THE UNIVERSITY OF HULL

A COMPUTER ALGORITHM FOR THE APPROXIMATE ANALYSIS OF A MULTISERVER QUEUEING NETWORK

being a Thesis submitted for the Degree of

Doctor of Philosophy

in the University of Hull

by

Frances Rose Gillet, B.A., M.Sc.

August 1983

SUMMARY

This thesis develops a computer algorithm that approximates the mean and standard deviation of the throughput time distribution in an open network of queues with any number of servers, general service time distributions and first come first served queueing discipline. The algorithm is based on the complete decomposition approach developed for a single-server queueing network. Using this method, all the transition processes of a network are assume to be renewal and the network is decomposed into its constituent queues which are analysed individually: the mean and variance of the waiting time distribution and the departure time distribution of a GI/G/n queue are required. The results are recomposed to represent the behaviour of the network as a whole.

No exact results are known for the variance of the departures from a GI/G/n queue, or the mean and variance of the waiting time distribution. In this thesis easily computable approximation formulae are developed for these quantities. The accuracy of the approximations is considerd in comparisons with exact and simulated results.

The approximation formulae are substituted in the complete decomposition algorith and estimates given by the algorithm for the throughput times in a number of networks are compared to values obtained by the simulation of the netwo The effect of the assumption of renewal transition processes on the accuracy of the algorithm is discussed.

The algorithm is applied in two case studies: it is used to predict the queueing times of ships unloading white fish at Mallaig Harbour, and to estimate the waiting times incurred in a job shop manufacturing Universal Joints.

CONTENTS

1. INTRODUCTION

1.1 QUEUEING NETWORKS

Queueing networks have proved to be useful models of systems in which jobs or customers may have to queue for more than one type of service. Such queueing situations occur in a number of fields including the traffic-flow within time-sharing and multiprogramming computer systems, communications networks and teletraffic. The manufacturing job shop can be considered as a network of queues in which the individual jobs demand operations from one or more machine centres before leaving the shop. A machine centre can be represented by a single-server or multiserver queue depending on the number of machines at the centre.

In most cases, the queueing network used to model a system is a simplification of the system, which adequately reflects the parameters that are of importance in predicting its performance. The system measures often of interest are the mean number in the system, the average delay at the individual service centres and the total time a customer, or job, can expect to spend in the system.

Due to the extent of its applications, much interest has been shown in the development of queueing network theory over the past twenty five years. As yet, only networks that satisfy the condition of local balance have been susceptible to exact analysis; consequently much of the more recent effort has been directed towards obtaining good approximations for the parameters of more complex systems.

1.2 THE NETWORK MODEL

A queueing network is a collection of service centres arranged in such a way that customers proceed from one queue to another in order to fulfil their service requirements. Each centre has an associated queue in which jobs may wait prior to receiving service. The centres are characterised by the number of servers, the distribution of service times, and the queue discipline, which determines the order in which jobs arriving at the centre receive service. Customers may be of different classes with varying priority levels and service requirements.

A network is termed 'open' if jobs are permitted to enter or leave the system, and 'closed' if the same jobs remain in the network at all times. Each class of job may follow a different route through the network; this can be specified by assigning a probability to the transition of a job from one centre to another. In most network studies, it is convenient to assume that the transitions between service centres occur instantaneously.

A queueing network is said to be in the steady-state, or equilibrium, if its state at any particular instant in time is independent of the time and state of the network when activities commenced.

In using a network to analyse a system of queues, closed form expressions may be obtained for the equilibrium probability of finding the network in a given state, and the utilisations of the centres and the average waiting and throughput times derived from these.

1.3 A REVIEW OF WORK ON QUEUEING NETWORKS

The simplest form of a queueing network is the tandem queue; here two service centres are in series and customers leaving the first queue immediately enter the next. Hence, the inter-departure times of the first queue are the inter-arrival times of the second. Burke [10] studied the output process of a queue with exponential inter-arrival and service time distributions, and proved that for any number of servers, the departures formed a Poisson process with the same rate as the arrival process. Reich [60] used the concept of reversible Markov chains to prove Burke's result, and he showed that the waiting times at the service centres of a tandem queue with first come first served queue disciplines are independent.

The first indication of the product form of the joint queue length distribution, derived later for more general networks, came from R.R.P. Jackson's result for exponential queues in series [29]. By proving it to be the unique solution of the steady-state equations, he showed that the joint queue length distribution was the product of the distributions obtained for the corresponding M/M/n queues.

J.R. Jackson [30] examined queueing networks with more flexible customer routings. He considered a network in which each centre consisted of a number of servers with exponentially distributed service times, and any queueing discipline that was independent of a customers processing time and route through the network. External arrivals to the system formed Poisson processes, and a customer's route was determined by a transition probability matrix. Jackson found that, as with queues in series, each service centre acted as if it were an isolated M/M/n queue.

In a later paper [31] he allowed the customer arrival process to depend on the total number in the system (thus including the possibility of closed networks) and the service rate at any centre to be a function of the number of servers at that centre. The joint queue length distribution was again proved to be of the product form, and it showed the queue lengths at the centres to be independent when the rates of the arrival processes remained constant. Unaware of Jackson's earlier work, Gordon and Newell [26] derived a more specific form of this result for the special caes of closed networks.

Kelly [33] extended the classical Markovian queueing network models of Jackson, and Gordon and Newell, to include networks of queues with customers of different types. The type of a customer was allowed to influence his route through the system, and under certain conditions, his service time distribution at the centres. Kelly considered the service centres to be single server but, by varying the service rate according to the number in the queue, a multiserver queue could be modelled. Under these conditions, a product form of the joint probability distribution was shown to be the unique solution of the local balance equations. In equilibrium, the local balance equations of a network equate the rate of flow into each possible state of the network to the rate of flow out of that state.

More general results were obtained by Baskett, Chandy, Muntz and Palacios [4], and were further considered by Reiser and Kobayashi [62]; a variety of customer types and different kind of service centres were included, in order to model central processors, data channels, terminals and routing delays in computer systems. Various queue disciplines were allowed and customers could change type on transition from one centre to another. If a centre had FCFS queue discipline, the only service time distribution allowed was the negative exponential distribution, at centres using any of the other allowed disciplines any service time distribution with a rational Laplace transform could be considered, and represented as series of exponential stages. By making use of the local balance equations a product form solution was derived for the joint queue length distribution.

Kelly [34] proved a similar result to be a consequence of his earlier work. He applied certain constraints to the forms of the allowable queue disciplines (thus excluding the FCFS case), and showed that for all networks, with service time distributions that could be expressed as a finite combination of gamma distributions, there was product form solution for the queue length distribution. He conjectured that any general distribution could be approximated to a required degree of accuracy by a mixture of gamma distributions, Barbour [3] was later to prove this conjecture.

Using a method of partial decomposition, Chandy, Herzog and Woo [11] showed the state probabilities of a queueing network to be of the product form, with parameters that could be calculated exactly for any network satisfying local balance conditions. A direct analogy with Norton's result in electrical and circuit theory led them to conclude that, with regard to queue lengths and waiting time distributions, any subsystem of a queueing network, with a single node as input and a single node as output, could be considered to be in an equivalent network in which all queues not in the system were replaced by one composite queue.

Although product form solutions can be easily expressed for many queueing networks, the actual computation of the parameters can be costly, due to the number of operations required to evaluate the normalising constants. Bruell and Balbo [8] have developed algorithms to analyse all multiple job class networks with the service time distributions and queue disciplines for which Baskett et al. were able to prove a product form solution.

More recent exact analysis of Markovian networks has concentrated on the flows within a network. Kelly [34] and Beutler and Melamed [5] have used the reversibility of Markov processes to indicate some instances when the transition processes are Poisson. Walrand and Varaiya [73] extended these results using the theory of martingales to produce necessary and sufficient conditions for the flows in a network to form Poisson processes.

Approximate Results for Queueing Networks

At present, only networks that satisfy local balance conditions can be analysed by computationally efficient algorithms, and Muntz [53] believes that the methods used to develop product form solutions have reached their limits. To satisfy local balance conditions, sevice centres with general service time distributions are limited in the type of queue disciplines that are allowed; processor-sharing and no-queueing disciplines, which are of use in computer systems modelling, can be considered. If FCFS discipline is to be considered the service times must be exponentially distributed.

FCFS and shortest processing time (SPT) are queueing disciplines that are often employed in the scheduling of manufacturing job shops, no exact results are known for networks of centres with general service time distributions and these queue disciplines. Simulation can be used to estimate the waiting times in networks that cannot be analysed exactly. However, it is an expensive tool and attention has been turned to the development of approximate methods of solutions that demand less computer time.

To obtain exact solutions for a queueing network model it may be necessary to make questionable assumptions about the service time distributions or queue disciplines at the service centres. Chandy and Sauer [13] have shown that in many cases systems can be modelled by queueing networks, relying on more credible assumptions, that can be solved approximately, and provide good performance estimates at low cost. Three major approaches have been used to obtain approximate results for networks of queues: the methods of diffusion, partial decomposition and complete decomposition.

<u>The Diffusion Approximation</u>

The diffusion approximation was proposed by Newell [54] in his consideration of rush-hour queues, and it is most useful in situations of heavy traffic. The approximation relies on the Central Limit theorem; its basic arguement being that the number of customers N(t) in a queue at time t will tend to become normally distributed as t is increased. The process N(t) is approximated by a continuous-path diffusion process with a distribution described by a diffusion equation with appropriate boundary conditions.

Kobayashi [46] considered a network of single server queues with general service time distributions and FCFS queue disciplines and, by imposing reflecting barriers on a multidimensional diffusion processs, arrived at an approximate joint queue length distribution of the product form. The queueing theory result for the probability of an empty queue was used to modify the form of the distribution so that it provided better approximations in light traffic conditions. Reiser and Kobayashi [61] assessed the accuracy of the diffusion approximation in a comparison of the approximations with the results of the simulations of various queueing networks. The approximation was most accurate when the coefficients of variation of the service time distributions were close to one and utilisations were high.

With a different treatment of the boundary conditions, Gelenbe [24] was able to obtain a solution to the diffusion equation that took into account the distribution of the idle time of a queue and did not necessitate an modification of the results for light traffic behaviour. The model yielded some queueing theory results exactly and performed well when utilisations were low.

The Partial Decomposition Approach

Partial decomposition involves the exact analysis of a subsystem of a queueing network while regarding the rest of he network as a composite queue. The equivalence of the subsystem and composite queue to the original network relies on a direct analogy with Norton's theorem for electrical circuits. Chandy, Herzog and Woo [11] made use of this technique to obtain exact product form results for the joint queue length distributions for networks that sastified local balance conditions. An iterative method based on Norton's reduction was later developed by Chandy, Herzog and Woo [12] to approximate performance values of closed queueing networks of single server queues with general service times distributions and FCFS queue disciplines. This method was applied to central server models by Sauer and Chandy [66] and Sevcik, Levy, Tripathi and Zahorjan [67]. The techniques of Sevcik et al. proved to be more accurate, but more complex and costly, than those of Sauer and Chandy, but simper and more economical than, though not as accurate, as those of Chandy, Herzog and Woo.

The Complete Decomposition Approach

In the complete decomposition approach, a queueing network is decomposed into its constituent service centres and each queue is analysed individually. In order to do this it is necessary to determine the interactions between the arrival and departures processes within the network, obtain the mean and variance of the waiting times at the queues given their arrival process, and recompose these results to represent the behaviour of the network as a whole. Reiser and Kobayashi [61] combined this method with a diffusion approximation for the behaviour of each queue to analyse open and closed networks of single server queues having general service distributions and FCFS queue disciplines.

Shum [70] suggested that the arrival process to each service centre in a closed network be approximated by a Poisson process;

hence each centre would act as an M/G/1 queue for which the moments of the waiting time distribution are known. This approximation is most useful for complex networks which involve many compositions and decompositions of traffic streams as it can be shown that the combination of a large number of renewal processes tends to produce a Poisson process.

Kuhn [43] used the complete decomposition approach to analyse open networks of single server FCFS queues; he assumed all the transition processes to be renewal, and used an approximation of Kramer and Lagenbach-Belz [42] to evaluate the average waiting times at the GI/G/1 queues and hence estimate the overall throughput time of the network. Shanthikumar [68] simplified the decomposition method, obtained a more accurate approximation for the mean wait in a GI/G/1 queue, and extended Kuhn's method so that queueing networks with shortest processing time queue disciplines could be modelled. A comparison with simulation results showed the approximation algorithm to perform well for a variety of queueing networks.

1.4 AIMS OF THE THESIS

Good approximations now exist for networks of single server queues with general service time distributions and the queue disciplines most commonly used for the scheduling of operations. As yet, there are no results for a multiserver network with general service time distributions and FCFS queueing discipline. Such networks have many practical applications including the modelling of manufacturing job shops, and the monitoring of calls in a telephone switchboard system.

The complete decomposition approach employed by Shanthikumar [68] to obtain approximations for a network of single server queues can be applied to networks of queues with more than one server provided various chahacteristics of the multiserver queues are known. Very few exact results have been derived for GI/G/n queues, those that are known tend to take the form of upper and lower bounds.

The aim of this thesis is to produce a computer algorithm for the approximate analysis of an open network of multiserver queues with general service time distributions and FCFS discipline. Approximations are developed for the coefficient of variation of the departure process, and the mean and variance of the waiting time distribution in a GI/G/n queue. The approximations are incorporated into the complete decomposition algorithm and estimates of the mean and variance of the throughput times in various networks are compared with the results of simulations of the networks.

2 THE COMPLETE DECOMPOSITION ALGORITHM

2.1 INTRODUCTION

The method of complete decomposition used by Shanthikumar [68] for the approximate analysis of networks of single server queues, with general service time distributions and first come first served (FCFS) queueing disciplines, could be applied to networks of queues with more than one server, provided the first two moments of the departure and waiting time distributions of the individual queues were known.

The complete decomposition approach consists of three steps:

1. The analysis of the interactions of the arrival and departure streams within the network.
2. The decomposition of the network and analysis of the individual queues.
3. The recomposition of the results to obtain estimates of the mean and standard deviation of the throughput time distribution.

Steps 1 and 3 were the same for networks of single and multiserver queues, once the mean and standard deviation of the departure time and waiting time distributions were known. Step 2 involved the evaluation of these quantities. As no exact results have yet been derived, it was necessary to develop approximate expressions for the variance of the departure process and the mean and variance of the waiting time distribution of a GI/G/n queue.

In the complete decomposition algorithm only the first two moments of all processes involved were considered. It was therefore implicitly assumed that the shapes of the arrival and service time distributions, described by their higher moments had little influence on queueing times. This assumption has been shown to be true for the M/G/1 queue; the exact value of the mean waiting time is given by the Pollaczek-Khintchine equation and is dependent only on the mean and variance of the service time distribution.

2.2 THE QUEUEING NETWORK

The general model considered was an M-node network; each node corresponding to a service centre i with n_i servers. The service time distribution at each centre was represented by its mean s_i, and coefficient of variation C_{s_i}; all queue disciplines were assumed to be FCFS. External arrivals to centre i were characterised by the mean arrival rate λ_i, and coefficient of variation C_{e_i}. On completion of service at centre at centre i, a customer proceeded to centre j with probability p_{ij}. It was convenient to assume $p_{ii}=0$ for i=1,2,....,M: cases of feedback could be included by increasing the service time requirement of the customer in question ([71]).

A customer left the network after completion of service at centre i with probability $1-\sum_{j=1}^{M} p_{ij}$.

λ_i represented the total arrival rate to centre i, including external and internal arrivals, and was defined ([30]) by the set of equations:

$$\lambda_i=\gamma_i+\sum_{j=1}^{M}\lambda_j p_{ji} \qquad i=1,2,\ldots,M.$$

In an open queueing network, λ_i is independent of the service time distribution and equal to the product of $E[N_i]$, the expected number of visits made by a customer to centre i, and $\sum_{j=1}^{M}\gamma_j$, the total rate of external arrivals to the network, that is

$$\lambda_i = E[N_i] \cdot \sum_{j=1}^{M} \gamma_j = \lambda E[N_i] \qquad i=1,2,\ldots,M$$

where

$$\lambda = \sum_{j=1}^{M} \gamma_j .$$

2.3 THE RENEWAL ASSUMPTION

The development of the complete decomposition was based on the assumption that the external arrivals to the network and all customer transition processes within the network were renewal, and hence each service centre could be represented by a GI/G/n queue. This was only true in a limited number of cases, and a necessary condition was that all departure processes from the service centres were themselves renewal processes. In most instances successive departures from a GI/G/n queue are correlated and the validity of the renewal assumption is questionable. The extent of the correlation present in the transition processes and its effect on the accuracy of the complete decomposition approximation are discussed in Chapter 4.

