

The Distribution of Cycle Times in Tree-like Networks of Queues

P. G. Harrison

Department of Computing, Imperial College of Science and Technology, 180 Queen's Gate, London SW7 2BZ, UK

The time delays experienced by tasks in computer systems are a prime interest for both the user community and installation management. Thus their prediction becomes an important objective for the computer performance analyst. Cycle time in scheduling systems and response time, an aggregation of cycle times, in interactive systems are typical examples. The statistical characteristics of time delays have been represented predominantly by simulation models. In analytical models, based on queueing network analysis, normally only their mean values have been derived using Little's law. An exact derivation is presented for the distribution of cycle times in so-called tree-like queueing networks. The analysis is performed for a choice of network structure which avoids the need for explicit tagging of some test customer. Thus expansion of the state space is not necessary. Cycle time distribution is derived in the form of its Laplace transform, from which its moments follow. Further, a recurrence relation for a uniformly convergent discrete representation of the distribution may be determined in a similar manner. Numerical examples show how the distribution of cycle time and its standard deviation vary as the population of a network increases, and how the exact formulae may be used to validate other types of model, such as approximate analytic or simulation.

1. INTRODUCTION

Prediction of characteristics of the time delays experienced by tasks in a computer system or messages in a communication network is of great importance to the performance analyst. For example, in a process control system one could predict the probability of occurrence of a system fault caused by failure of a scheduler to complete a cycle of work in some minimum specified time, determined in real time by exogenous events. Similarly in communication network modelling, the probability of a message transmission taking longer than some given time may be required. Aggregation of a sequence of time delays such as these is sometimes the main interest. For example, in the case of an interactive computer system, a major requirement of the user community is optimization of response time which, even if it has a rather large expected value, may still be tolerable if it is fairly consistent.

Time delays have been studied in terms of their *mean values*, typically using Little's law, making the necessary independence assumptions, e.g. Refs 1 and 2. However, the mean value of a time delay alone is frequently insufficient. In the examples given above, for instance, higher moments are necessary to give the standard deviation for response time and in polling systems quantiles are also required. To date, more detailed characteristics have been investigated, in the main, by use of simulation techniques, see for example the collection of papers by Iglehart and Shedler summarized in Ref. 3. Theoretical studies of the probability distributions of time delays have tended to be limited to their Laplace transforms, and thence moments, and networks of restricted structure.⁴⁻⁶ Unfortunately analytic inversion of the Laplace transforms is possible only for very simple networks and numerical inversion is difficult as well as time consuming in view of the smoothing effect of the transformation which obscures important details; properties of the distribution's tail in particular.

Chow⁴ derives the cycle time distribution for cyclic networks of two exponential, first-come-first-served (FCFS) servers, extending the result to central server networks in Ref. 5. His analysis is based on the observation that the behaviour of the second server in a cycle, given the queue length there on arrival of some customer, is that of the server taken in isolation. An analysis in continuous time derives the probability distribution of the queue length faced at the second server, conditional on that existing initially (on arrival) at the first. The exact result follows via a complex integral expression. In Ref. 6 the Laplace transform is derived for cycle time distribution in cyclic networks of any number of exponential, FCFS servers.

In this paper we present an *exact* analysis of the *cycle time distribution* for customers in a class of networks with cyclic properties. This class is chosen so that at all stages in the computations involved, whatever the state of the network, the position of some *test customer* is known. In particular, given the initial distribution of customers among the service centres, and the initial position of the test customer, this is true of *cyclic* queueing networks if all servers have FCFS queueing discipline.

Thus cyclic networks are considered first in the next section, and the analysis is generalized to the tree-like class of networks.

In Section 3, a solution is derived for the Laplace transform of cycle time distribution; its mathematical equivalent, but of limited practical value. Results are then obtained for the moments of the distribution. A uniformly convergent discrete form approximation may be derived in the same way as in our continuous time analysis, using generating functions in place of Laplace transforms, the details appearing elsewhere.⁷

Applications of the analysis are considered in Section 4, and some numerical examples are given which show how the distribution, its mean and variance change as a network's population increases, and how these quantities can be used as standards against which to assess alternative, inexact models.

2. CYCLE TIMES AND THE EQUIVALENT OPEN NETWORK

2.1 Cyclic networks

The first step in the analysis is to consider the corresponding *tandem* network consisting of the servers in the same sequence, but with the last no longer connected to the first. There are no external arrivals, and departures from the network occur at the last service centre; all states are, then, transient. Of course the corresponding tandem network is not unique since the choice of the first server is arbitrary.

The method consists of the following steps:

1. On arrival of a test customer at server 1 in the closed network, the equilibrium probability distribution for the state space of the network is assumed. Thus the result presented by Mitrani and Sevcik⁸ can be applied. This states that the state space probability distribution seen by the arriving customer is the same as the equilibrium distribution for the same network with itself removed.
2. The corresponding open network is now considered. The cycle time in the closed network is the same as the time taken for the test customer to depart from the open network if the assumption is made that returning customers joining queues behind the test customer can have no effect on the rate of progress of the test customer through the network; i.e. departed customers can be disregarded.

In other words it must not be possible for customers to be *overtaken* by other customers, i.e. the cyclic ordering of customers must be invariant, and service rates must be unaffected by the addition of new customers to queues. This is equivalent to demanding constant service rates. Order invariance is ensured by FCFS queueing discipline at all centres together with the existence of only one path in the network. Note that PS* discipline is precluded by both the order invariance and service rates requirements.

The invariance of order implicitly tags the test customer in the open network in that it is always the leftmost (furthest from departure) and its position is therefore always known; in *any* network state. Such implicit tagging, although not possible for networks of the most general type, results in a much smaller state space than would be required in the direct analysis of the Markov process with an additional state space dimension included for the 'tagging' information, as in Ref. 3 for example. Thus it reduces computational complexity.

With the assumptions listed above, the cycle time for the test customer is identical to the time taken for the open network to enter the state with zero customers at all centres. Now, the network can empty in this way by passing through any of a (finite) number of (finite) sequences of transitions between (transient) states. Thus the cycle time distribution, *conditional* on any particular sequence, is the convolution of the distributions of the sojourn times for each state in that sequence, by the Markov property. The *unconditional* cycle time distribution is then a weighted sum of convolutions of state

sojourn time distributions, the weights being the probabilities of occurrence of the corresponding sequences of states. The formal analysis of cyclic networks proceeds on this basis, and may be extended to queueing networks of more general structure as described in the following section.

2.2 Generalization to tree-like networks

The method outlined in Section 2.1 relies primarily on the *order invariance* property of customers in the network, so allowing the position of the test customer to be known in any state and an equivalent *open* network with no arrivals to be analysed. Clearly such an approach can be applied to a much greater class of networks, although not to networks of arbitrarily interconnected service centres. In this section, cycle time distribution is considered for the class of *tree-like* networks, defined as follows.

