mean waiting-time approximation (2.10) with simulation and with Kuehn's approximation. $N=16$ queues, $\Lambda=1$, $\lambda_1=\cdots=\lambda_{16}=1/16$, all service-time distributions negative exponential with $\beta_1=\beta_7$, $\beta_2=\cdots=\beta_6=\beta_8=\cdots=\beta_{16}=(1/3)\beta_1$. All switch-over times equal to 0.05.

| $\rho$ | 0.3 | 0.5 | 0.8 |
|---|---|---|---|
| $Ew_1$ simulation | 0.823 | 1.697 | 8.78 |
| $Ew_1$ approx. (2.10) | 0.831 | 1.742 | 10.06 |
| Error % | 1.0 | 2.7 | 14.6 |
| $Ew_1$ approx. Kuehn | 0.796 | 1.513 | 7.35 |
| $Ew_{2-6}$ simulation * | 0.793 | 1.591 | 7.98 |
| $Ew_2$ approx. (2.10) | 0.797 | 1.590 | 7.54 |
| Error % | 0.5 | -0.1 | -5.5 |
| $Ew_2$ approx. Kuehn | 0.752 | 1.301 | 4.58 |
| $Ew_7$ simulation | 0.833 | 1.720 | 8.90 |
| $Ew_7$ approx. (2.10) | 0.831 | 1.742 | 10.06 |
| Error % | -0.2 | 1.3 | 11.8 |
| $Ew_7$ approx. Kuehn | 0.796 | 1.513 | 7.35 |
| $Ew_{8-16}$ simulation * | 0.793 | 1.591 | 7.91 |
| $Ew_8$ approx. (2.10) | 0.797 | 1.590 | 7.54 |
| Error % | 0.5 | -0.1 | -4.6 |
| $Ew_8$ approx. Kuehn | 0.752 | 1.301 | 4.58 |

* The results represent mean waiting times averaged over the corresponding group of queues

TABLE 4

Comparison of the mean waiting-time approximation (2.10) with simulation and with Kuehn's approximation. $N=16$ queues, $\Lambda=1$, $\lambda_1=...=\lambda_4=0.16$, $\lambda_5=...=\lambda_{16}=0.03$; all service-time distributions negative exponential with identical means. All switch-over times equal to 0.05.

| $\rho$ | 0.3 | 0.5 | 0.8 |
|---|---|---|---|
| $Ew_{1-4}$ simulation * | 0.898 | 1.929 | 17.66 |
| $Ew_1$ approx. (2.10) | 0.897 | 1.884 | 16.87 |
| Error % | -0.1 | -2.3 | -4.2 |
| $Ew_1$ approx. Kuehn | 0.863 | 1.646 | 12.02 |
| $Ew_{5-16}$ simulation * | 0.717 | 1.267 | 3.57 |
| $Ew_5$ approx. (2.10) | 0.720 | 1.307 | 3.14 |
| Error % | 0.4 | 3.2 | -12.0 |
| $Ew_5$ approx. Kuehn | 0.689 | 1.122 | 3.36 |

TABLE 5

Comparison of the mean waiting-time approximation (2.10) with simulation and with Kuehn's approximation. $N=16$ queues, $\Lambda=1$, $\lambda_1=0.6$, $\lambda_2=\cdots=\lambda_{16}=2/75$; all service-time distributions negative exponential with identical means. All switch-over times equal to 0.01.

| $\rho$ | 0.3 | 0.5 | 0.8 |
|---|---|---|---|
| $Ew_1$ simulation | 0.330 | 1.015 | 9.71 |
| $Ew_1$ approx. (2.10) | 0.321 | 0.996 | 9.79 |
| Error % | -2.7 | -1.9 | 0.9 |
| $Ew_1$ approx. Kuehn | 0.302 | 0.850 | 6.93 |
| $Ew_{2-16}$ simulation * | 0.222 | 0.495 | 1.35 |
| $Ew_2$ approx. (2.10) | 0.224 | 0.521 | 1.24 |
| Error % | 0.9 | 5.3 | -8.1 |
| $Ew_2$ approx. Kuehn | 0.205 | 0.418 | 1.21 |

* The results represent mean waiting times averaged over the corresponding group of queues

TABLE 6

Comparison of the mean waiting-time approximation (2.10) with simulation and with Kuehn's approximation. $N=16$ queues, $\Lambda=1$, $\lambda_1=\lambda_7=0.15$, $\lambda_2=\cdots=\lambda_6=\lambda_8=\cdots=\lambda_{16}=0.05$; service-time distributions at $Q_2,\ldots,Q_6,Q_8,\ldots,Q_{16}$ negative exponential with identical means; service-time distribution at $Q_1$ Erlang-4 with $\beta_1=6\beta_2$; service-time distribution at $Q_7$ two-stage hyperexponential $q(1-\exp-t/m_1)+(1-q)(1-\exp-t/m_2)$ with $q=0.8873$, $m_1=0.5635\times\beta_7$, $m_2=4.4365\times\beta_7$, $\beta_7=6\beta_2$, $\beta_7^{(2)}=5\beta_7^2$. All switch-over times equal to 0.05.

| $\rho$ | 0.3 | 0.5 | 0.8 |
|---|---|---|---|
| $Ew_1$ simulation | 1.198 | 3.253 | 41.26 |
| $Ew_1$ approx. (2.10) | 1.224 | 3.271 | 33.84 |
| Error % | 2.2 | 0.6 | -18.0 |
| $Ew_1$ approx. Kuehn | 1.153 | 2.755 | 23.02 |
| $Ew_{2-6}$ simulation * | 0.946 | 2.011 | 6.27 |
| $Ew_2$ approx. (2.10) | 0.940 | 2.027 | 4.90 |
| Error % | -0.6 | 0.8 | -21.9 |
| $Ew_2$ approx. Kuehn | 0.868 | 1.610 | 4.72 |
| $Ew_7$ simulation | 1.247 | 3.335 | 39.21 |
| $Ew_7$ approx. (2.10) | 1.224 | 3.271 | 33.84 |
| Error % | -1.8 | -1.9 | -13.7 |
| $Ew_7$ approx. Kuehn | 1.153 | 2.755 | 23.02 |
| $Ew_{8-16}$ simulation * | 0.922 | 1.902 | 6.17 |
| $Ew_8$ approx. (2.10) | 0.940 | 2.027 | 4.90 |
| Error % | 2.0 | 6.6 | -20.6 |
| $Ew_8$ approx. Kuehn | 0.868 | 1.610 | 4.72 |

* The results represent mean waiting times averaged over the corresponding group of queues