2.4 THE COMPLETE DECOMPOSITION APPROACH

Step 1

Sevcik et al. [67] suggested expressions for the coefficients of variation of the composition and decomposition of renewal processes:

Decomposition

If C_j is the coefficient of variation of an arrival stream created by choosing jobs with probability $q_j > 0$ from an original stream with coefficient of variation C, then

$$C_j^2 = q_j C^2 + (1 - q_j).$$

The result is exact when the original stream is a renewal process and a comparison with simulation results ([43]) has shown it to be a good approximation for non-renewal processes.

Composition

The coefficient of variation C of the compositon of M renewal processes can be approximated by

$$C = \sum_{j=1}^{M} \frac{\lambda_j C_j}{\lambda_c}$$

where C_j is the coefficient of variation of the j^{th} component stream, λ_j is the arrival rate of stream j, and $\lambda_c = \sum_{j=1}^{M} \lambda_j$.

This result is only exact when all the component streams are Poisson processes.

Combining the results for the decomposition and composition of renewal processes; let C_{d_j} be the coefficient of variation of the departures from centre j in an M-node network then, if C_{a_i} is the coefficient of variation of the arrivals to centre i, the transitions from centre i to centre j will be of rate $\lambda_j p_{ji}$, and

$$C_{a_i}^2 = \sum_{j=1}^{M} \frac{\lambda_j p_{ji}}{\lambda_i}(p_{ji} C_d^2 + (1 - p_{ji})) + \frac{\gamma_i}{\lambda_i} C_e^2 \quad i=1,2,\ldots,M, \qquad (1)$$

the term $\frac{\gamma_i}{\lambda_i} C_{e_i}^2$ corresponding the external arrivals to centre i.

Step 2

Under the renewal assumption, each service centre of the network was considered to form a GI/G/n queue and the variance of the inter-arrival distribution was evaluated using the expression obtained in Step 1. For the set of equations (1) to be solved, an approximation formula for the coefficient of variation of the departures was required, and in order to implement Step 3, expressions were required for the mean and standard deviation of the waiting time in a GI/G/n queue. Approximations that were dependent on only the first two moments of the arrival and service time distributions were developed for these quantities; these are discussed in Chapters 4, 6 and 7.

Step 3

The recomposition method employed by Shanthikumar [68] was applicable to networks of queues with more than one server: the location of a customer after each transition was represented by a finite state Markov chain with transition and absorption states, and an expression was derived for the distribution of N_j, the number of visits made to centre j by an arbitrary customer before leaving the network, and N_{ij}, the number of visits made to a centre j by a customer whose first service was at centre i. It was assumed that the times spent at a service centre on each visit were independently and identically distributed, with mean $E[T_i]$ and it was shown that, if T is the total time spent by a customer in the network, then

Whatever he did for single server would be done in multiserver. Strong assumption

$$E[T]=\sum_{i=1}^{M} E[N_i]E[T_i]$$

where

$$E[N_i]=\frac{\lambda_i}{\sum_{j=1}^{M}\gamma_j}$$

and

$$Var[T]=\sum_{i=1}^{M} E[N_i]Var[T_i]+\sum_{i=1}^{M}\sum_{j=1}^{M} Cov[N_i,N_j]E[T_i]E[T_j]$$

where

$$Cov[N_i,N_j]=E[N_j]E[N_{ji}]+E[N_i]E[N_{ij}]-E[N_i]E[N_j] \quad i\neq j$$
$$Cov[N_i,N_i]=Var[N_i]=E[N_i](2E[N_{ii}]-E[N_i]-1) \qquad (2)$$

and

$$(E[N_{ij}])=(I-P)^{-1} \quad i,j=1,2,\ldots,M;$$

I being the identity matrix, and P the matrix of transition probabilities.

2.5 THE DISTRIBUTION OF THE TIME IN THE SYSTEM

Shanthikumar [68] conjectured that the distribution of the throughput time T could be approximated by a truncated normal distribution with mean and variance given by the complete decomposition algorithm. He offered the justification that T could be considered as the sum of a large number of random variables and consequently the Central Limit Theorem could be applied. This approximation to the throughput time distribution is therefore most reasonable for large networks.

2.6 THE APPROXIMATION ALGORITHM

The steps used to develop the computer algorithm for the approximate analysis of a multiserver queueing network are outlined below. The approximation program is listed in Appendix 2. The program was used to evaluate queueing times in the case studies described in Chapters 9 and 10. The input data required was: M -the number of service centres,

n_i -the number of servers at centre i,

i -the rate of external arrivals to centre i,

C_{e_i} -the coefficient of variation of the external arrivals to centre i,

s_i -the mean service time at centre i,

C_{s_i} -the coefficent of variation of the service times at centre i,

(p_{ij}) -the matrix of transition probabilities,

for $i,j=1,2,\ldots,M$.

Step 1

Evaluation of C_{a_i}, the coefficient of variation of the arrivals to centre i, for each i, by:

a) Evaluation of λ_i, the arrival rate to centre i

$$\lambda_i=\gamma_i+\sum_{j=1}^{M}\lambda_j p_{ji} \qquad i=1,2,\ldots,M$$

Defining the 1xM matrices:

$A=(\lambda_i)$ and $C=(\gamma_i)$

then

$$A=C+A*P$$

therefore

$$A=C*(I-P)^{-1}$$

where I is the identity matrix.

b) Evaluation of an approximate expression for C_{d_j}, the coefficient of variation of the departures from centre j, for each j.

c) Substitution of the expression for C_{d_j} in the set of equations (1) to obtain C_{a_i} for each i.

Step 2

Evaluation of the approximations for $E[W_i]$, the mean waiting time at centre i, and $Var[W_i]$, the variance of the waiting time at centre i, for each i.

Step 3

Approximation of E[T], the mean throughput time, and Var[T], the varaince of the throughput time distribution, by:

a) Evaluation of $E[T_i]$, the mean throughput time, and $Var[T_i]$, the variance of the throughput time at centre i, using:

$$E[T_i]=E[W_i]+\frac{s_i}{n_i}$$

$$Var[T_i]=Var[W_i]+C_{s_i}^2\left(\frac{s_i}{n_i}\right)^2 \qquad i=1,2,\ldots,M.$$

b) Evaluation of $E[N_i]$, the expected number of visits made to centre i by an arbitrary customer:

$$E[N_i]=\frac{\lambda_i}{\sum_{j=1}^{M}\gamma_j} \qquad i=1,2,\ldots,M$$

Evaluation of $E[N_{ij}]$, the expected number of visits made to centre j by a customer whose first service was at centre i:

$$W=(E[N_{ij}])=(I-P)^{-1} \qquad i=1,2,\ldots,M$$

c) Evaluation of $Cov[N_i,N_j]$, the covariance of N_i from equations (2), for each i and j.

d) Evaluation of E[T] and Var[T] from the equations:

$$E[T]=\sum_{i=1}^{M}E[N_i]E[T_i]$$

$$Var[T]=\sum_{i=1}^{M}E[N_i]Var[T_i]+\sum_{i=1}^{M}\sum_{j=1}^{M}Cov[N_i,N_j]E[T_i]E[T_j].$$

3 VALIDATION OF THE SIMULATION PROGRAMS

3.1 INTRODUCTION

Very few exact results are known for the characteristics of a GI/G/n queue, or for networks of such queues. Consequently, the accuracy of the approximations to the mean and variance of the throughput times, given by the complete decomposition algorithm, were assessed by a comparison with simulation results; and approximation formulae for the coefficient of variation of the departure process and the variance of the waiting time distribution of a GI/G/n queue were obtained by applying linear regression techniques to the data obtained from simulations of the queue.

A simulation of a GI/G/n queue with first come first served (FCFS) queueing discipline was written in FORTRAN77 and run on the Harris S135. This simulation was incorporated into a larger program which modelled a network of GI/G/n queues, with the routes of jobs in the network determined by a matrix of transition probabilities.

The general arrival and service time distributions of the queues were represented by gamma distributions with the appropriate means and standard deviations. In the development of the complete decomposition algorithm, no assumptions were made about the shape of the arrival and service time distributions; the algorithm was dependent only on the first two moments of the distributions, and it was conjectured that the higher moments had little influence on queueing times. However, in applications it is common to have unimodal frequency distributions, with coefficients of variation between 0 and 1, that can be readily approximated by gamma type frequency curves.

The regenerative method of simulation and a sequential stopping rule were used to determine the number of departures in a single run of the simulations; in the interests of economy an upper limit was imposed on the number of departures allowed. Confidence intervals were estimated for the values of the mean waiting times in the systems. Flow diagrams of the queueing systems are given in Figures 3.1 and 3.2. The simulation programs are listed in Appendices 1.1 and 1.2.

3.2 THE PROCESS GENERATORS

The gamma distributed inter-arrival and service times were generated using an algorithm developed by Atkinson [2]. The algorithm is based on a composition/rejection method in which the range of the random variables is split into two parts at the mode of the distribution. It is most efficient for gamma distributions with squared coefficients of variation between 0.25 and 1.

Uniformly distributed random variables were required by the algorithm and were also used to determine the route of a job through the network from the transition probability matrix. The NAG library subroutine G05CAF was used to generate these variables. The subroutine used a multiplicative congruential method and had a cycle length of the order of 2^{57}. To avoid impairing the statistical properties of a sequence of random variables, it was recommended that the number in the sequence should not exceed the square root of the cycle length. The subroutine G05CCF was called to provide an initial value for each sequence, and this yielded different subsequent sequences of random numbers in different runs of the calling program. The number of uniform random variables required for each of the simulation runs was recorded. The maximum value observed, of 3,302,393, was less than 2^{22} and it therefore seemed reasonable to assume that the lengths of all the sequences generated were less than the square route of their cycle lengths.

3.3 REGENERATIVE SIMULATION

Two major problems exist in estimating the parameters of the steady state behaviour of a queueing system by means of simulation; these being the statistical dependence between successive observations and the inability of the simulator to begin the system in the steady state. Crane and Iglehart [18] suggested the regenerative method of simulation to avoid such difficulties. This necessitated finding a random grouping of observations which provided independently and identically distributed blocks from the start of the simulation. The key requirement in obtaining the blocks was that the system simulated returned to a single state infinitely often, and that the mean time between returns was finite. In a multiserver queueing network the busy period structure provided such a sequence of blocks. The simulation was started with the system in the empty state, and a regenerative cycle was completed on each return to the empty state; all observations were averaged within the regenerative cycles.

Figure 3.1 Flow Diagram for the Simulation of a FCFS Multiserver Queue

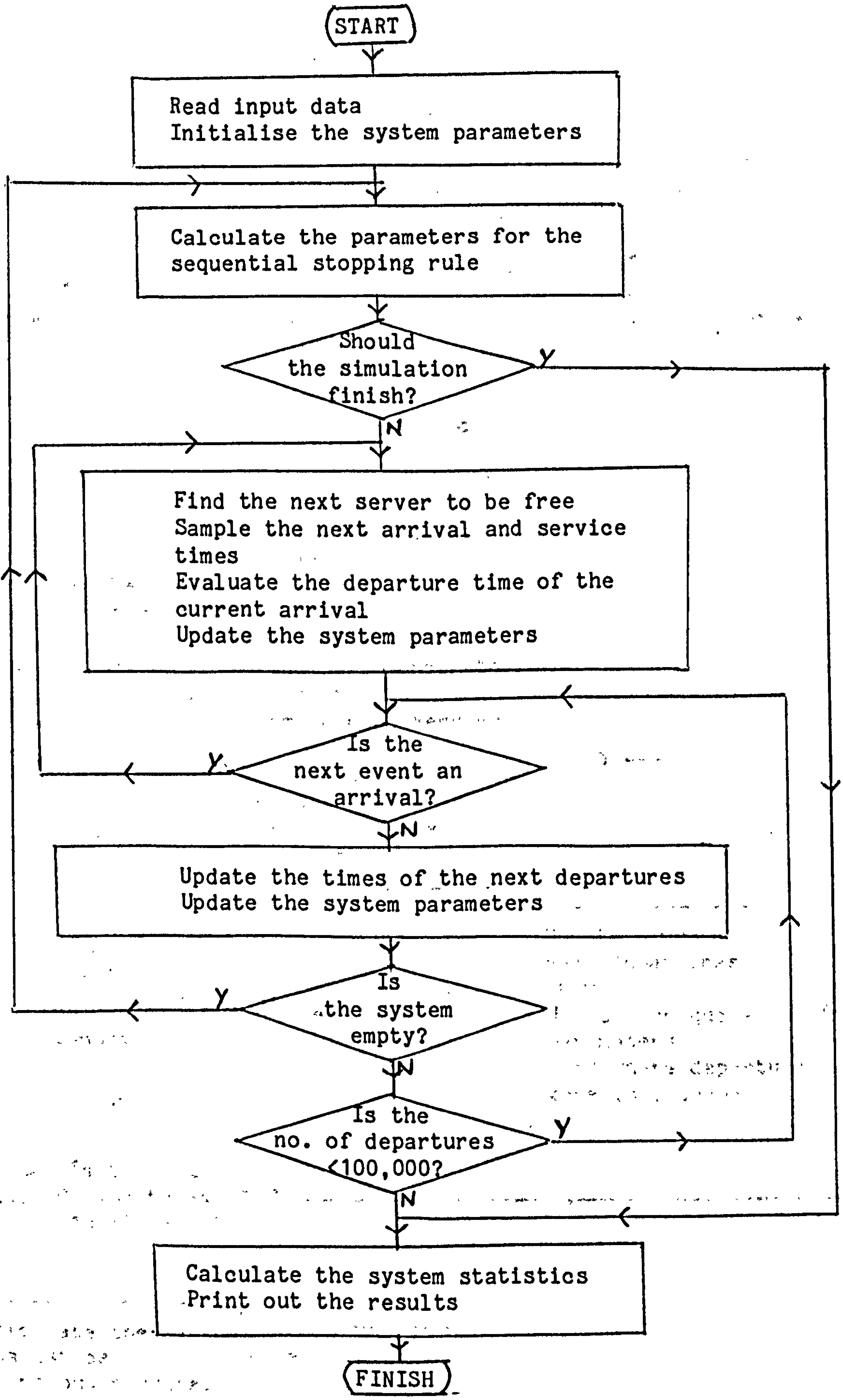
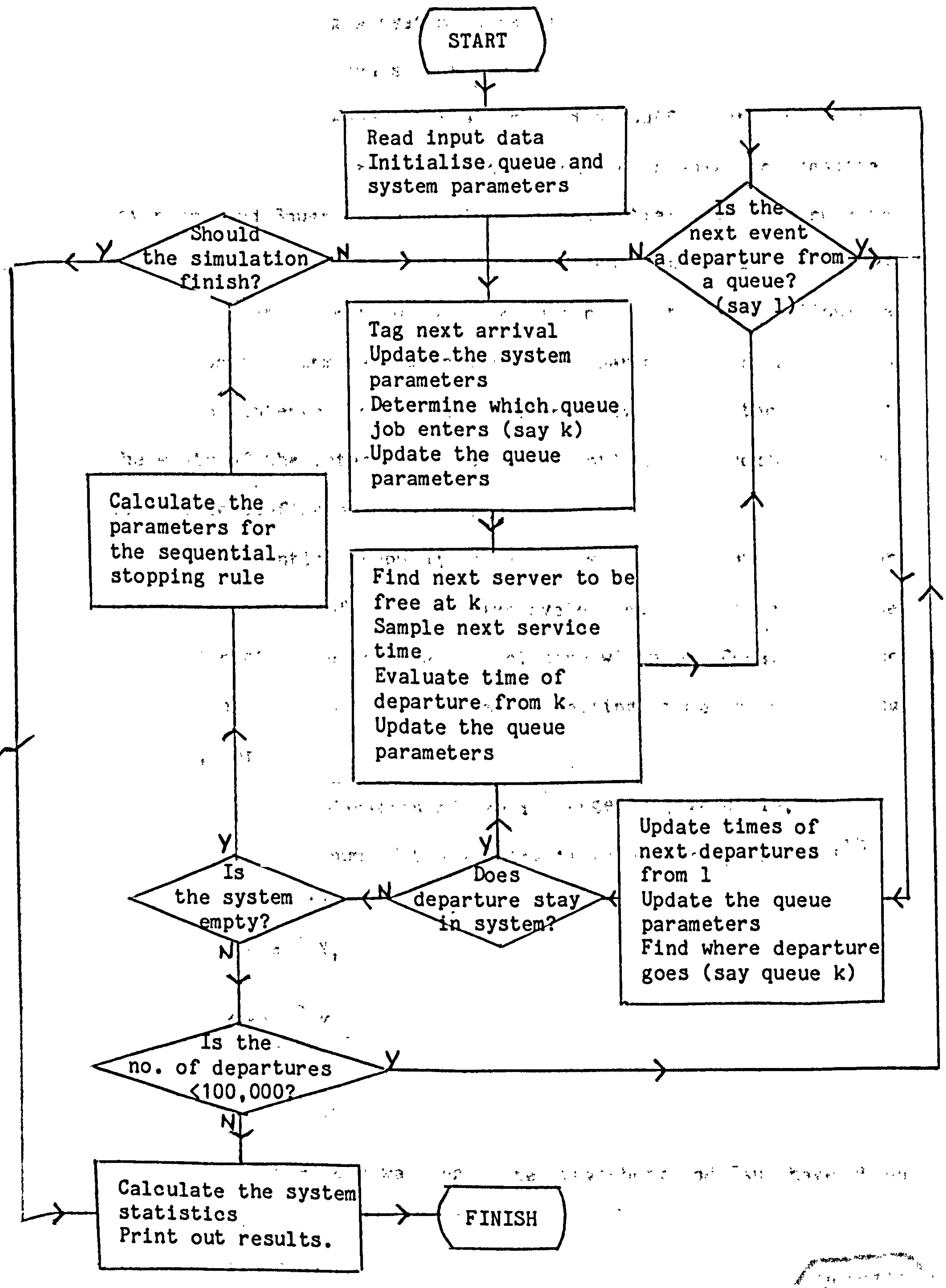