Broadly speaking, a tree-like network is, as its name suggests, one which has branches with no loops or paths between separate branches. Formally a tree-like network, abbreviated to tree, is either

- (i) Null, or
- (ii) a tandem network (*root segment*) the last centre, *b*, of which is connected to more than one tree—the *primary subtrees* in the sense that a customer departing from centre *b* passes directly to the first centre of a primary subtree according to the network's routing probabilities.

The *leaves* of the tree are those centres which are connected to a null tree. In particular, then, a tandem network is a tree by virtue of being connected to two (or more) null primary subtrees from its last centre.

A *closed tree-like network* is one in which the leaves are all connected back to the top of the tree, i.e. on departure from a leaf centre, a customer may proceed directly to the first centre in the root segment. Thus, cyclic networks are a special case of closed tree-like networks. An example of a closed tree-like network, showing a primary sub-tree, is shown in Fig. 1.

The cycle time in a closed tree-like network is the time elapsed between successive arrivals of a customer at the first centre in the root segment. This is equivalent, assuming instantaneous passage between centres, to the time elapsed between arrival at the first root segment centre and departure from the corresponding open network.

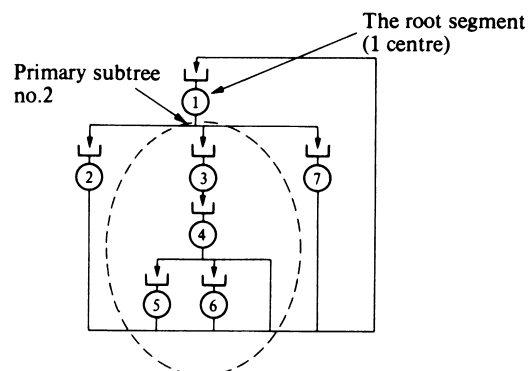


Figure 1. A closed tree-like queueing network.

* PS: Processor sharing.

The order invariance property of cyclic networks is preserved in the form of a *non-overtaking* property in tree-like networks. Given the path taken by the test customer in the network and the FCFS queueing discipline of the servers, no customer *behind* the test customer in the latter's path can subsequently enter a centre in that path *ahead* of the test customer. Thus the network can be analysed recursively as a composition of tandem networks of the type considered in Section 2.1: tagging is still implicit. Intuitively, tree-like networks are the most general that can be solved by this method; otherwise the non-overtaking property will not hold. This is proved to be the case in Ref. 9, yielding a formal definition of tree-like networks.

In the following section the notation used in the analysis is defined and the basic recursive result for the cycle time distribution is given in terms of convolutions of the distributions of the time intervals required by the test customer to pass through each (linear) segment of the tree. From this result, expressions for the Laplace transform, and so the moments of the cycle time distribution, are given in Section 3.3.

Although the required result is a mixture of convolutions of negative exponential distributions, the problem is to determine the weights for the components in this mixture, and to keep track of the parameters of the constituent exponential functions. It is shown in the following section that these all derive from the inverse of the matrix which is the difference between the identity and a trivial generalization of the state transition matrices, $(I - T^*)^{-1}$. Because of the lower triangular form of T^* , this may be computed by a simple algorithm. Despite its limitations for direct practical use, the Laplace transform is the principal result; not only academically, but also as the function from which the moments follow exactly by a straightforward algorithm, and as the basis for a parallel argument which yields a discrete form approximation to the result.

In any exact analysis of the distribution of the time elapsed between a network's entering specified states, every possible sequence of intermediate states must be considered, each making its own additive contribution. Aggregation leads to some form of averaging, and a result which must then be approximate, even if adequate for use in practice. This is in contrast to the case of solving for marginal queue length probabilities, which are by definition aggregates, relating neither to individual customers nor states.

The resulting computational complexity limits the domain of numerical application to some extent, but certainly permits validation of alternative model types, such as simulation and approximate analytic formulae. In the case of the latter, the procedure is clear, a precise error being computable in any given test. For simulation, the validation process is a little less simple. Simulation never yields precise predictions, only estimates for the quantities of interest, given along with estimated error ranges and their associated levels of confidence. Thus it is necessary to validate not only the model's design and implementation as above, but also the inherent imprecision. For any specified confidence level, the width of the error range depends on the size of the simulation sample, i.e. on the number of simulated events, and so on the length of the simulation run. It is therefore necessary to establish confidence in a simulation with respect to

known results according to appropriate statistical tests. The present analysis provides just such known results, or *standards*. Note, incidentally, that simulation estimates all require quite restricting underlying assumptions, and the confidence band widths are typically inversely proportional to the square root of the simulated run time, rendering simulation highly inefficient as a predictive model.

3. DERIVATION OF THE MATHEMATICAL FORMULAE

3.1 Notation

Consider a tree-like network, A , with r primary subtrees labelled (arbitrarily) A_1, A_2, \dots, A_r ($r \geq 0$). If $r = 0$, there are no non-null primary subtrees and the network is tandem.

Let A consist of a total of M centres and have root segment B consisting of b centres: B_1, B_2, \dots, B_b . Let the j th primary sub-tree of A have M_j centres ($1 \leq j \leq r$). Centre $c \in A$ is numbered $m(c)$ defined as

$$m(c) = \begin{cases} i & \text{if } c = B_i (1 \leq i \leq b) \\ m_s(c) + b + \sum_{j=1}^{s-1} M_j & \text{if } c \in A_s (1 \leq s \leq r) \\ & \text{where } M_j \text{ is the} \\ & \text{number of centres} \\ & \text{in subtree } A_j (1 \leq j \leq r). \end{cases}$$

where centre $c \in A_s$ is numbered $m_s(c)$ with respect to A_s .

The mapping m between the set of servers in A and $\{1, 2, \dots, M\}$ is easily seen to be a 1-1 correspondence.

The network A is assumed to be of the Jackson-Gordon-Newell type,^{10,11} all the servers having FCFS queueing discipline and service times with negative exponential probability distributions.

Let the state space of A under this centre enumeration and for a population of N customers be denoted by S and given by

$$S = \left\{ \mathbf{n} \mid \sum_{i=1}^M n_i \leq N; n_i \geq 0, 1 \leq i \leq M \right\}$$

where if $\mathbf{n} \in S$, n_i is the number of customers at the i th numbered centre, $1 \leq i \leq M$.

Let the set of valid initial states be denoted by $S^{(1)}$ and defined by

$$S^{(1)} = \left\{ \mathbf{n} \mid \mathbf{n} \in S; \sum_{i=1}^M n_i = N; n_1 > 0 \right\}$$

which represents a state with a total of N customers and the test customer at (the back of the queue of) the first centre in the root segment of A .

In order to proceed with the recursive analysis, it is also necessary to define one more subset of states, namely those which can introduce the test customer into a primary subtree after one state transition. Define $E \subseteq S$ by

$$E = \{ \mathbf{n} \mid \mathbf{n} \in S; n_i = 0, 1 \leq i < b; n_b = 1 \}$$

Hence E consists of states with only one customer left in the root segment, B , of A ; at its last centre. Because of the FCFS queueing discipline, this must be the test customer.

Let the service rate of centre i be μ_i ($1 \leq i \leq M$), a constant for the reasons explained in Section 2, and define θ, ϕ, λ by

$\theta(u, v)$ = number of centre from which a departure causes a transition $u \rightarrow v$ ($u, v \in S$)

$\phi(u, v)$ = number of centre at which a customer arrives on a transition $u \rightarrow v$ ($u, v \in S$)

where $\theta(u, v), \phi(u, v)$ are undefined if a one-step transition $u \rightarrow v$ is invalid,

$$\lambda(u) = \sum_{\substack{1 \leq j \leq M \\ u_j > 0}} \mu_j, \text{ the total service rate in state } u \in S.$$

The state transition matrix for the embedded Markov chain, T , may be derived from the instantaneous transition rate matrix or the balance equations for A as

$$T_{uv} = \begin{cases} \frac{\mu_{\theta(u,v)} P_{\theta(u,v), \phi(u,v)}}{\lambda(u)} & \text{if a one-step transition } u \rightarrow v \text{ is valid} \\ 0 & \text{otherwise} \end{cases}$$

where p is the routing probability matrix of A , so that for a transition, $u \rightarrow v$, caused by a customer moving *within* a segment, the factor would be absent in the expression for T_{uv} .

By the exponential assumption, the time spent in state $u \in S$ also has exponential distribution, F_u say, with mean $\{\lambda(u)\}^{-1}$ and so has Laplace transform

$$D_u(s) = \frac{\lambda(u)}{s + \lambda(u)}$$

The modified transition matrix, T^* , is defined by $T_{uv}^* = T_{uv} D_u$, i.e.

$$T_{uv}^* = \begin{cases} \frac{\mu_{\theta(u,v)} P_{\theta(u,v), \phi(u,v)}}{s + \lambda(u)} & \text{if a one-step transition } u \rightarrow v \text{ is valid} \\ 0 & \text{otherwise} \end{cases}$$

Let the probability distribution function for the time to pass through A on some stochastically chosen path, conditional on initial state $\alpha \in S^{(1)}$, be $G(t|\alpha)$ with the unconditional distribution function for an initial equilibrium state distribution being $G(t)$. Let these distributions have Laplace transforms $L(s|\alpha)$ and $L(s)$, respectively.

Let the random variable for the time taken for the network A to reach state β from state α ($\alpha, \beta \in S$) be denoted by $\tau_{\alpha\beta}$ and define

$$H_{\alpha\beta}(t) = \Pr(\tau_{\alpha\beta} \leq t) \quad (\alpha \neq \beta)$$

$$H_{\alpha\alpha}(t) = 1$$

3.2 The basic recurrence relation

Proposition 1. The conditional cycle time distribution, $G(t|\alpha)$, in a tree-like queueing network, A , with b root segment centres, r primary subtrees, N customers initially

and start state $\alpha \in S^{(1)}$ in which the test customer is at the first numbered centre of A , is given by

$$G(t|\alpha) = \begin{cases} H_{\alpha 0} & (r = 0) \\ \sum_{\beta \in E} \sum_{j=1}^r \frac{P_j \mu_b}{\lambda(\beta)} G_j(t|\beta^{(j)}) \oplus F_\beta \oplus H_{\alpha\beta} & \text{otherwise} \end{cases}$$

where

$G_j(t|\gamma)$ is the cycle time distribution for subtree A_j conditional on initial state γ in the state space of A_j , state 0 is that representing zero customers at all centres $P_j = p_{bK_j}$, the routing probability, in which

$$K_j = 1 + b + \sum_{i=1}^{j-1} M_i \quad (1 \leq j \leq r \neq 0)$$

$$\beta_i^{(j)} = \beta_{i+K_j-1}^{(j)} \quad (1 \leq i \leq M_j)$$

$\beta^{(j)} \in S$ is the state succeeding $\beta \in S$ entered by transit of a customer from centre b to the first centre in subtree A_j ($1 \leq j \leq r$). $\beta^{(j)}$ is the same as $\beta^{(j)}$ with its components not corresponding to centres in subtree A_j removed.

Proof. Let the random variable for the time taken for the test customer to leave the network from state $\beta \in S$ be denoted by Y_β . Also, let δ_β denote the random variable for the network's sojourn time in state β .

Then, for $\alpha \in S^{(1)}$ and $\alpha \neq E$, $r \neq 0$,

$$G(t|\alpha) = \sum_{\beta \in E} \int_0^t \Pr \left[\begin{array}{l} Y_\beta \leq t - u \text{ and transition from} \\ \text{state } \beta \text{ caused by test customer} \\ \text{service completion} \end{array} \right] \times d\Pr(\tau_{\alpha\beta} \leq u|\beta)$$

since for any path in A taken by the test customer, some state $\beta \in E$ must be entered just before passage of the test customer to a primary sub-tree, and using the fact that $\tau_{\alpha\beta}$ is a Markov time. Thus,

$$G(t|\alpha) = \sum_{\beta \in E} \int_0^t \left\{ \sum_{j=1}^r \Pr \left[\begin{array}{l} Y_\beta \leq t - u \text{ and test customer} \\ \text{enters primary subtree} \\ j \text{ after state } \beta \end{array} \right] \right\} \times d\Pr(\tau_{\alpha\beta} \leq u|\beta)$$

Now, the probability distribution of δ_β and the probability that the next state transition is $\beta \rightarrow \beta^{(j)}$ are independent by the Markov property, so the expression $\{ \}$ may be written as

$$\sum_{j=1}^r \frac{P_j \mu_b}{\lambda(\beta)} \int_0^{t-u} G(t-u-v|\beta^{(j)}) \quad d\Pr(\delta_\beta \leq v)$$

state transition probability $\beta \rightarrow \beta^{(j)}$ distribution of time from state $\beta^{(j)}$ to end of path distribution of time spent in state β

where $\beta^{(j)}$ is, as defined in the proposition statement, the state succeeding β caused by transit of a customer from centre b (last in the root segment) to the first centre in the j th primary subtree's root segment, numbered K_j say; $P_j = p_{bK_j}$, the associated routing probability. By definition,

$$K_j = 1 + b + \sum_{i=1}^{j-1} M_i \quad (1 \leq j \leq r)$$

Thus the expression $\{ \}$ is

$$\sum_{j=1}^r \frac{P_j \mu_b}{\lambda(\beta)} [G_j(\beta^{(j)}) \oplus F_\beta](t - u)$$

where, as required,

$$\beta_i^{(j)} = \beta_{i+\kappa_j-1}^{(j)} \quad (1 \leq i \leq M_j)$$

are the components of the state space vector for A_j taken in isolation corresponding to state $\beta^{(j)}$ in A . The sub-trees A_j may be considered separately in this way since all transitions in other sub-trees A_k ($1 \leq k \neq j \leq r$) are stochastically independent. Thus,

$$G(t|\alpha) = \sum_{\beta \in E} \sum_{j=1}^r \frac{P_j \mu_b}{\lambda(\beta)} G_j(\beta^{(j)}) \oplus F_\beta \oplus H_{\alpha\beta}$$

for $r \neq 0$ and $\alpha \notin E$.