Figure 3.2 Flow Diagram for the Simulation of a Network of FCFS Multiserver Queues



3.4 THE SEQUENTIAL STOPPING RULE

When simulating a system it is desirable to obtain both point and confidence intervals for the parameters of interest. The width of the confidence intervals should be sufficiently small for useful conclusions to be drawn from the simulation results. Lavenberg and Sauer [44] developed a sequential stopping rule to control the width of the estimated confidence intervals when using the regenerative method of simulation. The rule allows a simulation to terminate when simulated parameter has achieved a confidence interval of a specified relative width (the ratio of the width of the interval to its mid-point) with approximately the percentage confidence required.

The sequential stopping rule determines a minimum value for the number $N(\alpha,\gamma)$ of regenerative cycles necessary to approximate a $100\times\alpha\%$ confidence interval of relative width γ. Considering the confidence interval for the mean waiting time in a queueing network, define:

X_i =the duration of the i^{th} regenerative cycle,

Y_i =the sum of the waiting times occurring in the i^{th} regenerative cycle,

$$X(n)=\frac{\sum_{i=1}^{n} X_i}{n}$$

$$Y(n)=\frac{\sum_{i=1}^{n} Y_i}{n}$$

$$r(n)=\frac{Y(n)}{X(n)}$$

proportion

If r is the mean waiting time, Lavenberg and Saur have shown that

$$\lim_{n\to\infty} r(n)=r$$

Let

$$S_x^2(n)=\frac{\sum_{i=1}^{n}[X_i-X(n)]^2}{n-1}$$

$$S_y^2(n)=\frac{\sum_{i=1}^{n}[Y_i-Y(n)]^2}{n-1}$$

$$S_{xy}(n)=\frac{\sum_{i=1}^{n}[X_i-X(n)][Y_i-Y(n)]}{n-1},$$

and

$$S^2(n)=S_y^2(n)-2r(n)S_{xy}(n)+r^2(n)S_x^2(n).$$

Then, if

$$N(\alpha,\gamma)=\min\{n:\ n>2;\ S(n)>0;\ \frac{2\delta(n,\alpha)}{r(n)}<\gamma\},$$

where

$$\delta(n,\alpha)=\frac{\Phi^{-1}[(1+\alpha)/2]S(n)}{nX(n)}$$

and

$$\Phi(t)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{t}\exp\left(\frac{-u^2}{2}\right)du,$$

Lavenberg and Sauer proved that

$$I(\alpha,\gamma)=(r(N(\alpha,\gamma))-\ (N(\alpha,\gamma),\alpha),r(N(\alpha,\gamma))+\ (N(\alpha,\gamma),\alpha))$$

is approximately an 100xα% confidence interval for the mean waiting time and the relative width of the interval is less than γ.

Lavenberg and Sauer experimented with regenerative simulations of various queueing systems and found that for the 90% level of confidence considered, a relative width of 0.1 was sufficiently small to yield valid confidence intervals in most cases.

3.5 THE ACCURACY OF THE SIMULATION

The simulations were run using the sequential stopping rule with $\alpha=0.95$, and $\gamma=0.1$; being considered as a fraction of the mean waiting time in the systems. The number of departures per cycle, and the number of cycles required by the rule varied, with some very large values arising for networks with high utilisations and numbers of servers. In order to control the amount of computer time needed, an upper limit of 100,000 was imposed on the number of departures allowed in a single run of the simulations.

The estimates of various system parameters provided by the simulations were compared with exact values calculate by Sakasegawa [65] for $E_j/E_2/n$ queues. Exact results, evaluated by Shanthikumar [68] for networks of FCFS single-server queues, with two and four service centres, were used to test the accuracy of the network simulation. Results were given for flow shops; in which all jobs follow the same route through the network and visit each service centre once, and symmetric shops; in which, on completion of a service, the probabilities of a job visiting any of the other service centres or leaving the network are equal.

Tables 3.3-3.6 show that in all cases, the exact values of the mean waiting times in the systems were contained in the 95% confidence intervals estimated for the simulated values, and the estimates of the other parameters were close to their exact values.

Table 3.3 Simulation and Exact Results for $E_j/E_k/1$ Queues

	$E_2/E_2/1$ Simulation	$E_2/E_2/1$ Exact	$E_5/E_2/1$ Simulation	$E_5/E_2/1$ Exact
Utilisation	0.6009	0.6000	0.5999	0.6000
Squared coefficient of variation of arrivals	0.5034	0.5000	0.2006	0.2000
Average wait (in units of mean service time)	0.6484±0.0323	0.6306	0.3446±0.0172	0.3488
Mean number in the system	0.3693	0.3784	0.2058	0.2094
Standard deviation of number in system	1.0673	1.1011	0.8235	0.8336
Probability of no wait	0.5332	0.5254	0.6503	0.6511
Probability all servers busy	0.5957	0.6000	0.5949	0.6000
Number of departures	47585	-	44115	-

Table 3.4 Simulation and Exact Results for $E_j/E_k/5$ Queues

	$E_2/E_2/5$ Simulation	$E_2/E_2/5$ Exact	$E_5/E_2/5$ Simulation	$E_5/E_2/5$ Exact
Utilisation	0.5989	0.6000	0.5985	0.6000
Squared coefficient of variation of arrivals	0.4998	0.5000	0.2010	0.2000
Average wait (in units of mean service time)	0.2035±0.0133	0.1994	0.0842±0.0064	0.0837
Mean number in the system	0.1263	0.1196	0.0518	0.0502
Standard deviation of number in system	1.6076	1.5903	1.3189	1.3111
Probability of no wait	0.8623	0.8612	0.9244	0.9242
Probability all servers busy	0.1719	0.1741	0.1276	0.1275
Number of departures	100000	-	100000	-

TABLE 3.5 Simulation and Exact Results for M-centre Flow Shops

	M=2 Simulation	M=2 Exact	M=4 Simulation	M=4 Exact
Utilisation	0.5982	0.6000	0.5996	0.6000
Squared coefficient of variation of arrivals	0.9978	1.0000	0.9989	1.0000
Mean throughput time (in units of mean service time)	4.9025±0.2407	5.0000	10.1209±0.4527	10.0000
Standard deviation of throughput time	3.5013	3.5355	4.9792	5.0000
Mean number in system	2.9304	3.0000	5.9070	6.0000
Number of departures	14149	-	20136	-

Table 3.6 Simulation and Exact Results for M-centre Symmetric Shops

	M=2 Simulation	M=2 Exact	M=4 Simulation	M=4 Exact
Utilisation	0.5991	0.6000	0.6002	0.6000
Squared coefficient of variation of arrivals	1.0754	1.0000	1.0797	1.0000
Mean throughput time (in units of mean service time)	5.0096±0.2432	5.0000	9.8477±0.4260	10.0000
Mean number in system	3.0541	3.0000	5.9268	6.0000
Number of departures	18963	-	11940	-

4. AN APPROXIMATION FOR THE COEFFICIENT OF VARIATION OF THE DEPARTURES FROM A GI/G/N QUEUE

4.1 INTRODUCTION

In order to apply the complete decomposition algorithm to a network of multiserver queues, an expression was required for the coefficient of variation of the departures from a GI/G/n queue. There are few exact results known for the departure processes of queueing systems. Burke [10] showed that the departures from an M/M/n queue formed a Poisson stream. A simple proof of this result was also produced by Reich [60], who went on to show that the departures from an $E_2/E_2/1$ system did not have an E_2 distribution. Mirasol [51] proved that the output of an $M/G/\infty$ queue was Poisson, and the very limiting conditions necessary for the departures of an M/G/1 queue to form a renewal process were established by Disney, Farell and De Morais [22]. Cox [16] used renewal theory to show that as the number of servers was increased the output of a queue tended to form a Poisson process.

The following expression for the squared coefficient of variation (SCV) of the departure process Cd^2 of a GI/G/1 queue was derived by Marshall [48]:

$$C_d^2 = C_a^2 + 2u^2 C_s^2 - 2u(1-u)\frac{E(W)}{s}$$

exact formula.

where C_a and C_s were the coefficients of variation of the arrival and service time distributions respectively, s was the mean service time, u the utilisation, and E(W) was the mean waiting time of a customer in the queue. Kuhn substituted the Kramer and Lagenbach-Belz [42] approximation for E(W)

$$E(W) \simeq \frac{su}{2(1-u)}(C_a^2 + C_s^2)\, g(u, C_a^2, C_s^2)$$

where

$$g(u, C_a^2, C_s^2) = \exp\left\{\frac{-2(1-u)(1-C_a^2)^2}{3u(C_a^2+C_s^2)}\right\} \quad C_a^2<1$$

to obtain the expression

$$C_d^2 = C_a^2 + 2u^2 C_s^2 - u(C_a^2 + C_s^2)\, g(u, C_a^2, C_s^2).$$

From this he produced the simplified approximation formula

$$C_d^2 \approx (1-u^2)C_a^2 + u^2 C_s^2$$

used by Shanthikumar [68] in the complete decomposition algorithm for a single-server network.

Due to the complexity of the output process of a queueing system with more than one server, no simple form has been derived for the coefficient of variation of the departures from a GI/G/n queue. To obtain an expression for substitution in the complete decomposition algorithm, a statistical analysis was performed on the coefficients of variation of the departures of a range of GI/G/n queues. The data was produced by the simulation program listed in Appendix 1.1.

4.2 GLIM

The statistical package GLIM (Generalised Linear Interactive Modelling) was used to fit a regression model which explained the variation in the squared coefficient of variation of the departures from a GI/G/n queue in terms of its utilisation, number of servers, and coefficients of variation of its arrival and service time distributions. GLIM could be operated interactively and was useful for exploratary work; allowing rapid adaptation of the experimental models fitted to the data.

Fitting a Model with GLIM

A dependent y-variate was declared along with a number of explanatory x-variables. The generalised linear model fitted was of the form

$$y_i = \mu_i + \varepsilon_i$$

It consisted of a systematic component μ_i, and one random component ε_i; the distribution of ε_i being a member of an exponential family that included the normal, binomial, Poisson, chi-squared and gamma distributions. The explanatory variables entered as a sum of their effects

$$\eta_i = \sum_j \beta_j x_j .$$

The mean η_i was functionally related to $_i$

$$\eta_i = g(\mu_i);$$

g could be defined as the identity link $(\eta_i = \mu_i)$, or one of the other available link functions including log and reciprocal.

The method of maximum likelihood was used to estimate the linear parameters β_j, and hence the linear predictors η_i, and fitted values μ_i. The problem was then to obtain the best 'trade-off' between the number of explanatory variables to be included in the linear structure, and the ability of the model to represent the data. This was done by an analysis of the residuals of the model in question and a comparison of the deviances of different models. The deviance of a model was calculated for each regression fitted, and in the case of multiple regression with a normal error distribution and identity link function, it was the residual sum of squares.

4.3 OBTAINING THE APPROXIMATION FORMULAE

Values of the squared coefficient of variation of the departures from GI/G/n queues were obtained by simulating the systems with squared coefficients of variation of arrival and service time distributions in the range (0,1); combinations of the values 0.1,0.3,0.5,0.7 and 0.9 being considered to produce a sample of twenty-five for each utilisation and number of servers. Utilisations were increased from 0.1 to 0.9 in steps of 0.1, the values 0.15, 0.85 and 0.95 were included later to give a clearer picture of the behaviour of the coefficient of variation of departures at the extremes of utilisation. Queues with up to ten servers were simulated. On examination, the values of C_d^2 showed little variation in form for values of n exceeding seven. Systems with twenty servers were also simulated to ensure that the results remained consistent for large n.

When using GLIM to fit regression models to the data, a normal error structure and identity link structure were assumed. For each instance of u and n, C_d^2 was declared the dependent y-variate and various products of C_a and C_s were considered as the explanatory x-variables for the models. Analysis of the resultant deviances indicated that the variation in C_d^2 could be adequately explained by a model of the form

$$C_d^2=k+fC_a^2+gC_s^2+hC_aC_s+iC_s \qquad (1)$$

k,f,g,h and i being dependent on the values of u and n.

The effect of u on the coefficients of (1) was examined for each value of n, and a model of the form

$$1-C_d^2=f(n)(1-u)(1-C_a^2)+g(n)u^2(1-C_s^2)+h(u)(1-u)C_s(1-C_a) \qquad (2)$$

seemed to best represent the data.

An investigation into the dependence of the coefficients of (2) on n, provided the following expression

$$1-C_d^2=(1.60-0.09n)(1-u)(1-C_a^2)+\frac{1.47u^2}{n}(1-C_s^2)$$
$$+(1.03-0.03n)(1-u)C_s(1-C_a) \quad (F1)$$

The standard deviations of error produced on fitting (F1) were calculated for each instance of u and n. Table 4.1 shows that the standard errors were less than 0.05, for values of n in the range [2,10], and utilisations >0.2. However, for utilisations $\leqslant$0.2, and all utilisations when n took the value one or twenty, standard errors of up to 0.40 indicated that (F1) could not always be considered a good estimator Of C_d^2.

The forms of the coefficients of (1) were re-examined: a simultaneous consideration of u and n as explanatory variables for the coefficients f,g,h, and i resulted in the model

$$1-C_d^2=(1.37-0.04n)(1-u)(1-C_a^2)+\frac{1.13u^2}{n}(1-C_s^2)$$
$$-0.65(1-u)C_s(1-C_a); \quad (F2)$$

an expression similar in form to (F1). The standard deviations of error of the model given in Table 4.2 showed it to be slightly less accurate than (F1) for utilisations in the range [0.2,0.6] when n took values from five to ten, but to have a lower overall average.