For $\alpha \in E$, the same reasoning and resulting equation applies using the result that

$$H_{\alpha\alpha}(t) = 1 \quad (t \geq 0, \alpha \in S)$$

For $r = 0$, A is a tandem network, and so

$$G(t|\alpha) = H_{\alpha 0}(t)$$

This completes the proof.

Note that the tree-like property of the network A allows the position of the test customer to be determined at all stages in the recursive computation, using the partition of each route (or sequence of states entered by the network) via the states $\beta \in E$. The test customer is, therefore, *implicitly tagged*.

Note further that the terms $H_{\alpha\beta}$ are the time delay distributions corresponding to the test customer's passing into service at the last centre in the root segment of the tree-like network. Thus, in order to compute any properties of cycle time distribution, the corresponding properties of $H_{\alpha\beta}$ must first be derived. The analysis of the following sections proceeds thus.

3.3 The Laplace transform and moments

The Laplace transform is simply obtained using the result of Proposition 1 and the moments by differentiation. Corresponding results for the distribution $H_{\alpha\beta}(\alpha, \beta \in S)$ are first derived in the Lemma which follows the notation given below.

For $s, t \in S$, define

$$R_{\alpha\beta} = \{(i_1, i_2, \dots, i_n) | n \in \mathbb{Z}^+; i_j \in S, 1 \leq j \leq n; i_1 = \alpha; i_n = \beta; T_{i_k i_{k+1}} \neq 0, 1 \leq k < n\}^*$$

i.e. the set of all sequences of states entered, or routes, from state α to state β . If

$$\mathbf{i} = (i_1, i_2, \dots, i_n) \in R_{\alpha\beta}$$

then let $|\mathbf{i}| = n$, the number of steps in the route \mathbf{i} . Let

$$R'_{\alpha\beta} = \{\mathbf{r} | \mathbf{r} \in R_{\alpha\beta}; r_i \neq \beta, 1 \leq i < |\mathbf{r}|\},$$

the subset of loop-free routes. For tree-like networks, $R'_{\alpha\beta} \equiv R_{\alpha\beta}$, all states being transient. A formal proof of this statement is given in Appendix A.

Let $H_{\alpha\beta}(t)$ have Laplace transform $L_{\alpha\beta}(s)$.

* \mathbb{Z}^+ is the set of positive integers.

Lemma 1.

$$L_{\alpha\beta}(s) = (I - T^*)_{\alpha\beta}^{-1}$$

where I is the unit matrix.

Proof.

$$H_{\alpha\beta}(t) = \sum_{\mathbf{r} \in R'_{\alpha\beta}} \Pr(\mathbf{r} | \alpha, \beta) \Pr(\tau_{\alpha\beta} \leq t | \mathbf{r})$$

since the end states α, β are implied by the route \mathbf{r} , and where

$$\Pr(\mathbf{r} | \alpha, \beta) = \Pr\{\mathbf{r} | r_1 = \alpha, r_{|\mathbf{r}|} = \beta\}.$$

Now,

$$\Pr(\tau_{\alpha\beta} \leq t | \mathbf{r}) = \Pr\left\{\sum_{i=1}^{|\mathbf{r}|-1} \delta_{r_i} \leq t\right\}$$

and

$$\Pr(\mathbf{r} | \alpha, \beta) = \prod_{i=1}^{|\mathbf{r}|-1} T_{r_i r_{i+1}}$$

Thus, by the Markov property,

$$H_{\alpha\beta}(t) = \sum_{\mathbf{r} \in R'_{\alpha\beta}} \left\{ \prod_{i=1}^{|\mathbf{r}|-1} T_{r_i r_{i+1}} \right\} \bigoplus_{j=1}^{|\mathbf{r}|-1} F_{r_j} \quad (1)$$

and so

$$L_{\alpha\beta}(s) = \sum_{\mathbf{r} \in R'_{\alpha\beta}} \prod_{i=1}^{|\mathbf{r}|-1} T_{r_i r_{i+1}} D_{r_i}(s)$$

where $L_{\gamma\gamma}(s) \equiv 1, \forall s \geq 0, \gamma \in S$.

This result applies to networks in general; in fact to any such Markov process. Rearranging and dropping the prime on $R_{\alpha\beta}$ for tree-like networks,

$$L_{\alpha\beta}(s) = \sum_{k=1}^{\infty} \sum_{\substack{\mathbf{r} \in R_{\alpha\beta} \\ |\mathbf{r}|=k}} \prod_{i=1}^{k-1} T_{r_i r_{i+1}}^*$$

For $k \geq 3$,

$$\sum_{\substack{\mathbf{r} \in R_{\alpha\beta} \\ |\mathbf{r}|=k}} \prod_{i=1}^{k-1} T_{r_i r_{i+1}}^* = \sum_{\gamma_{k-1} \in S} \left\{ \sum_{\substack{\mathbf{r}' \in R_{\alpha\gamma_{k-1}} \\ |\mathbf{r}'|=k-1}} \prod_{i=1}^{k-2} T_{r'_i r'_{i+1}}^* \right\} T_{\gamma_{k-1} \beta}^*$$

for if $\nexists \mathbf{r} \in R_{\alpha\beta}$ with $r_{k-1} = \gamma_{k-1}$ then either no one-step transition $\gamma_{k-1} \rightarrow \beta$ exists so that $T_{\gamma_{k-1} \beta}^* = 0$, or $R_{\alpha\gamma_{k-1}} = \emptyset$.†

By a simple inductive argument, this yields the result

$$\sum_{\gamma_2 \in S} \dots \sum_{\gamma_{k-2} \in S} \sum_{\gamma_{k-1} \in S} T_{\alpha\gamma_2}^* T_{\gamma_2\gamma_3}^* \dots T_{\gamma_{k-2}\gamma_{k-1}}^* T_{\gamma_{k-1}\beta}^* = \{T^*\}_{\alpha\beta}^{k-1}$$

and so

$$L_{\alpha\beta}(s) = \sum_{k=1}^{\infty} \{T^*\}_{\alpha\beta}^{k-1}$$

since $L_{\alpha\alpha}(s) \equiv 1$ and $L_{\alpha\beta}(s) = T_{\alpha\beta}^*$ for a one-step transition $\alpha \rightarrow \beta$ by definition. Thus,

$$L_{\alpha\beta}(s) = [I - T^*]_{\alpha\beta}^{-1}$$

† \emptyset is the empty set.

since T is a stochastic matrix and for $s > 0$, $D_u(s) < 1 \forall u \in S$, so the series converges. In fact here $\exists k \in \mathbb{Z}^+$ s.t. $(T^*)^k = 0$, since ultimately the network has no customers and can have no transitions.