Table 4.1 Estimated Standard Deviations of Error of (F1)

n \ u	0.1	0.15	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.85	0.9	0.95
1	0.09	0.09	0.09	0.09	0.08	0.09	0.11	0.15	0.18	0.21	0.24	0.22
2	0.08	0.07	0.06	0.05	0.05	0.04	0.04	0.03	0.03	0.03	0.03	0.04
3	0.08	0.06	0.05	0.05	0.04	0.04	0.03	0.03	0.02	0.02	0.03	0.06
4	0.08	0.05	0.04	0.04	0.04	0.05	0.04	0.03	0.02	0.03	0.03	0.05
5	0.08	0.05	0.03	0.04	0.05	0.04	0.05	0.04	0.02	0.03	0.03	0.05
6	0.10	0.06	0.03	0.04	0.05	0.05	0.05	0.04	0.02	0.02	0.03	0.05
7	0.11	0.06	0.03	0.03	0.05	0.05	0.05	0.03	0.02	0.02	0.03	0.04
8	0.12	0.07	0.04	0.03	0.04	0.04	0.04	0.03	0.02	0.02	0.03	0.04
9	0.14	0.09	0.05	0.03	0.03	0.04	0.04	0.03	0.02	0.02	0.02	0.03
10	0.16	0.10	0.06	0.03	0.03	0.03	0.03	0.03	0.02	0.03	0.03	0.04
20	0.40	0.33	0.30	0.23	0.18	0.15	0.12	0.10	0.06	0.06	0.04	0.04

Average standard deviation of error=0.064

Table 4.2 Estimated Standard Deviations of Error of (F2)

n \ u	0.1	0.15	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.85	0.9	0.95
1	0.07	0.07	0.07	0.09	0.09	0.08	0.07	0.05	0.04	0.04	0.05	0.07
2	0.06	0.06	0.05	0.04	0.04	0.04	0.04	0.05	0.06	0.07	0.07	0.08
3	0.05	0.04	0.03	0.03	0.03	0.03	0.03	0.03	0.05	0.06	0.08	0.09
4	0.04	0.02	0.03	0.05	0.06	0.05	0.04	0.03	0.04	0.05	0.07	0.09
5	0.03	0.03	0.04	0.07	0.08	0.06	0.06	0.04	0.03	0.04	0.06	0.09
6	0.03	0.03	0.06	0.08	0.09	0.08	0.07	0.04	0.02	0.03	0.05	0.08
7	0.03	0.04	0.06	0.09	0.10	0.09	0.07	0.05	0.03	0.03	0.04	0.07
8	0.03	0.04	0.07	0.10	0.10	0.09	0.07	0.05	0.03	0.02	0.04	0.06
9	0.03	0.04	0.07	0.10	0.10	0.09	0.08	0.05	0.02	0.02	0.03	0.06
10	0.04	0.04	0.07	0.10	0.10	0.09	0.07	0.05	0.03	0.02	0.04	0.05
20	0.12	0.06	0.04	0.03	0.04	0.04	0.03	0.03	0.03	0.03	0.03	0.03

Average standard deviation of error=0.054

4.4 SUBSTITUTION OF THE APPROXIMATION FORMULAE IN THE ALGORITHM

Estimates of C_{a_i}, the squared coefficient of variation of the arrivals to the i^{th} centre in an M-centre queueing network, were obtained by substituting each of three approximation formulae for C_{d_j}, $(j=1,2,\ldots,M)$, in the complete decomposition algorithm.

The first approximation considered was the one developed by Kuhn and used by Shanthikumar in the analysis of single-server networks:

$$C_d^2=(1-u^2)C_a^2+u^2C_s^2 \qquad \text{(K)}$$

The other two formulae were (F1) and (F2), obtained in the previous section.

The expressions were substituted in the set of equations

$$C_{a_i}^2=\sum_{j=1}^{M}\frac{\lambda_j p_{ji}}{\gamma_i}(p_{ji}C_{d_j}^2+(1-p_{ji}))+\frac{\lambda_i}{\gamma_i}C_{e_i}^2 \quad i=1,2,\ldots,M \qquad (3)$$

derived in Chapter 2, and the equations were solved for the $C_{a_i}^2$.

Substitution of (K) gave a linear matrix equation which was solved by matrix inversion. Substitution of (F1) and (F2) produced quadratic matrix equations, and it was necessary to use an iterative method to find solutions.

The approximate squared coefficients of variation of arrival streams given by the algorithm, using each of the formulae, were compared with values obtained by simulating various types of networks.

The Networks

Queueing networks with two and three service centres and varying numbers of servers and transition probabilities were simulated. All the the service centres of a network were considered to have the same number of servers and average service times. External arrivals to the networks were assumed to be Poisson distributed.

Networks of two single-server centres with average service times of 0.6 and 0.8, and two three-server centres with average service times of 1.8 and 2.4, were simulated with the following transition probability matrices:

Flow $$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

Symmetric $$\begin{pmatrix} 0 & 1/2 \\ 1/2 & 0 \end{pmatrix}$$

Asymmetric $$\begin{pmatrix} 0 & 2/3 \\ 1/3 & 0 \end{pmatrix}$$

Three-centre networks of single-server centres with average service times of 0.6 and 0.8, and of six-server centres with average service times of 3.6 and 4.8 were considered with the transition probability matrices:

Flow $$\begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}$$

Symmetric $$\begin{pmatrix} 0 & 1/3 & 1/3 \\ 1/3 & 0 & 1/3 \\ 1/3 & 1/3 & 0 \end{pmatrix}$$

$$\text{Asymmetric} \begin{pmatrix} 0 & 1/2 & 1/4 \\ 1/4 & 0 & 1/2 \\ 1/4 & 1/2 & 0 \end{pmatrix}$$

Accuracy of the Approximations

The percentage errors given on comparison of the approximations produced by the algorithm, using each of the three formulae for C_d^2, with the simulated values of the squared coefficients of variation of the arrival streams of the networks are shown in Tables 4.3-4.5. In networks with single-server centres, the errors given by substitution of (K) and (F2) were of similar magnitude, and in general less than those given by (F1). In the networks with more than one server at the centres, substitution of (F1) and (F2) provided estimates of the squared coefficients of variation of the arrival streams that were within 5% of the simulation results, errors of up to 65% were observed when (K) was used as the approximation for C_d^2.

The means of the absolute percentage errors for each type of network are listed in Table 4.6. Substitution of the approximation formula (F2) appeared to produce the most consistently accurate results. (F2) was therefore selected as the optimum formula for use in the complete decomposition analysis of multiserver networks.

Table 4.3 % Errors of the Approximations to the SCV of the arrivals to the M^{th} Centre of an M-Centre Flow Shop

M	No. of Servers	C_s^2	u=0.6 Simulated Value	u=0.6 % Errors (K)	u=0.6 % Errors (F1)	u=0.6 % Errors (F2)	u=0.8 Simulated Value	u=0.8 % Errors (K)	u=0.8 % Errors (F1)	u=0.8 % Errors (F2)
2	1	0.2	0.71	-0.14	-19.21	-5.46	0.49	0.41	-48.97	-13.17
		0.5	0.84	-2.03	-12.19	-4.90	0.68	0.00	-22.05	-6.03
		1.0	0.99	0.81	0.81	0.81	0.99	-0.60	-0.60	-0.60
	3	0.2	0.88	-18.63	-1.83	1.83	0.75	-34.67	0.27	8.03
		0.5	0.92	-11.26	-1.29	0.87	0.84	-18.75	0.72	4.78
		1.0	1.01	-1.19	-1.19	-1.19	0.98	2.35	2.35	2.35
3	1	0.2	0.64	-18.04	-43.39	-18.81	0.42	-27.96	-84.60	-31.75
		0.5	0.76	-7.73	-19.40	-7.21	0.62	-9.46	-31.57	-10.10
		1.0	0.99	1.32	1.32	1.32	0.99	0.91	0.91	0.91
	6	0.2	0.94	-43.64	-3.42	-1.38	0.85	-64.07	0.83	4.41
		0.5	0.98	-27.69	-3.28	-2.15	0.93	-65.13	-2.47	-0.43
		1.0	1.00	0.40	0.40	0.40	1.01	1.09	1.09	1.09

Table 4.4 % Errors of the Approximations to the SCV of the arrivals to the M^{th} Centre of an M-Centre Symmetric Shop

M	No. of Servers	C_s^2	u=0.6 Simulated Value	u=0.6 % Errors (K)	u=0.6 % Errors (F1)	u=0.6 % Errors (F2)	u=0.8 Simulated Value	u=0.8 % Errors (K)	u=0.8 % Errors (F1)	u=0.8 % Errors (F2)
2	1	0.2	1.03	-10.92	-14.32	-11.50	0.90	-4.02	-10.73	-5.74
		0.5	1.07	-11.75	-13.71	-12.03	0.94	-3.00	-6.91	-3.83
		1.0	1.11	-10.23	-10.23	-10.23	1.02	-2.15	-2.15	-2.15
	3	0.2	0.98	-6.45	-1.74	-0.82	0.92	-6.32	1.85	3.49
		0.5	0.99	-4.35	-1.31	-0.81	0.95	-3.70	1.27	2.22
		1.0	1.04	-3.75	-3.75	-3.75	1.02	-1.57	-1.57	-1.57
3	1	0.2	1.01	-8.14	-11.22	-8.74	0.92	-4.26	-11.17	-5.90
		0.5	1.04	-8.36	-10.19	-8.65	0.95	-2.94	-6.41	-3.68
		1.0	1.12	-10.95	-10.95	-10.95	0.99	1.21	1.21	1.21
	6	0.2	0.98	-5.32	0.61	1.02	0.97	-9.22	0.62	1.35
		0.5	1.01	-5.27	-1.69	-1.41	0.98	-5.91	0.10	0.51
		1.0	0.99	0.70	0.70	0.70	1.01	0.79	0.79	0.79

Table 4.5 % Errors of the Approximation to the SCV of the arrivals to the M^{th} Centre of an M-Centre Asymmetric shop

No. of M	Servers	C_s^2	u=0.45 Simulated Value	% Errors (K)	(F1)	(F2)	u=0.6 Simulated Value	% Errors (K)	(F1)	(F2)
2	1	0.2	1.03	-8.09	-10.42	-8.48	0.89	1.57	-2.81	0.90
		0.5	1.02	-5.67	-6.95	-5.77	0.95	-1.16	-3.47	-1.37
		1.0	1.07	-6.47	-6.37	-6.37	1.01	-1.19	-1.19	-1.19
	3	0.2	1.01	-6.82	-3.75	-3.16	0.98	-7.37	-2.05	-1.02
		0.5	1.01	-4.27	-2.28	-1.89	0.98	-3.58	-0.21	0.41
		1.0	1.04	-3.38	-3.38	-3.38	1.00	-0.60	-0.60	-0.60
3	1	0.2	1.04	-10.39	-12.90	-10.78	0.88	-1.13	-7.48	-2.61
		0.5	1.03	-7.47	-8.53	-7.57	0.92	-0.33	-4.01	-1.08
		1.0	1.14	-13.20	-12.94	-12.94	1.02	-1.96	-1.96	-1.96
	6	0.2	1.02	-8.99	-4.01	-3.71	0.96	-9.54	0.62	1.34
		0.5	0.98	-2.95	0.51	0.61	0.98	6.31	-0.10	0.41
		1.0	1.02	-3.03	-2.64	-2.54	1.00	0.30	0.30	0.30

Table 4.6 Average Absolute % Errors of the Approximations to the SCV of Arrivals

Type of Network	u	(K)	(F1)	(F2)
Flow	0.6	11.07	8.98	3.86
	0.8	18.78	16.37	6.97
Symmetric	0.6	7.18	6.70	5.88
	0.8	3.76	3.73	2.70
Asymmetric	0.45	6.72	6.22	5.61
	0.6	2.92	2.07	1.10
Overall average		8.41	7.35	4.35

4.5 THE RENEWAL ASSUMPTION

A fundamental assumption in obtaining the equations for the squared coefficients of the arrival streams in a network was that all the transitions between the centres of the network formed renewal processes. However, the departures from most queues, with the exception of the M/M/n sytem, are correlated to some extent. Correlation of the arrival streams can also result from the patterns of the transitions within a network -the simulation results listed in Appendix 3 show that the arrival streams in networks of M/M/n queues were significantly correlated when the transition probabilities were symmetric.

Tables 4.7-4.9 give the lag one correlations of the simulated arrivals to the M^{th} service centre of a number of M-server networks, and the percentage errors of the estimates given by the algorithm, with (F2) estimating the coefficients of variation of the departures. Box and Jenkins [6] have shown autocorrelations to be significant if they exceed $1.96/\sqrt{N}$, where N is the number of observations. The number of arrivals generated in each run of the simulation was determined by a sequential stoppping rule, and varied from 20,000 to the imposed upper limit of 100,000. In the case of the least number of arrivals simulated, correlations exceeding 0.014 were significant, hence, in all the simulations any correlations of more than 0.014 were considered significant. It can be seen from Tables 4.7-4.9 that percentage errors greater than 5% only arose when the arrival streams exhibited significant correlations. The approximation algorithm appeared to over-estimate the coefficients of variation of arrival streams which were negatively correlated, and under-estimate it when significant positive correlations were evident.

4.5 THE RENEWAL ASSUMPTION

A fundamental assumption in obtaining the equations for the squared coefficients of the arrival streams in a network was that all the transitions between the centres of the network formed renewal processes. However, the departures from most queues, with the exception of the M/M/n sytem, are correlated to some extent. Correlation of the arrival streams can also result from the patterns of the transitions within a network -the simulation results listed in Appendix 3 show that the arrival streams in networks of M/M/n queues were significantly correlated when the transition probabilities were symmetric.

Tables 4.7-4.9 give the lag one correlations of the simulated arrivals to the M^{th} service centre of a number of M-server networks, and the percentage errors of the estimates given by the algorithm, with (F2) estimating the coefficients of variation of the departures. Box and Jenkins [6] have shown autocorrelations to be significant if they exceed $1.96/\sqrt{N}$, where N is the number of observations. The number of arrivals generated in each run of the simulation was determined by a sequential stoppping rule, and varied from 20,000 to the imposed upper limit of 100,000. In the case of the least number of arrivals simulated, correlations exceeding 0.014 were significant, hence, in all the simulations any correlations of more than 0.014 were considered significant. It can be seen from Tables 4.7-4.9 that percentage errors greater than 5% only arose when the arrival streams exhibited significant correlations. The approximation algorithm appeared to over-estimate the coefficients of variation of arrival streams which were negatively correlated, and under-estimate it when significant positive correlations were evident.