A simpler, but non-constructive, proof which avoids the need to consider the set $R_{\alpha\beta}$ is as follows.

For $\alpha \neq \beta$,

$$L_{\alpha\beta} = \sum_{\gamma \in S} T_{\alpha\gamma}^* L_{\gamma\beta}$$

Now, $T_{\alpha\gamma}^* L_{\gamma\alpha} = 0$ for all $\alpha, \gamma \in S$ since if there is a route $\alpha \rightarrow \gamma$ there is no route $\gamma \rightarrow \alpha$, otherwise α and γ would be recurrent states, contradicting the fact that all states are transient in tree-like networks. Thus we may write

$$L_{\alpha\alpha} = 1 + \sum_{\gamma \in S} T_{\alpha\gamma}^* L_{\gamma\alpha}$$

the summation term yielding zero contribution. Hence,

$$L = I + T^* \cdot L$$

as required.

Theorem 1.

$$L(s|\alpha) = \begin{cases} [I - T^*]_{\alpha\alpha}^{-1} & (r = 0) \\ \sum_{\beta \in E} \sum_{j=1}^r \frac{P_j \mu_b}{s + \lambda(\beta)} [I - T^*]_{\alpha\beta}^{-1} L_j(s|\beta^{(j)}) & \text{otherwise} \end{cases}$$

where $L_j(s|\gamma)$ is the Laplace transform of $G_j(t|\gamma)$, $1 \leq j \leq r$.

Assuming an initial equilibrium state space probability distribution $P(\alpha)$, $\alpha \in S^{(I)}$,

$$L(s) = \sum_{\alpha \in S^{(I)}} P(\alpha) L(s|\alpha)$$

The proof follows directly from Proposition 1 and Lemma 1, using the fact that F_β has Laplace transform

$$D_\beta(s) = \lambda(\beta) \{s + \lambda(\beta)\}^{-1}$$

By the result in Ref. 8, for $\mathbf{n} \in S^{(I)}$,

$$P(\mathbf{n}) = \frac{\mu_1 \prod_{i=1}^M \{e_i/\mu_i\}^{n_i}}{e_i G(N-1)}$$

where $G(N-1)$ is the normalizing constant for the closed network, A , with one customer removed and $\{e_i | 1 \leq i \leq M\}$ is such that

$$e_i = \sum_{j=1}^M e_j p_{ji} \quad (1 \leq i \leq M)$$

Since all states are transient, by suitable enumeration, with *centres* numbered as defined in Section 3.2, the transition matrix T , is lower triangular.⁹ Thus $(I - T^*)^{-1}$ may be computed by a simple back substitution process. Moreover, not every column of the inverted matrix is required; only

$$\{(I - T^*)_{\alpha\beta}^{-1} | \beta \in E\}$$

Indeed, for *cyclic* networks only the *first* column is required. As a result, both numerical precision and computational efficiency are enhanced.

The moments of cycle time distribution may be derived by differentiation of the Laplace transform as follows.

Corollary. Let the p th moment of cycle time distribution for the tree-like network, A , of N customers and r primary sub-trees, conditional on start state $\alpha \in S'$ be denoted by $M(p|\alpha)$ where

$$S' = \bigcup_{n=1}^N S_n^{(I)}$$

in which S_n is the state space for network A having population n .

Let

$$X^{(p)} = \sum_{\substack{\sum_{i=1}^{|\mathbf{m}|} m_i = p \\ m_i > 0}} \left\{ \prod_{i=1}^{|\mathbf{m}|} (I - T)^{-1} T'(m_i) \right\} (I - T)^{-1} \quad (p > 0)$$

$$X^{(0)} = I$$

where $|\mathbf{m}|$ is the number of components in the vector \mathbf{m} and

$$T'_{\alpha\beta}(m_i) = \frac{T_{\alpha\beta}}{\{\lambda(\alpha)\}^{m_i}} \quad (\alpha, \beta \in S)$$

Then, for $\alpha \in S'$,

$$M(p|\alpha) = \begin{cases} p! X_{\alpha\alpha}^{(p)} & (r = 0) \\ p! \sum_{\beta \in E} \sum_{j=1}^r P_j \mu_b \sum_{\substack{u+v+w=p \\ u, v, w \geq 0}} X_{\alpha\beta}^{(u)} \{\lambda(\beta)\}^{-v-1} \\ \quad \times M_j(w|\beta^{(j)}) \{w!\}^{-1} & (r > 0) \end{cases}$$

where M_j is the moment function for the primary subtree A_j .

The unconditional p th moment of the cycle time distribution is then

$$M(p) = \sum_{\alpha \in S^{(I)}} P(\alpha) M(p|\alpha)$$

Proof.

$$M(p|\alpha) = (-)^p \left[\frac{d^p}{ds^p} L(s|\alpha) \right]_{s=0} \quad (\alpha \in S')$$

For $r = 0$,

$$M(p|\alpha) = (-)^p \left[\frac{d^p}{ds^p} (I - T^*)_{\alpha\alpha}^{-1} \right]_{s=0} = p! X_{\alpha\alpha}^{(p)}$$

by the result derived in Appendix B which yields

$$\left[\frac{d^m T_{\alpha\beta}^*}{ds^m} \right]_{s=0} = (-)^m m! T'_{\alpha\beta}(m) \quad (\alpha, \beta \in S)$$

For $r > 0$, by Theorem 1,

$$\begin{aligned} M(p|\alpha) &= (-)^p \sum_{\beta \in E} \sum_{j=1}^r P_j \mu_b \\ &\quad \times \left[\frac{d^p}{ds^p} \left\{ (I - T^*)_{\alpha\beta}^{-1} \frac{1}{s + \lambda(\beta)} L_j(s|\beta^{(j)}) \right\} \right]_{s=0} \\ &= (-)^p \sum_{\beta \in E} \sum_{j=1}^r P_j \mu_b p! \\ &\quad \times \sum_{\substack{u+v+w=p \\ u, v, w \geq 0}} \frac{F_{\alpha\beta}^{(u)}(0) (-)^v L_j^{(w)}(0|\beta^{(j)})}{u! \{\lambda(\beta)\}^{v+1} w!} \end{aligned}$$

where the n th derivative of a function of s , $F(s)$ say, is denoted by $F^{(n)}(s)$ and Leibnitz's theorem for repeated differentiation of a product has been used. Here, $F = (I - T^*)^{-1}$ so that the result follows as for the case $r = 0$.