Table 4.7 Correlation and the % Errors of the Approximation to the SCV of the Arrivals to the M^{th} Centre in an M-Centre Flow Shop

M	No. of Servers	C_s^2	u=0.6 Correlation	u=0.6 % Error	u=0.8 Correlation	u=0.8 % Error
2	1	0.2	0.049	-5.46	0.076	-13.17
		0.5	0.029	-4.90	0.027	-6.03
		1.0	-0.008	0.81	-0.007	-0.60
	3	0.2	-0.022	1.83	-0.097	8.03
		0.5	-0.022	0.87	-0.045	4.78
		1.0	-0.006	-1.19	0.009	2.35
3	1	0.2	0.064	-18.81	0.061	-31.37
		0.5	0.024	-7.21	0.025	-10.10
		1.0	0.000	1.32	-0.002	0.91
	6	0.2	-0.026	-1.38	-0.071	4.41
		0.5	-0.025	-2.15	-0.046	-0.43
		1.0	0.004	0.40	-0.007	1.09

Table 4.8 Correlation and the % Errors of the Approximation to the SCV of the Arrivals to the M^{th} Centre in an M-Centre Symmetric Shop

M	No. of Servers	C_s^2	u=0.6 Correlation	u=0.6 % Error	u=0.8 Correlation	u=0.8 % Error
2	1	0.2	0.051	-11.50	0.029	-5.74
		0.5	0.044	-12.03	0.016	-3.83
		1.0	0.035	-10.23	0.015	-2.15
	3	0.2	0.014	-0.82	-0.004	3.49
		0.5	0.018	-0.81	0.012	2.22
		1.0	0.030	-3.75	0.013	-1.57
3	1	0.2	0.053	-8.74	0.021	-5.90
		0.5	0.054	-8.65	0.006	-3.68
		1.0	0.034	-10.95	0.019	-1.21
	6	0.2	-0.017	1.02	-0.025	1.35
		0.5	0.091	-1.41	-0.005	0.51
		1.0	0.020	0.70	0.007	0.79

Table 4.9 Correlation and the % Errors of the Approximation to the SCV of the Arrivals to the M^{th} Centre in an M-Centre Asymmetric Shop

M	No. of Servers	C_s^2	u=0.45 Correlation	u=0.45 % Error	u=0.6 Correlation	u=0.6 % Error
2	1	0.2	0.040	-8.48	0.017	0.90
		0.5	0.043	-5.77	0.005	-1.37
		1.0	0.034	-6.37	0.011	-1.19
	3	0.2	0.010	-3.16	-0.017	-1.02
		0.5	0.014	-1.89	-0.010	0.41
		1.0	0.022	-3.38	0.008	-0.60
3	1	0.2	0.046	-10.78	0.015	-2.61
		0.5	0.032	-7.57	0.008	-1.08
		1.0	0.037	-12.94	0.004	-1.96
	6	0.2	0.011	-3.71	-0.016	1.34
		0.5	0.014	0.61	-0.015	0.41
		1.0	0.008	-2.64	-0.010	0.30

4.6 CONCLUSIONS

Substitution of the formula (F2) in the complete decomposition algorithm provides good approximations to the squared coefficients of variation of the arrival streams in most multiserver networks. However, the renewal assumption implicit in the algorithm is not always valid, and a consideration of the correlation of the transition processes within a network may result in a greater degree of accuracy being attained.

5 CORRELATIONS IN QUEUEING NETWORKS

5.1 INTRODUCTION

An arrival stream to a service centre of a queueing network may consist of external arrivals to the system and departures from other centres of the network. The characteristics of an arrival stream are thus dependent on those of its component transition streams. The complete decomposition algorithm relies on the assumption that all the transition processes within a network are renewal: the expressions for the composition and decomposition of the transition streams given by Sevcik et al. [67] are only exact for some renewal processes, and the approximations for the parameters of the individual queues of the network were developed for GI/G/n queues. A necessary though, as was observed in the previous chapter (p.44), not always sufficient condition for the transition streams of a network to form renewal processes is that all the departures from a service centre be uncorrelated.

5.2 CORRELATION OF DEPARTURE PROCESSES

It is only in a limited number of circumstances that the output of a queue forms a renewal process. Burke [10] showed that the departure process of an M/M/n queue was Poisson. Disney, Farell and De Morais [22] considered GI/G/1 queues and proved independent departures were only produced by M/M/1 queues, M/D/1 queues with waiting space for one customer, and M/G/1 queues with no waiting space or all service times zero. Cox [17] showed that for an $M/E_k/1$ queue, correlation of the departure process was positive for k>1, and negative for k<1. An upper bound for the correlations of an $M/E_k/1$ queue was given by Jenkins [32]; he found that the highest correlations were produced by the M/D/1 system, and proved $0.5e^{-1} \simeq 0.183$ to be an upper limit for all values of k and utilisations. Considering the M/G/1 queue, Daley [21] concluded that correlations were either positive and small, or negative. He showed that $0.5e^{-1}$ formed an upper bound, but found no lower bound greater than -1.

Shimshak and Sphicas [69] studied the effect of the correlation of the arrival process on the waiting time of a queue with exponentially distributed service times; the arrivals were the departures from an $M/E_k/1$ system. In an comparison of the exact numerical results for the average wait in the second queue, when its arrival were assumed to be independent, with the values obtained from a simulation of the system, it became clear that the independence assumption caused an under-estimation of the waiting time. The discrepancy increased with the degree of correlation of the arrival process, and was greatest when both queues had high utilisations and k was large. Errors in the estimation of the mean waiting time of up to 50% occurred under the independence assumption.

In developing the complete decomposition algorithm for a single-server network, Shanthikumar [68] was concerned with the correlation of the ouput of GI/G/1 queues. Due to the difficulty in obtaining numerical results, he used simulation to investigate the departure processes of GI/G/1 queues with shortest processing time queueing disciplines. It was found that in general, correlation increased with utilisation: correlations were low for queues with Poisson arrivals and low utilisations; in all other cases the validity of the renewal assumption was questionable.

Shanthikumar simulated single-server flow and symmetric job shops with various parameters and FCFS and SPT queueing disciplines. For both disciplines, correlations between successive departures were seen to be higher in symmetric shops than flow shops, when all service centres had utilisations of 0.6, but with utilisations of 0.8 the flow shops exhibited the larger correlations, while symmetric shops with many service centres produced very small correlations, so proving it to be a case where the renewal assumption would hold good.

5.3 CORRELATION IN A MULTISERVER QUEUEING NETWORK

The consideration of the accuracy of the estimation of the coefficient of variation of the arrival streams of a network, in Chapter 4, indicated that the presence of significant correlations had an adverse influence on the performance of the approximation algorithm. In all cases where the approximation over-estimated C_a^2 by more than 5%, the arrivals were negatively correlated; similarly, under-estimations of C_a^2 occurred when large positive correlations were evident. It would therefore be an advantage to be able to predict when such significant correlations would arise in the transition streams of a network, and incorporate this as a factor in the complete decomposition model. However, relaxation of the renewal assumption greatly increases the complexity of an analytical investigation of queueing properties, and a comprehensive study of the effect of correlation on the paramaters of a queueing network using simulation and linear regression techniques is made impractical by the number of different types of network to be considered.

The largest correlations were observed in the arrival streams of flow shops. Here the departures from one service centre formed the arrival stream to the next, and the correlation of the arrivals were those of the output of the preceeding queue. For this reason, the correlation of the departure processes of some GI/G/n queues has been examined to establish, at least for flow shops, when the approximation is likely to over or under-estimate the coefficients of variation of the arrival streams of a network.

5.4 CORRELATION OF THE OUTPUT OF $E_j/E_k/n$ QUEUES

A range of $E_j/E_k/n$ queues were simulated, and the lag one autocorrelations of the departure processes were estimated and are shown in Figures 5.1-5.7. Both positive and negative values resulted, the maximum value of 0.123 was observed for an $M/E_k/1$ queue, with a utilisation of 0.9, and the minimum of -0.398 was given by an $E_2/E_{10}/2$ system, with u=0.9. The dependence of the degree of correlation in the output of the queues on the values of n,u,j and k was investigated.

Positive correlations only arose when single-server queues were simulated. Figure 5.1 shows that all the correlations given by $M/E_k/1$ systems were positive. This was the result proved by Cox [17] for all $M/E_k/1$ queues with k>1. Some positive correlations were also evident in the departure processes of $E_2/E_k/1$ systems when k was large (Figure 5.2). In all other instances the departure streams were negatively correlated, with the minimum values occurring when n was equal to two, and the output approaching a renewal process as the number of servers increased (Figures 5.2 and 5.3). Cox [16] used renewal theory to show that the departure process of a queue tended to form a Poisson stream as the number of servers was increased. For $E_j/E_k/n$ queues with j less than k, the correlation of the departure processes generally increased as the utilisation of the queue was increased (Figures 5.4 and 5.5). When j was equal to k, increasing u had little affect on the dependence of the departures (Figures 5.5 and 5.6). When j was greater than k, the correlation decreased as the utilisation was increased (Figures 5.6 and 5.7).

In all instances the magnitude of the correlation of the output of an $E_j/E_k/n$ queue was increased when either j or k was increased.

5.5 CONCLUSIONS

When large correlations were evident in the arrival streams of queueing networks, the complete decomposition algorithm did not provide accurate estimates of the squared coefficients of variation of the arrivals. The greatest degree of correlation was observed in the transition processes of flow shops where the level of interaction between service centres was low. In such cases knowledge of the correlation of the output processes of the individual queues can give some insight into whether the coefficients of variation of the arrival streams will be over or under-estimated by the algorithm: if the arrival stream to a centre is composed of departures from single-server queues, with little variation in service times and high utilisations, then correlations will be positive, and the coefficient of variation is likely to be under-estimated by more than 5%, conversely, queues with two or more servers and fairly constant inter-arrival or service times will produce negatively correlated arrival streams and the approximation algorithm will tend to over-estimate the coefficients of variation of the arrivals.