An algorithm to compute moments from these formulae is illustrated by the explicit expansion of the following terms

$$\begin{aligned} X^{(1)} &= (I - T)^{-1} T'(1) (I - T)^{-1} \\ X^{(2)} &= (I - T)^{-1} T'(2) (I - T)^{-1} \\ &\quad + (I - T)^{-1} T'(1) (I - T)^{-1} T'(1) (I - T)^{-1} \end{aligned}$$

and for $r > 0$

$$\begin{aligned} M(2|\alpha) &= 2 \sum_{\beta \in E} \sum_{j=1}^r \frac{P_j \mu_b}{\lambda(\beta)} \{ X_{\alpha\beta}^{(2)} \\ &\quad + 1/\lambda(\beta)^2 + \frac{1}{2} M_j(2|\beta^{(j)}) + X_{\alpha\beta}^{(1)} M_j(1|\beta^{(j)}) \\ &\quad + M_j(1|\beta^{(j)})/\lambda(\beta) + X_{\alpha\beta}^{(1)}/\lambda(\beta) \} \end{aligned}$$

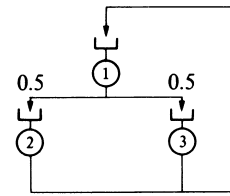
In theory, then, any number of moments of the cycle time distribution may be computed. But in practice, unoptimized recursion may be excessively inefficient, since the moments for the subtrees of A may be recomputed many times, and many large associated data structures must be stored simultaneously. However, for the first two moments, the computations involved are not prohibitive for networks with simple structure. Typically no higher moments will be required, particularly if, for example, the central limit theorem is to be applied for prediction of response time (an aggregation of cycle times¹²) distribution, under suitable independence assumptions.

4. NUMERICAL RESULTS

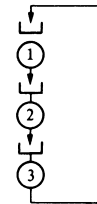
The formulae derived in the previous section for the moments of cycle time distribution are simply expressed as algorithms and programmed for execution on a computer. Moreover, a similar approach may be used to derive an approximate discrete form for the actual distribution of cycle time.⁷ The negative exponential state sojourn times are first expressed in discrete form by geometric distributions. Their convolutions, corresponding to test customer routes through the network, are determined using Z-transforms (corresponding to the use of Laplace transforms in the continuous time domain), and numerical computation may then be performed directly as simple summations. Approximations are introduced at an early stage in this analysis, and expressions are subsequently manipulated. Thus an error analysis becomes essential, and it is shown in Ref. 7 that the result is uniformly convergent to the exact cycle time distribution as the time axis mesh size approaches zero.

The choice of network structure results in an implementation which is efficient compared to the 'brute force' approach involving tagging a customer explicitly and working with holding time distributions in an extended Markovian state space.¹³ However, the algorithms are still very complex computationally and permit only relatively small networks to be analysed—the largest in our selection of tests has seven servers and four customers. This applies particularly to the representation of the distribution in discrete form as a histogram, especially when a small mesh size is required. The mesh used in our

A Simplest non-cyclic case



B Cyclic



C More complex

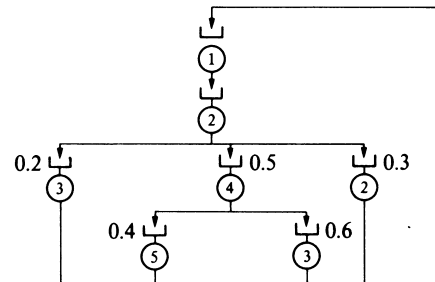


Figure 2. Test network specifications.

examples has size equal to one tenth of the associated mean value.

From a series of tests, the following three are presented to show the way in which the distribution of cycle time changes as a network's parameters vary and to illustrate how the accuracy of an approximate method may be assessed by comparison with standards.

The networks are all Markovian, closed, with FCFS exponential servers, and defined in Fig. 2. The service rates and routing probabilities are shown in the circles and by the arrows, respectively. The mean and standard deviation of cycle time distribution were calculated for a customer population of four in each case, with the results for population two and six also given for case A (Table 1). The distributions were also computed in histogram form by the algorithm given in Ref. 7, and may be seen in Figs 3–5. Thus, from Fig. 3 and Table 1, the effect on cycle time distribution of changing the number of customers in network A can be seen. Similar results were also obtained by the approximate method of Ref. 12 and are included in Table 1; the corresponding histograms appear with the exact plots in Figs 4 and 5. Accuracy of the approximate method may be assessed for these test cases via its predictions for standard deviation and

Table 1. Summary statistics for test cases

Case	Popu- lation	Mean	Standard deviation (exact) (approx)		Percentage error	Maximum difference	Per- centile
A	2	2.186	1.451	1.488	2.6	0.030	54
A	4	4.026	2.035	2.115	3.9	0.033	62
A	6	6.003	2.491	2.574	3.4	0.013	77
B	4	4.245	1.962	2.187	11.5	0.041	14
C	4	4.246	1.995	2.171	8.8	0.066	26

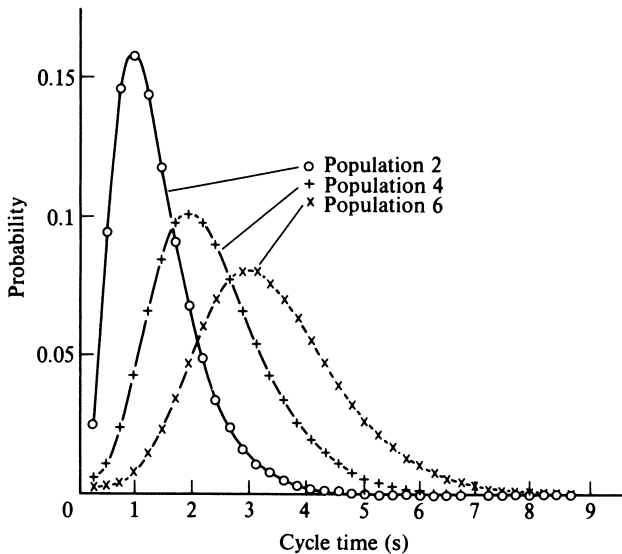


Figure 3. Cycle time distribution for network A with 2, 4, 6 customers.

percentiles. It should be noted here that this particular approximation yields exact mean cycle times. Precision is therefore indicated in Table 1 by the percentage error in the standard deviation and by the maximum difference between the cumulative discrete distributions (along with the corresponding quantile), and accuracy can also be judged by inspection of the Figures.

It can be seen, particularly from Fig. 3, how the distribution becomes more Normal as the numbers of customers in the networks increase—corresponding to a greater number of convolutions and better validity of the central limit theorem's assumptions. Moreover, for smaller populations, the shapes of the curves are characteristic of lower-order Erlang density functions, corresponding to (mixtures of) small numbers of convolutions of negative exponentials. It may also be seen (Figs 4, 5) that the approximate method yields a 'flatter' density function, with correspondingly larger variance (Table 1). This results from a greater degree of averaging, arising from the assumption of independence between servers which underlies the approximation in this particular model. The effect is less pronounced for the more complex network C (Fig. 5) in which less correlation between customers is to be expected than in, say, cyclic networks in which their ordering is invariant under FCFS queueing discipline.