Figure 5.1
Correlation of departures from an $M/E_k/1$ queue

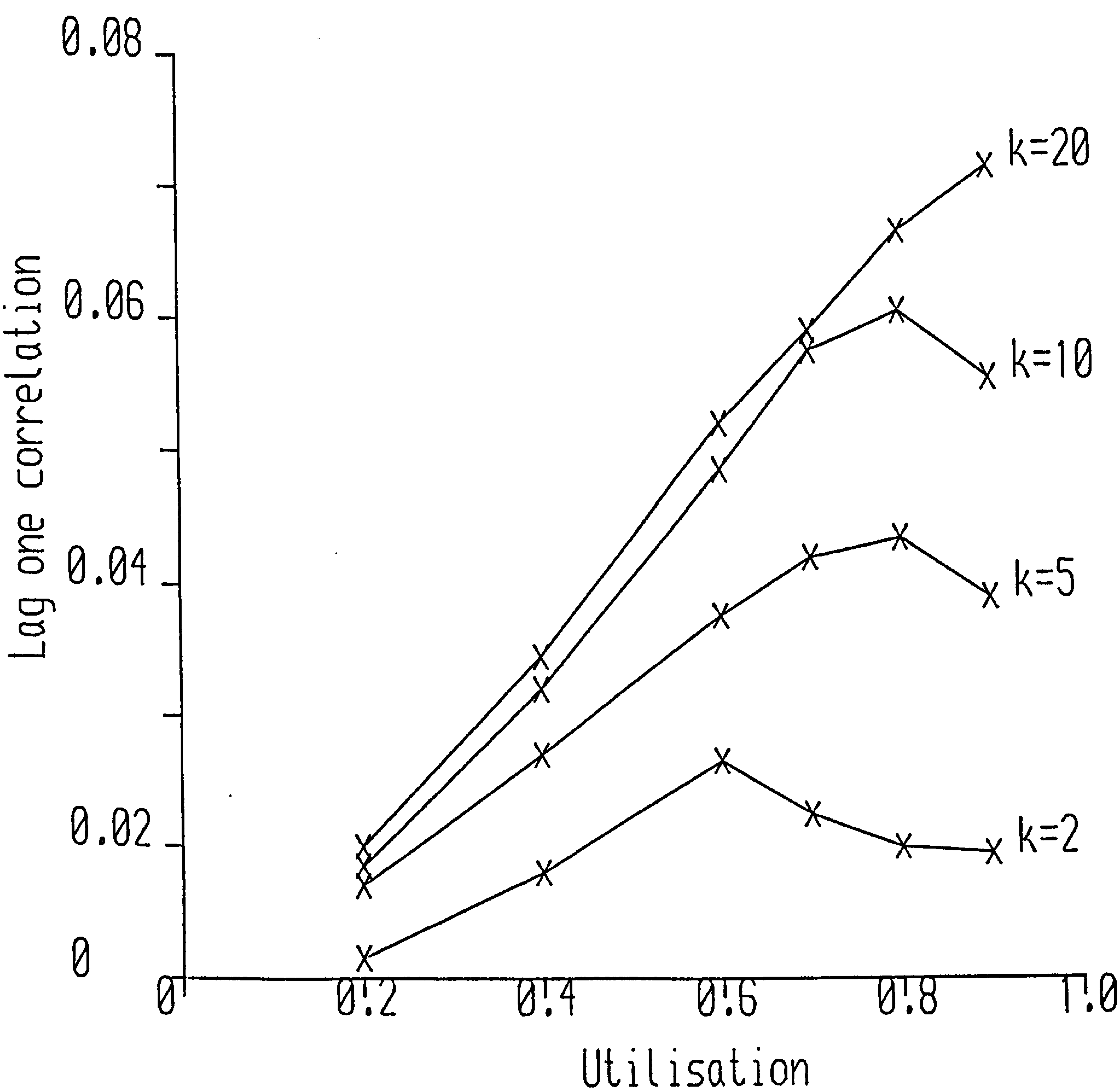


Figure 5.2
Correlation of departures from an $E_2/E_k/n$ queue

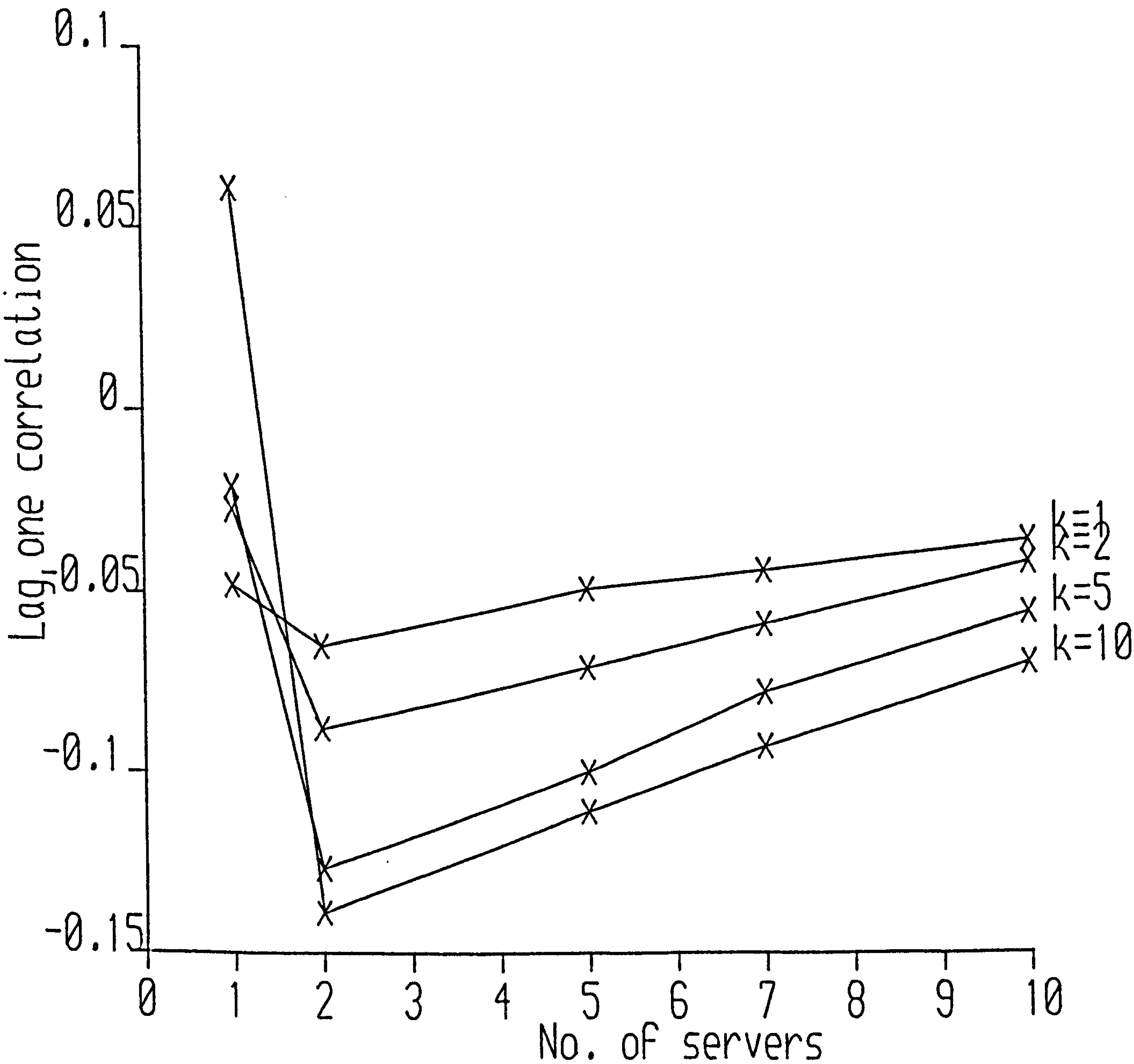


Figure 5.3

Correlation of departures from an $E_j/E_2/n$ queue

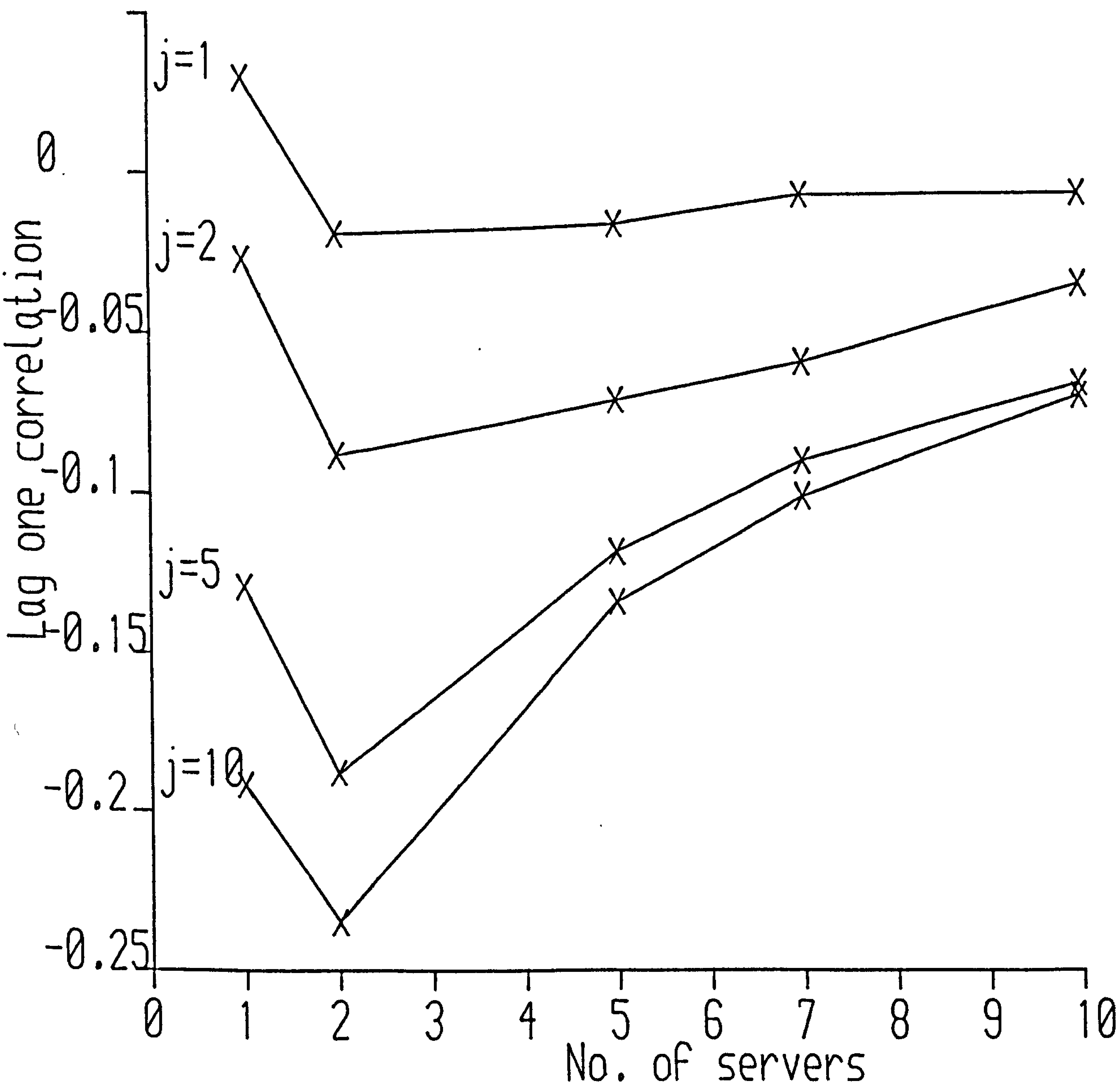


Figure 5.4
Correlation of departures from an $E_j/M/2$ queue

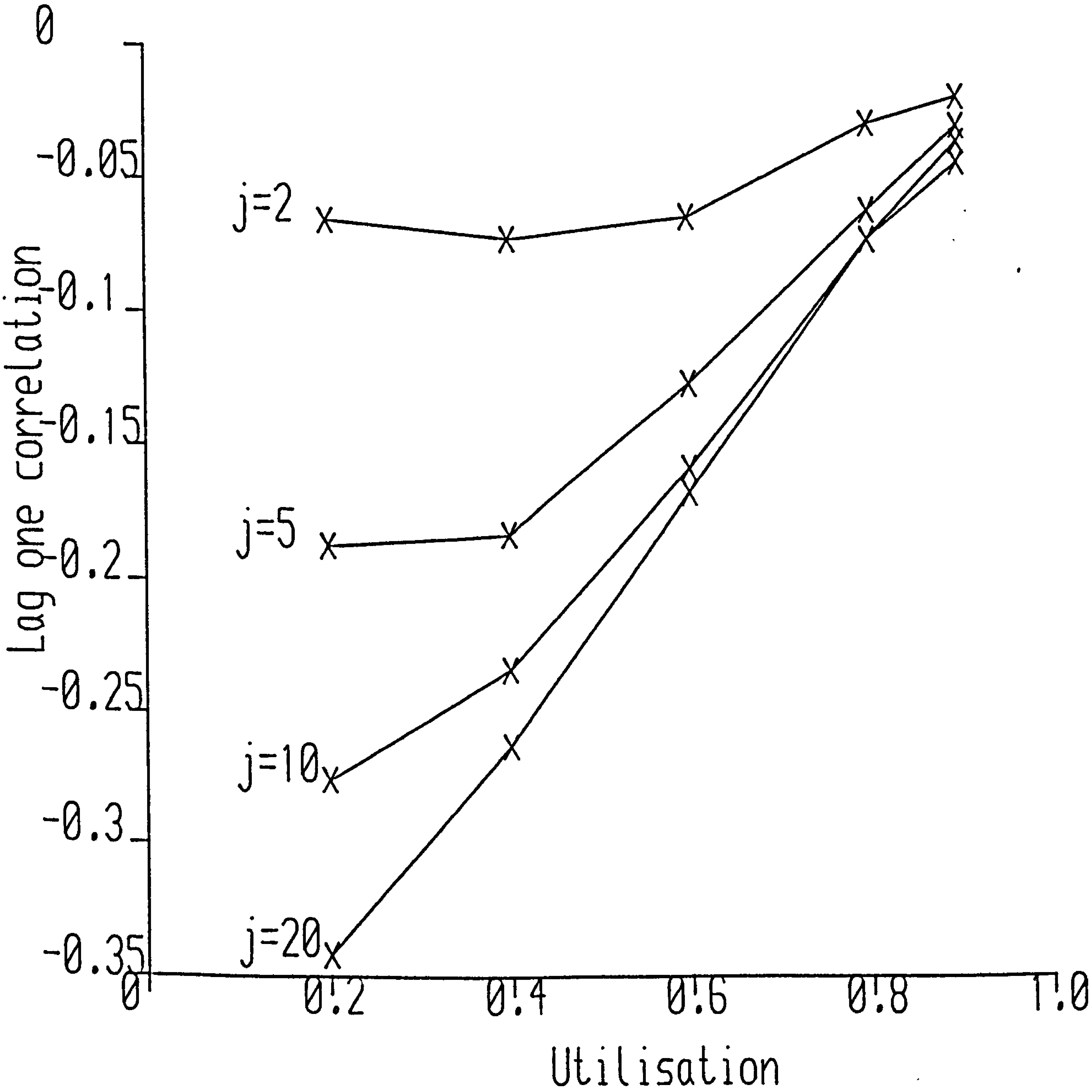


Figure 5.5
Correlation of departures from an $E_j/E_2/2$ queue

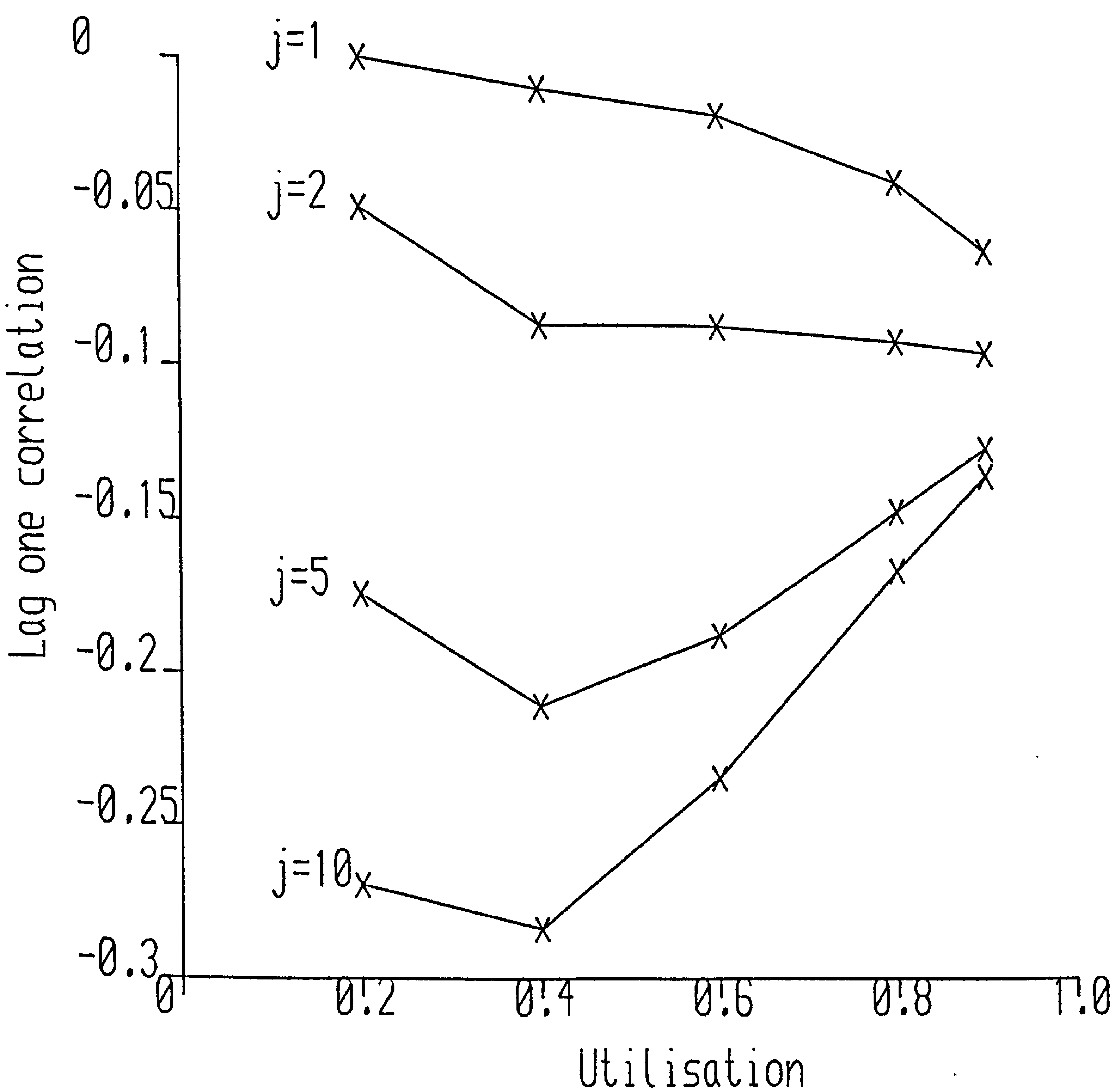


Figure 5.6
Correlation of departures from an $E_2/E_K/2$ queue

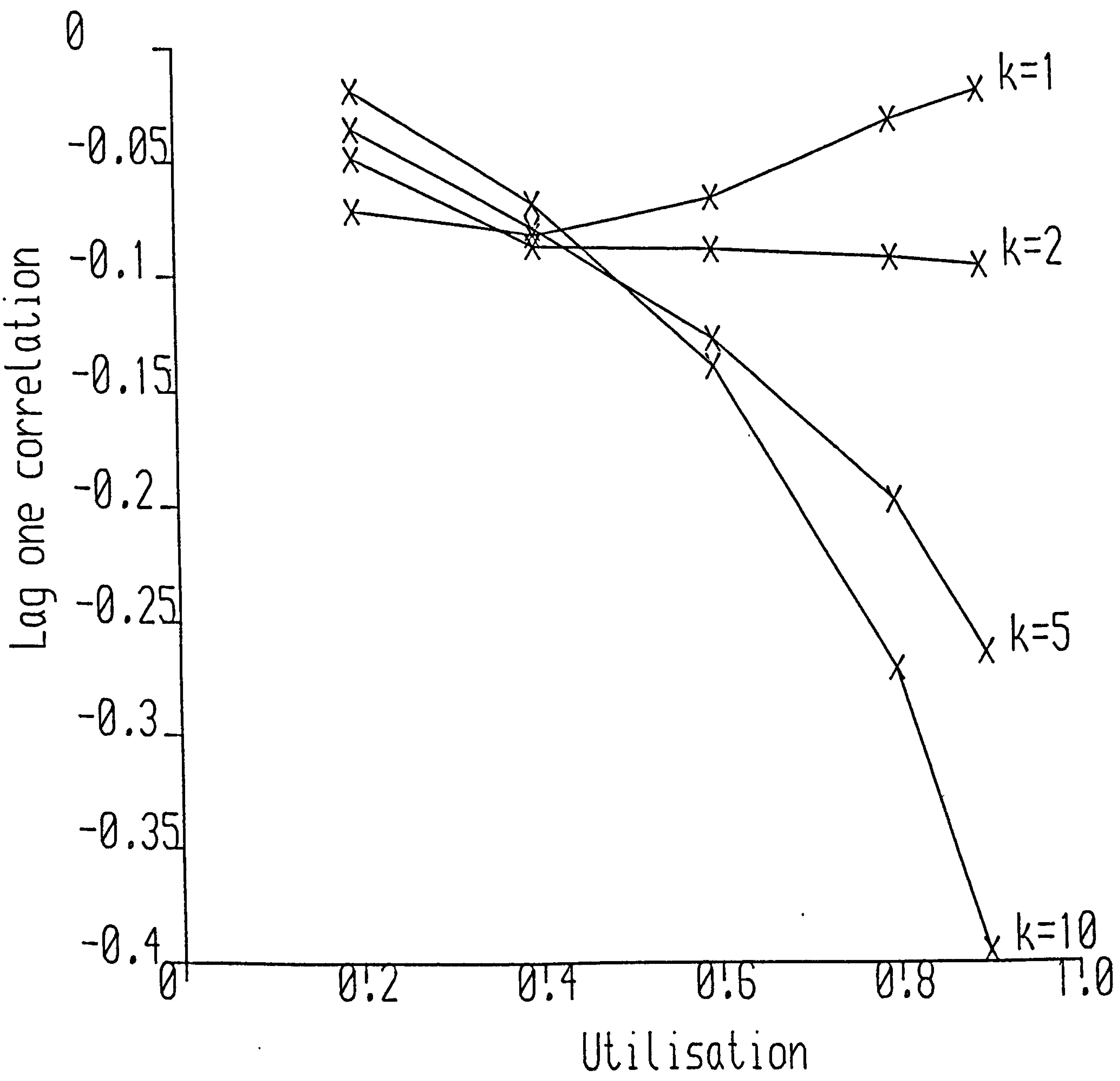
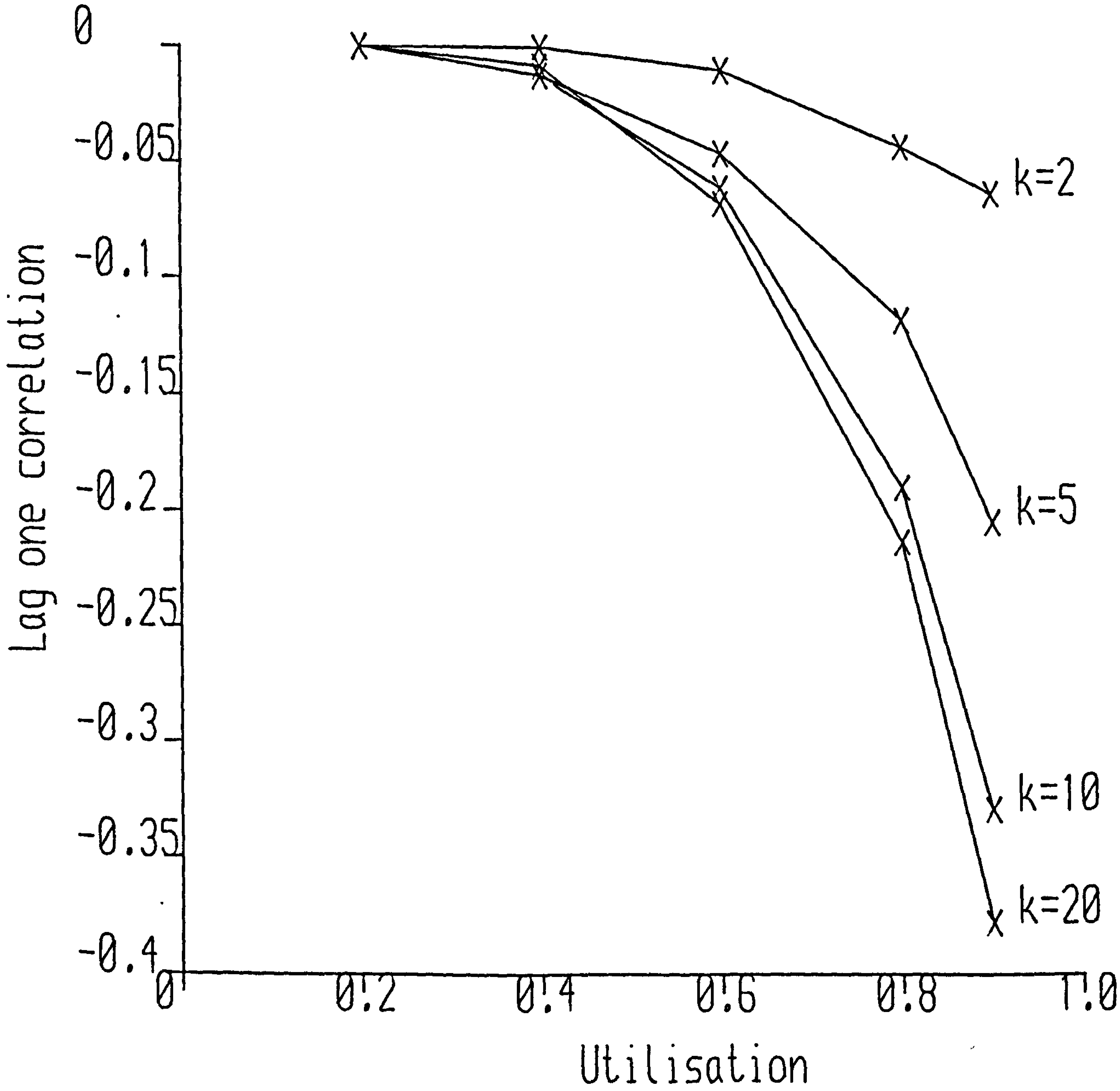