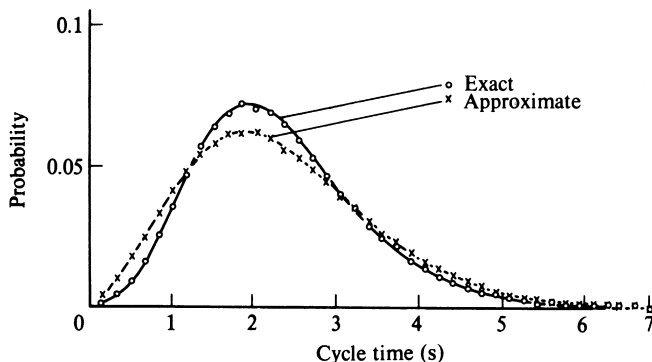


Figure 4. Exact and approximate histograms for cycle time distribution in network B with 4 customers.

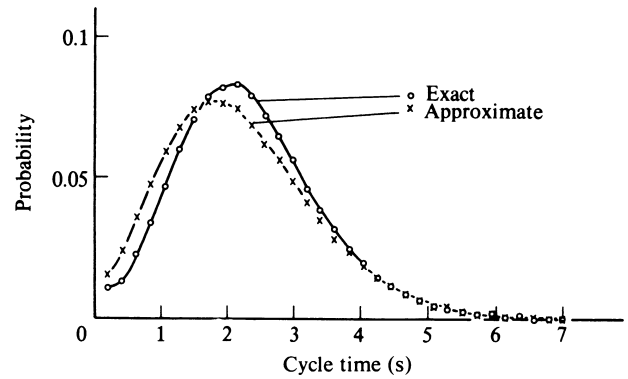


Figure 5. Exact and approximate histograms for cycle time distribution in network C with 4 customers.

5. CONCLUSION

The probability distribution of cycle times has been derived for the tree-like class of queueing networks, in terms of its Laplace transform, its moments and as a uniformly convergent discrete-time representation. The importance of this result lies in the modelling and prediction of time delay characteristics in a variety of physical systems such as communications and computer networks. From the moments, predictions for response time distribution may also be made. Response time is represented by the sum of several successive cycle times for some customer. If these constituent cycle times are assumed to be independently distributed, as indicated in Ref. 9 for example, the central limit theorem may be applied so that the response time distribution, conditional on the number of cycles, is asymptotically Normal. Thus only the first two moments of cycle time distribution are needed, simplifying the computations required. Perhaps more importantly, from the histogram representation, percentiles may be computed, allowing for example, the probability of system failure to be predicted, rather than merely an aggregate description in terms of mean time between failures and perhaps its standard deviation. By the convergence property, the precision of any such estimate is, in theory, arbitrary (over the whole of the distribution in view of the uniformity), although in practice limited by the computing resources available to handle sufficiently small mesh sizes.

The computations involved in the implementation of the formulae derived here are undeniably complex. As a result, the main practical use of the theory is its application to simple network structures to provide standards by which to assess the precision of alternative, inexact solutions, such as simulation or approximate formulae. The tests of section 4 illustrate the methodology for such comparisons, and demonstrate its viability. The ultimate goal of this research direction would be to create performance tools for practical use from the more efficient, but approximate models, the accuracy of which would be adequate according to systematic validation.

Acknowledgements

The author wishes to thank I P Sharp Associates Limited for making available their APL system for the implementation of the formulae from which all the numerical results were obtained. Thanks are also due to the referee whose review has resulted in a much better structured, more readable and relevant paper.

APPENDIX A

Proof of non-existence of loops in tree-like networks

In the notation of Section 3.3: for $\alpha, \beta \in S$,

$$R_{\alpha\beta} = R'_{\alpha\beta}$$

Proof

1. Trivial for $\beta = 0$, the state with zero customers.
2. Otherwise, suppose the result is false and assume the centre enumeration given in Section 3.1.

Then if a customer passes from centre i to centre j , $i < j$, $\exists r \in R_{\beta\beta}$ with $|r| > 1$ and integers i, k s.t.

$$\begin{aligned} r_{2,n} &= \beta_n \quad (n \neq i, k) \\ r_{2,i} &= \beta_i - 1 \\ r_{2,k} &= \beta_k + 1 \end{aligned}$$

i.e. the state transition $\beta \rightarrow r_2$ is caused by passage of a customer from centre i to centre k . Choose r with minimum such i .

Now, since $r \in R_{\beta\beta}$, \exists integers $n > 2, j < i$ s.t.

$$\begin{aligned} r_{n,i} &= r_{2,i} \\ r_{n,j} &> 0 \\ r_{n+1,i} &= r_{n,i} + 1 = \beta_i \\ r_{n+1,j} &= r_{n,j} - 1 \end{aligned}$$

i.e. the state transition $r_n \rightarrow r_{n+1}$ is caused by passage of a customer from centre j to centre i . Suppose first that $\beta_j = r_{1,j} > 0$ and consider the sequence of states, r^* , defined by

$$\begin{aligned} r_1^* &= r_1 \\ r_{m,l}^* &= r_{m,l} \quad (l \neq i, j, k) \end{aligned}$$

$$\begin{aligned} r_{m,i}^* &= r_{m,i} + 2 \\ r_{m,j}^* &= r_{m,j} - 1 \\ r_{m,k}^* &= r_{m,k} - 1 \end{aligned}$$

where $2 \leq m \leq n$, so that, in particular, the transition $r_1^* \rightarrow r_2^*$ is caused by passage of a customer from centre j to centre i , subsequent transitions being caused by the same passages as in route r ;

$$\begin{aligned} r_{n+1,l}^* &= r_{n,l}^* \quad (l \neq i, k) \\ r_{n+1,i}^* &= r_{n,i}^* - 1 \\ r_{n+1,k}^* &= r_{n,k}^* + 1 \end{aligned}$$

so that the transition $r_n^* \rightarrow r_{n+1}^*$ is caused by passage of a customer from centre i to centre k . Thus $r_{n+1}^* = r_{n+1}$ and

$$r_m^* = r_m \quad (n+1 \leq m \leq |r|)$$

Thus, $r^* \in R_{\beta\beta}$ contradicting the definition of i since $j < i$.

If $\beta_j = 0$, \exists integers $h < j$ and $p < n$ such that the transition $r_p \rightarrow r_{p+1}$ is caused by passage of a customer from centre h to centre j . Assume $\beta_h > 0$ and consider the sequence of states r^{**} defined by

$$\left. \begin{aligned} r_1^{**} &= r_1^* \\ r_{m,l}^{**} &= r_{m,l}^* \quad (l \neq i, h, j) \\ r_{m,i}^{**} &= r_{m,i}^* - 1 \\ r_{m,h}^{**} &= r_{m,h}^* - 1 \\ r_{m,j}^{**} &= r_{m,j}^* + 2 \\ r_m^{**} &= r_m^* \quad (p+1 \leq m \leq |r|) \end{aligned} \right\} \quad 2 \leq m \leq p$$

Then $r^{**} \in R_{\beta\beta}$, again contradicting the definition of i since $h < i$. A simple inductive argument completes the proof for the case $\beta_h = 0$.