Figure 5.7
Correlation of departures from an $M/E_k/2$ queue



6 APPROXIMATIONS FOR THE MEAN WAIT IN A GI/G/n QUEUE

6.1 INTRODUCTION

The complete decmposition algorithm provides approximations for the throughput time distribution in a multiserver queueing network by a composition of the mean and variance of the queueing times at the individual service centres. The centres were assumed to form GI/G/n queues, and only the first two moments of the arrival time distribution were known. An expression was therefore required for the mean waiting time in a GI/G/n queue which was dependent only on the mean and standard deviation of the inter-arrival times.

6.2 THE MEAN WAIT IN A SINGLE-SERVER QUEUE

Exact results for the average waiting time in a single-server queue are well-known and can be easily calculated for the M/M/1 and M/D/1 systems but there is no closed form result for D/M/1. The Pollazcek-Khintchine formula evaluates the mean wait in an M/G/1 queue. Kendall [35] derived the conditional distribution of the waiting time in a GI/M/n queue, given the queue is not empty, and Prabhu [59] developed a method of solution for the $GI/E_K/1$ system. Lindley [46] obtained an integral equation for the waiting time of a customer in a steady-state GI/G/1 queue. The equation does not yield easily calculable results, and, consequently, approximations to the mean wait in a GI/G/1 queue have been developed. Approximation formulae have the advantage of providing explicit solutions for complex systems without resort to tedious numerical methods or simulation.

To obtain an expression for the average wait at the individual service centres in a single-server queueing network. Shanthikumar [68] compared the accuracy of a number of approximations for the mean wait in a GI/G/1 system, with exact results available for $GE_k/E_1/1$ queues. He concluded that, for greatest accuracy, the form of the approximation used should change with the values of the coefficients of variation of the arrival and service time distributions, and produced a table recommending the approximations for use in different cases.

6.3 THE MEAN WAIT IN A MULTISERVER QUEUE

Exact expressions for the mean wait in queues with more than one server have been established in some cases; these include the M/M/n queue, a complicated expression for the M/D/n system [19], Kendall's derivation for the GI/M/n queue and the result of Mirasol [51] for the $M/G/\infty$ system. By considering the steady-state probability equations, Mayhugh and McCormick [50] obtained numerical results for $M/E_3/2$, and Poyntz and Jackson [58] have given a method of solution for the $E_k/E_1/n$ queue though, except in the most trivial cases, a good deal of computer time is needed to obtain the results. Sakasegawa [65] used this method to tabulate a number of performance values of $E_k/E_2/n$ systems, with n and k ranging from one to twenty. All other exact results for GI/G/n queues take the form of inequalities such as the upper and lower bounds derived by Kingman [38] and Brumelle [9].

Recently a number of good approximations to the mean wait in an M/G/n queue have been proposed. Nozaki and Sheldon [55] and Hokstad [28] independently arrived at the same expression by a consideration of the differential difference equations of the joint equilibrium pobability distribution. Boxma, Cohen and Huffels [7] introduced a quantity termed 'the normed cooperation coefficient' in order to develop an accurate, though complex, approximation. By applying a recursive scheme and the consideration of different cases Tijms, Van Hoorn and Federgruen [72] produced a set of approximations, one of which agreed with that of Boxma et al., and another with Hokstad's expression.

6.4 APPROXIMATIONS FOR THE GI/G/n QUEUE

Several approximation formulae have been suggested for the mean wait in a GI/G/n system; these tend to be heuristic adaptations of results for less complex systems. The approximations known to the author are presented below, and their accuracy and ease of computation are considered.

Kingman

Kingman [37] derived a heavy traffic approximation for the average wait in a GI/M/n system and conjectured that the expression

$$W_K = \frac{s(C_a^2 + u^2 C_s)}{2nu(1-u)},$$

where s was the mean service time, u the utilisation, and C_a and C_s the coefficients of the inter-arrival and service time distributions respectively, should also hold for a generalised service time distribution. This conjecture was later established by Kollerstrom [41].

Page

Page [56] suggested linear interpolation on the coefficients of variation of the arrival and service time distributions of an $E_j/E_k/n$ queue to produce the formula

$$W_P=(1-C_a^2)C_s^2W_{D/M/n}+C_a^2(1-C_s^2)W_{M/D/n}+C_a^2C_s^2W_{M/M/n}$$

where $W_{D/M/n}$, $W_{M/D/n}$ and $W_{M/M/n}$ are the average waiting times in D/M/n, M/D/n and M/M/n queues respectively.

Rosenshine and Chandra

An approximation developed by Fraker [23] for σ_d^2, the variance of the departures from a GI/G/n queue was substituted in Marshall's [48] relation for GI/G/1:

$$W=\frac{\sigma_a^2+2\sigma_s^2-\sigma_d^2}{2a(1-u)}$$

where a was the mean arrival rate, to give

$$W=W_{M/M/1}\cdot Y$$

where

$$Y=C_s^2-0.5(1-C_a^2)-0.5(1-u)C_s^2(1-C_a^2)+0.5(1-C_s^2)$$
$$-(1-u)(1-C_s^2)(1-C_a^2)[0.25C_s^2+C_a^2].$$

Rosenshine and Chandra [63] hypothesised that a useful approximation to the average waiting time in a GI/G/n queue could be obtained by the analogy

$$W_{RC}=W_{M/M/n}\cdot Y;$$

$W_{M/M/n}$ being the average wait in an M/M/n system, given by

$$W_{M/M/n}=sn^{n-1}[n!(1-u)(\sum_{k=0}^{n-1}\frac{(nu)^k}{k!}+\frac{(nu)^n}{n!(1-u)}]^{-1}.$$

Sakasegawa

Sakasegawa [64] simplified the expression obtained on substitution of n=1 in Page's approximation, to produce the the estimate for the waiting time in a GI/G/1 queue

$$W=\frac{su(C_a^2+C_s^2)}{2(1-u)}.$$

From this he developed the similar expression for the mean wait in a GI/G/n system

$$W_S=\frac{su^{\sqrt{2(n+1)}}(C_a^2+C_s^2)}{2nu(1-u)}.$$

Allen-Cunneen

Allen and Cunneen [1] developed an approximation for GI/G/n queues involving the Erlang C formula, this is the probability of not having to wait in an M/M/n queue and is given by

$$C(n,u)=(nu)^n[n!(1-u)(\sum_{k=0}^{n-1}\frac{(nu)^k}{k!}+\frac{(nu)^n}{n!(1-u)})]^{-1}.$$

The Allen-Cunneen approximation is then

$$W_{AC}=\frac{sC(n,u)(C_a^2+C_s^2)}{2n(1-u)}.$$

6.5 A COMPARISON OF THE WAITING TIME APPROXIMATIONS.

The approximations of Kingman and Sakasegawa are of similar form and it can be seen that

$$\frac{W_K}{W_S}=\frac{C_a^2+u^2C_s^2}{u^{\sqrt{2(n+1)}}(C_a^2+C_s^2)}>\frac{u^2}{u^{\sqrt{2(n+1)}}}>1 \quad \text{for } u<1.$$

Thus, under equilibrium conditions, the estimate of the average wait in a GI/G/n queue given by W_K will always be greater than that given by W_S. In a comparison of appoximations obtained using

Sakasegawa's formula with exact values for the mean wait in a number of $E_j/E_k/n$ queues, Page [57] noted that the approximtion over-estimated the average waiting times except in some instances when j was equal to one. Hence, it can be concluded that Sakasegawa's formula will be more accurate than Kingman's in almost every case.

Using tabulated values for the mean wait in M/M/n, M/D/n and D/M/n queues, Page [57] showed his approximation to compare well with exact values for the mean wait in $E_j/E_2/n$ queues, given by Sakasegawa [78], and to be accurate whenever some degree of queueing occurred. However, the necessity of resorting to tables for values of the mean wait in M/D/n and D/M/n queues, limits the practical value of Page's approximation.

Cosmetatoes [14] developed simple approximation formulae for $W_{M/D/n}$ and $W_{D/M/n}$:

$$W_{M/D/n} \simeq 0.5[1+(1-u)(n-1)\frac{\sqrt{(4+5n)}-2}{16un}].W_{M/M/n}$$

and

$$W_{D/M/n} \simeq [1-(1-u)(n-1)\frac{\sqrt{(4+5n)}-2}{4un}](0.5-0.325(2-u)(1-u)).W_{M/M/n}$$

These approximations were shown to be good, with percentage errors of less than 1%, when waiting times were above 0.02 units of service time.

Substitution of Cosmetatoes' formulae and the exact expression for $W_{M/M/n}$ in W_P gives the self-contained approximation formula

$$W_{P'} = W_{M/M/n}.X$$

where

$$X=C_a^2C_s^2+0.5C_a^2(1-C_s^2)[1+(1-u)(n-1)\frac{\sqrt{(4+5n)}-2}{16un}]$$

$$+C_s^2(1-C_a^2)[1-(1-u)(n-1)\frac{\sqrt{(4+5n)}-2}{4un}](0.5-0.325(2-u)(1-u)).$$

It has been observed that the Rosenshine and Chandra approximation can be written in the form

$$W_{RC}=W_{M/M/n}.Y.$$

The Allen-Cunneen approximation can also be expressed in terms of the mean wait in an M/M/n system, becoming

$$W_{AC}=W_{M/M/n}.Z$$

where

$$Z=\frac{C_a^2+C_s^2}{2}.$$

In the special case of $C_a^2=1$, both the Allen-Cunneen approximation and Rosenshine and Chandra's expression reduce to Martin's [48] estimate for the average waiting time in an M/G/n queue.

The approximations $W_{P'}$, W_S and W_{AC} were compared to exact values of the mean wait, given by Sakasegawa [65] for $M/E_2/n$, $E_2/E_2/n$ and $E_5/E_2/n$, and by Hillier and Yu [27] for $E_2/M/n$, with various utilisations and numbers of servers. The percentage errors obtained are shown in Table 6.1. The errors produced by Page's approximation have been included for the $E_2/E_2/n$ system, and it can be seen that the introduction of approximate values for the average waits in M/D/n and D/M/n queues, to give the formula $W_{P'}$, has little effect on the accuracy. Except in the case of the $M/E_2/n$ queue, the approximations are over-estimates of the mean waiting time with

$$W_{P'} < W_{AC} < W_S$$

throughout. When C_a^2 was equal to one some under-estimation of the waiting time was evident, but a consideration of the absolute values of the percentage errors showed $W_{P'}$ to provide the most accurate results.

Table 6.2 compares the percentage errors of $W_{P'}$ and W_{RC} with the exact values of the waiting times. Both approximations showed similar trends, with large errors arising when utilisations were low. However, in most cases the errors exceeding 20% corresponded to average waits of less than 0.1 units of service time and the values of the actual deviations involved were small. Over the tabulated range, the maximum value of the absolute deviations occurred in the estimation of the mean wait in the $E_5/E_2/2$ queue, with a utilisation of 0.95; here the exact value was 3.1068 units of service time, compared with 3.1574 given by $W_{P'}$, a difference of 0.0506 units of average service time. The approximations performed best for queues with values of C_a^2 or C_s^2 equal to one, when $W_{P'}$ produced the better results. In the systems with less variable arrival and service time distributions the errors of both approximations were of similar magnitude.

The waiting times in $E_j/E_k/n$ queues with utilisations of 0.5 or less were low, and the percentage errors of the approximations were not a good indication of their accuracy, as small deviations from the exact values produdced large relative errors. The values of the average percentage errors for utilisations greater than 0.5 were averaged for each value of n and are given in Table 6.3.