APPENDIX B

Multiple differentiation of a (weighted) sum of products

Let $F(s) = [I - A(s)]^{-1}$. Then,

$$F^{(p)}(s) \equiv \frac{d^p}{ds^p} F(s) = p! \sum_{\substack{\sum_{i=1}^{|m|} m_i = p \\ m_j > 0 \\ 1 \leq j \leq |m|}} \left\{ \prod_{i=1}^{|m|} \frac{(I - A)^{-1} A^{(m_i)}}{m_i!} \right\} \times (I - A)^{-1}$$

where $A^{(m)} \equiv d^m A / ds^m$ and $|m|$ is the number of components in m so that $|m| \leq p$.

Proof

The proof is by induction on p .

For the case $p = 1$, the formula gives

$$\begin{aligned} F^{(1)}(s) &= \sum_{m_1=1}^1 1! \prod_{i=1}^1 (I - A)^{-1} A^{(1)} (I - A)^{-1} \\ &= (I - A)^{-1} A^{(1)} (I - A)^{-1} \end{aligned}$$

which is true, since for matrix $M(s)$,

$$\frac{d}{ds} \{M(s)\}^{-1} = -M^{-1} \frac{dM}{ds} M^{-1}$$

Now assume the result is true for derivatives up to the p th.

$$\begin{aligned} F^{(p+1)}(s) &= \frac{d}{ds} F^{(p)}(s) \\ &= X + Y \text{ say} \end{aligned}$$

where $X =$

$$\sum_{\substack{\sum_{i=1}^{|m|} m_i = p \\ \& m_j > 0}} p! \sum_{j=1}^{|m|} \left\{ \prod_{i=1}^{j-1} \frac{(I - A)^{-1} A^{(m_i)}}{m_i!} \right\} \frac{(I - A)^{-1} A^{(m_j+1)}}{m_j!} \times \left\{ \prod_{i=j+1}^{|m|} \frac{(I - A)^{-1} A^{(m_i)}}{m_i!} \right\} (I - A)^{-1}$$

and $Y =$

$$\sum_{\substack{\sum_{i=1}^{|m|} m_i = p \\ \& m_j > 0}} p! \sum_{j=1}^{|m|+1} \left\{ \prod_{i=1}^{j-1} \frac{(I - A)^{-1} A^{(m_i)}}{m_i!} \right\} \frac{(I - A)^{-1} A^{(1)}}{1!} \times \left\{ \prod_{i=j}^{|m|} \frac{(I - A)^{-1} A^{(m_i)}}{m_i!} \right\} (I - A)^{-1}$$

with slightly abbreviated notation.

First consider X and for each m, j define \mathbf{n} by

$$\begin{aligned} n_i &= m_i \quad (1 \leq i \neq j \leq |m|) \\ n_j &= m_j + 1 \end{aligned}$$

Then,

$$X = \sum_{\substack{\sum n_i = p+1 \\ \& n_i > 0}} p! \sum_{\substack{j=1 \\ n_j > 1}}^{|n|} \left\{ n_j \prod_{i=1}^{|n|} \frac{(I-A)^{-1} A^{(n_i)}}{n_i!} \right\} (I-A)^{-1}$$

Now define $k^{(j)}$ for each m, j by

$$\begin{aligned} k_i^{(j)} &= m_i \quad (1 \leq i < j) \\ k_j^{(j)} &= 1 \\ k_i^{(j)} &= m_{i-1} \quad (j < i \leq |m| + 1) \end{aligned}$$

Then,

$$Y = \sum_{\substack{\sum k_i = p+1 \\ \& k_i > 0}} p! \sum_{\substack{j=1 \\ k_j > 1}}^{|n|} \left\{ k_j \prod_{i=1}^{|n|} \frac{(I-A)^{-1} A^{(k_i)}}{k_i!} \right\} (I-A)^{-1}$$

Thus, relabelling k_i by n_i in the expression for Y ,

$$X + Y = \sum_{\substack{\sum n_i = p+1 \\ \& n_i > 0}} p! \sum_{j=1}^{|n|} \left\{ n_j \prod_{i=1}^{|n|} \frac{(I-A)^{-1} A^{(n_i)}}{n_i!} \right\} (I-A)^{-1}$$

But, $\sum_{j=1}^{|n|} n_j = p+1$, so

$$\begin{aligned} F^{(p+1)}(s) &= X + Y \\ &= \sum_{\substack{\sum n_i = p+1 \\ \& n_i > 0}} (p+1)! \left\{ \prod_{i=1}^{|n|} \frac{(I-A)^{-1} A^{(n_i)}}{n_i!} \right\} (I-A)^{-1} \end{aligned}$$

as required.

REFERENCES

1. M. Reiser, Mean value analysis of queueing networks, a new look at an old problem. *4th Int. Symp. on Modelling and Performance Evaluation of Computer Systems*, Vienna, February (1979).
2. Y. Bard, Some extensions to multiclass queueing network analysis. *4th Int. Symp. on Modelling and Performance Evaluation of Computer Systems*, Vienna, February (1979).
3. D. L. Iglehart and G. S. Shedler, *Regenerative Simulation of Response Times in Networks of Queues*. Lecture Notes in control and Information Sciences, Springer-Verlag (1980).
4. W. M. Chow, The cycle time distribution of exponential cyclic queues. *JACM* **27** (2), 281–286 (1980).
5. W. M. Chow, The cycle time distribution of exponential central server queues. *IBM Research Report, RC 6765* (1977).
6. R. Schassberger and H. Daduna, The time for a round-trip in a cycle of exponential queues. *Proc. ORSA-TIMS Conference on Applied Probability—Computer Science, The Interface*, Florida—Atlantic University, Boca Raton, Florida (1981).
7. P. G. Harrison, A Convergent discrete form for time delay distributions in networks of queues. *Research report DOC 82/2*, Dept. of Computing, Imperial College, London; *Computer Performance* **3** (3), 167–175 (1982).
8. I. Mitrani and K. C. Sevcik, The distribution of queueing network states at input and output instants. *4th Int. Symp. on Modelling and Performance Evaluation of Computer Systems*, Vienna, February (1979).
9. P. G. Harrison, Representative queueing network models of computer systems in terms of time delay probability distributions. *PhD Thesis*, University of London, England (1979).
10. J. R. Jackson, Jobshop-like queueing systems. *Man Sci* **10** (1), 131–142 (1963).
11. W. J. Gordon and G. F. Newell, Closed queueing systems with exponential servers. *Oper. Res.* **15**, 254–265 (1967).
12. P. G. Harrison, Approximate analysis and prediction of time-delay distributions in networks of queues. *Computer Performance* **2** (3), 124–135 (1981).
13. P. S. Yu, Passage time distributions for a class of queueing networks: closed, open or mixed, with different classes of customers with applications to computer system modelling. *Tech. report no. 135*, Dept. of Electrical Engineering and Computer Science, Stanford University (1977).

Received November 1982