Table 6.1 % Errors of the Average Waiting Time Approximations for $E_j/E_K/n$ Queues

			n=2				n=5				n=10			
j	k	u	W_P	$W_{P'}$	W_{AC}	W_S	W_P	$W_{P'}$	W_{AC}	W_S	W_P	$W_{P'}$	W_{AC}	W_S
1	2	0.3	-	0.26	-3.76	21.27	-	2.08	-10.42	129.17	-	38.89	-30.56	8927.78
1	2	0.5	-	-0.43	-2.24	7.43	-	-0.58	-5.81	31.05	-	0.00	-10.00	93.33
1	2	0.7	-	-0.33	-1.10	2.31	-	-0.62	-2.98	6.63	-	-0.69	-4.81	15.12
1	2	0.9	-	-0.11	-0.31	0.36	-	-0.21	-0.83	0.32	-	-0.28	-1.38	-0.02
1	2	0.95	-	-0.05	-0.15	0.14	-	-0.10	-0.40	0.00	-	-0.13	-0.68	-0.42
2	2	0.3	69.23	70.34	110.60	166.58	270.85	253.80	314.29	1031.30	1365.00	1900.00	525.00	16698.20
2	2	0.5	21.76	22.00	41.15	55.07	55.00	53.51	85.11	156.3	141.71	130.32	176.92	515.24
2	2	0.7	7.87	7.72	16.42	20.46	14.81	12.83	27.43	40.00	27.50	26.83	45.67	76.00
2	2	0.9	1.81	1.80	4.09	4.79	2.79	2.68	6.10	7.20	4.10	4.14	8.86	10.00
2	2	0.95	0.72	0.84	1.92	2.22	1.16	1.18	2.75	3.30	1.70	1.71	3.77	4.00
5	2	0.3	-	233.96	552.83	724.53	-	1500.00	3900.00	9100.00	-	-	-	-
5	2	0.5	-	50.09	125.73	147.97	-	132.56	325.58	490.70	-	400.00	1150.00	2600.00
5	2	0.7	-	15.62	11.24	46.07	-	28.48	73.28	90.37	-	53.64	127.27	184.55
5	2	0.9	-	3.50	9.27	10.00	-	4.86	13.72	15.04	-	6.96	19.81	21.39
5	2	0.95	-	1.63	4.28	4.58	-	2.12	6.15	6.57	-	2.76	8.51	8.77
2	1	0.3	-	31.82	87.37	136.11	-	55.56	258.33	826.67	-	733.33	733.33	9733.33
2	1	0.5	-	7.65	33.83	47.00	-	18.79	70.74	137.55	-	54.55	145.45	427.27
2	1	0.7	-	1.50	13.75	17.66	-	4.80	24.28	36.58		10.38	40.25	69.92
2	1	0.9	-	0.20	3.44	4.14	-	0.51	5.47	6.69	-	0.90	7.94	9.43
2	1	0.95	-	0.09	1.62	1.92	-	0.18	2.52	2.93	-	0.28	3.57	3.81

Table 6.2 Exact values and % Errors of the Average Waiting Time Approximations for $E_j/E_k/n$ Queues

			n=2			n=5			n=10		
j	k	u	Exact	$W_{P'}$	W_{RC}	Exact	$W_{P'}$	W_{RC}	Exact	$W_{P'}$	W_{RC}
1	2	0.3	0.0771	0.26	-3.76	0.0048	2.08	-10.42	0.0001	38.89	-30.56
1	2	0.5	0.2556	-0.43	-2.24	0.0415	-0.58	-5.81	0.0060	0.00	-10.00
1	2	0.7	0.7286	-0.33	-1.10	0.1947	-0.62	-2.98	0.0582	-0.28	-1.38
1	2	0.9	3.2073	-0.11	-0.31	1.1533	-0.21	-0.83	0.5085	-0.28	-1.38
1	2	0.95	6.9527	-0.05	-0.15	2.6439	-0.10	-0.40	1.2465	-0.13	-0.68
2	2	0.3	0.0235	69.23	27.80	0.0065	253.80	161.64	0.0000	1900.00	1900.00
2	2	0.5	0.1181	21.76	15.00	0.0141	53.50	32.31	0.0013	130.32	106.50
2	2	0.7	0.4135	7.87	-5.90	0.0988	12.83	5.90	0.0254	26.83	20.90
2	2	0.9	2.0478	1.81	-1.76	0.7193	2.68	0.03	0.3082	4.14	2.38
2	2	0.95	4.5409	0.72	-0.94	1.7080	1.18	-0.24	0.7961	1.71	0.86
5	2	0.3	0.0053	236.85	123.65	0.0001	1500.00	1300.00	0.0000	-	-
5	2	0.5	0.0517	50.00	19.40	0.0043	130.84	121.61	0.0002	373.93	516.11
5	2	0.7	0.2381	15.64	1.31	0.0509	22.17	24.11	0.0110	53.63	83.09
5	2	0.9	1.3655	3.50	-1.03	0.4693	4.86	3.01	0.1955	6.93	8.47
5	2	0.95	3.1068	1.63	-0.64	1.1577	2.12	1.15	0.5326	2.83	3.39
2	1	0.3	0.0396	31.82	43.68	0.0012	55.56	63.63	0.0000	733.33	733.33
2	1	0.5	0.1868	7.65	11.51	0.0229	18.79	42.35	0.0022	54.55	104.55
2	1	0.7	0.6335	1.50	2.37	0.1520	4.80	11.84	0.0395	10.38	26.33
2	1	0.9	3.0910	0.20	0.00	1.0844	0.51	1.95	0.4646	0.90	4.35
2	1	0.95	6.8316	0.09	0.07	2.5686	0.18	0.81	1.1957	0.28	1.84

Table 6.3 Average Absolute % Errors of the Approximations (u>0.5)

	n=2	n=5	n=10
$W_{P'}$	2.788	4.355	2.150
W_{RC}	1.298	4.438	2.919

The approximations appeared to be of similar accuracy; W_{RC} being slightly the better for small n, and $W_{P'}$ for large n.

6.6 CONCLUSIONS

Of the approximation formulae considered, $W_{P'}$ and W_{RC} produced the best approximations to the mean wait in a GI/G/n queue, and were of a form that could be included in the complete decomposition algorithm. The accuracy of the two expressions was comparable in most cases considered. $W_{P'}$ gave better results for queues with coefficients of variation of the inter-arrival and service time distributions close to one, and for queues with large numbers of servers. This approximation was chosen for substitution in the algorithm for the analysis of a multiserver queueing network.

7 AN APPROXIMATION FOR THE VARIANCE OF THE WAITING TIME IN A GI/G/n QUEUE

7.1 INTRODUCTION

In the complete decomposition algorithm a queueing network is decomposed into its service centres which are assumed to form GI/G/n queues. The standard deviation of the throughput time in a network is estimated by a composition of the standard deviations of the waiting times at the individual queues. The algorithm only relies on the first two moments of the transition processes and a formula was required for the variance of the waiting time in a GI/G/n queue that could be expressed in terms of the mean and variance of the inter-arrival and service time distributions.

No exact results are available for the variance of the waiting time in a GI/G/n system, nor are there any good approximations known for queues with more than one server. The conditional waiting time distribution (given that a customer has to wait) in a GI/M/n queue can be shown to be negative exponential and, in heavy traffic conditions, the distribution of the unconditional waiting time can be expected to approach the exponential (Kleinrock [39]). This observation led Kingman [37] to conjecture that, in heavy traffic, the waiting time in a GI/G/n system could be considered to be exponentially distributed with the mean approximated by

$$W_K = \frac{s(C_a^2 + C_s^2)}{2nu(1-u)}$$

where s is the mean service time, u the utilisation, and C_a and C_s are the coefficients of variation of the inter-arrival and service time distributions respectively. The heavy traffic approximation for the standard deviation of the waiting time would also be given by W_K. Kingman [38] later suggested that W_K formed an upper bound for the waiting time in the general case; this has yet to be established.

Some bounds are known for the variance of the waiting time in a GI/G/1 queue. Kingman [36] developed an upper bound, and Mori [52] produced upper and lower bounds. Shanthikumar [68] suggested two approximations for the variance of the wait in the single-server queue. He used weighted combinations of the variances of the waiting times of systems for which exact results, or good approximations, were known. Corresponding results are not available for queues with more than one server and this method could not be applied to give an approximation for the GI/G/n system.

7.2 OBTAINING THE APPROXIMATION FORMULA

In order to obtain an approximation formula for the variance of the waiting time in a GI/G/n queue, linear regression techniques were applied to an extensive range of simulated data. The simulation program is listed in Appendix 1.1. This method was used to produce an expression for the coefficient of variation of the departure process of a GI/G/n system, and has been discussed in Chapter 4. The formula

$$Var(W) \simeq \frac{n^{1/4}}{s^2(1-u)^2}\{u^{\sqrt{3n}}(0.385C_a^2C_s^2+0.188C_a^4)+0.266u^{\sqrt{7n}}C_s^4\}$$

was reached; this provided a good representation of the data, whilst remaining comparatively simple in form.

7.3 THE ACCURACY OF THE APPROXIMATION

Values of the standard deviation of the waiting times given by the approximation were compared with exact results for $E_j/M/1$ queues, and with simulation results for a number of $E_j/E_k/n$ systems. The percentage errors are given in Tables 7.1-7.3.

The approximation appeared most accurate for $E_j/M/1$ queues with u equal to 0.6. In most other cases the errors were less than 10%; the approximation showed a tendency to over-estimate the standard deviation when u was equal to 0.6, and under-estimate it when u was equal to 0.8. Errors of up to 55% were observed for large values of n and j when the utilisation was 0.6; in these cases the values of the standard deviation were less than 0.5 units of the mean service time and the actual deviations involved were small.

Table 7.1 % Errors of the Approximation to the Standard Deviation of the Waiting Time in an $E_j/E_k/1$ Queue

		u=0.6		u=0.8	
j	k	Actual Value	% Error	Actual Value	% Error
1	1	2.291	0.49	4.899	-6.61
2	1	1.713	-0.40	3.711	-6.23
5	1	1.347	-0.40	2.990	-5.41
10	1	1.221	-0.26	2.746	-4.85
∞	1	1.091	0.16	2.500	-4.04
1	2	1.720*	1.07	3.606	-5.84
2	2	1.117*	3.41	2.321*	-1.38
5	2	0.720*	9.14	1.517*	7.51

* simulated value

Table 7.2 % Errors of the Approximation to the Standard Deviation of the Waiting time in an $E_j/E_k/5$ Queue

		u=0.6		u=0.8	
j	k	Simulated Value	% Error	Simulated Value	% Error
1	1	1.590	-2.73	4.633	-14.17
2	1	1.141	-4.11	3.457	-9.12
5	1	0.811	-1.43	2.475	0.83
10	1	0.641	8.19	2.326	-2.51
1	2	1.185	1.61	3.508	-8.09
2	2	0.703	9.91	2.143	-0.34
5	2	0.409	21.53	1.468	0.35

Table 7.3 % Errors of the Approximation to the Standard Devaition of the Waiting time in an $E_j/E_k/10$ Queue

		u=0.6		u=0.8	
j	k	Simulated Value	% Error	Simulated Value	% Error
1	1	1.187	-8.01	4.090	-6.66
2	1	0.690	8.94	2.634	-6.02
5	1	0.434	20.76	2.143	-0.05
10	1	0.283	54.11	1.908	0.59
1	2	0.862	0.19	2.791	-7.17
2	2	0.433	26.15	2.035	-6.20
5	2	0.218	55.72	1.188	8.29

8 APPROXIMATION OF THE THROUGHPUT TIME IN A MULTISERVER QUEUEING NETWORK

8.1 INTRODUCTION

The approximation formulae for the coefficient of variation of the departure process, and the mean and standard deviation of the waiting time distribution of a GI/G/n queue, obtained in Chapters 4,6 and 7, were substituted into the complete decompositon algorithm for the analysis of a multiserver queueing network. Approximations given by the algorithm for the mean and standard deviation of the throughput time were compared with results obtained from the simulation of a number of networks. 95% confidence intervals were estimated for the mean throughput time in networks with one or two servers at the centres. The simulation program is validated in Chapter 3 and listed in Appendix 1.2. The approximation algorithm is described in Chapter 4 and listed in Appendix 2. A breakdown of the simulation and approximation results for the individual queues of the networks is given in Appendix 3.

8.2 THE NETWORKS

Networks of four and ten service centres were used to assess the accuracy of the complete decomposition algorithm. To reduce the number of parameters needed to specify the networks, it was convenient to consider that all the centres of a network had the same service time distributions, utilisations and numbers of servers. External arrivals to the networks were assumed to be Poisson distributed.

Two types of networks were modelled; these were termed flow shops or symmetric shops according to the transition probabilities of the jobs passing through them: in a network of M identical service centres, numbered 1,2,...,M, a flow shop can be defined as one in which each job enters the network at centre 1, visits each service centre in turn, and leaves the network after completion of service at centre M, in a symmetric shop, jobs entering the shop are equally likely to require their first service from any of the centres, on completion of a service the probabilities of a job leaving the network or approaching any of the other service centres are equal.

Defining

p_{ij}= the probability that a job proceeds to centre j after completion of service at centre i $i,j=1,2,...,M$

$p_{M+1,j}$= the probability that a job has its first service at centre j $j=1,2,...,M$

$p_{i,M+1}$= the probability that a job leaves the network after completion of service at centre i $i=1,2,...,M$,

then an M-centre flow shop is represented by the (M+1)x(M+1) transition probability matrix

$$\begin{bmatrix} 0 & 1 & 0 & . & . & 0 \\ 0 & 0 & 1 & . & . & 0 \\ 0 & 0 & 0 & . & . & 0 \\ . & . & . & . & . & . \\ 0 & 0 & 0 & . & . & 1 \\ 1 & 0 & 0 & . & . & 0 \end{bmatrix}$$

and the transition probabilities of an M-centre symmetric shop are represented by the (M+1)x(M+1) matrix

$$\begin{bmatrix} 0 & 1/M & 1/M & . & . & 1/M \\ 1/M & 0 & 1/M & . & . & 1/M \\ 1/M & 1/M & 0 & . & . & 1/M \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ 1/M & 1/M & 1/M & . & . & 0 \end{bmatrix}$$

8.3 THE ACCURACY OF THE COMPLETE DECOMPOSITION ALGORITHM

It can be seen from Tables 8.1-8.7 that the approximation algorithm generally produced good estimates of the mean throughput time in a network. The algorithm performed best when the coefficients of variation of the service times were equal to one and, except for some instances when the coefficients of variation were equal to 0.2, all the approximations were within the 95% confidence intervals estimated for the simulation results. Most of the percentage errors calculated were less than 5%. The larger errors arose for flow shops with E_5 service time distributions and high utilisations; the maximum error of 10.80% was recorded for the network of four two-server centres with utilisations of 0.8.

The approximations to the standard deviation of error of the throughput time distribution were close to the simulated values in the ten centre networks and in the networks with four five-server centres. The approximation was not as accurate for networks with one or two-server centres; the greatest deviations occurred in the single-server networks, and errors of more than 20% were calculated for the flow shop with four single-server centres.

Table 8.1 Flow Shop with Four Single-Server Centres

u	C_s^2	Mean Throughput Time: Simulated	Approx.	% Error	Standard Deviation: Simulated	Approx.	% Error
	0.2	2.024±0.105	2.134	5.43	0.618	0.760	22.98
0.4	0.5	2.281±0.114	2.325	1.93	0.927	1.057	14.02
	1.0	2.653±0.128	2.667	0.53	1.318	1.484	12.59
	0.2	3.776±0.181	3.908	3.50	1.313	1.349	2.74
0.6	0.5	4.641±0.232	4.662	0.45	2.063	1.995	-3.25
	1.0	5.990±0.290	6.000	0.17	3.004	3.012	0.27
	0.2	7.847±0.383	7.528	-4.23	3.845	2.812	-26.86
0.8	0.5	11.029±0.495	10.634	-4.62	5.434	4.514	-16.93
	1.0	15.765±0.776	16.000	1.49	8.031	7.497	-6.65

Table 8.2 Symmetric Shop with Four Single-Server Centres

u	C_s^2	Mean Throughput Time: Simulated	Approx.	% Error	Standard Deviation: Simulated	Approx.	% Error
	0.2	2.220±0.111	2.221	0.05	2.085	2.100	0.72
0.4	0.5	2.427±0.117	2.386	-1.69	2.349	2.344	-0.21
	1.0	2.647±0.132	2.667	0.76	2.743	2.745	0.07
	0.2	4.502±0.225	4.431	-1.58	4.450	4.184	-6.36
0.6	0.5	5.215±0.250	5.012	-3.89	5.408	4.865	-10.04
	1.0	5.848±0.273	6.000	2.60	6.103	6.006	-1.59
	0.2	10.788±0.555	10.131	-6.09	11.689	9.630	-17.61
0.8	0.5	12.766±0.637	12.312	-3.56	13.844	11.909	-13.98
	1.0	16.015±0.774	16.000	-0.09	17.509	15.755	-10.02

Table 8.3 Flow Shop with Four Two-Server Centres

u	C_s^2	Mean Throughput Time Simulated	Approx.	% Error	Standard Deviation Simulated	Approx.	% Error
	0.2	3.488±0.170	3.553	1.86	0.852	0.914	7.28
0.4	0.5	3.588±0.175	3.647	1.64	1.260	1.342	6.51
	1.0	3.791±0.199	3.810	0.56	1.803	1.864	3.38
	0.2	5.929±0.278	6.206	4.67	1.625	1.694	4.25
0.6	0.5	6.508±0.297	6.678	2.55	2.431	2.436	0.21
	1.0	7.502±0.347	7.500	0.03	3.511	3.435	2.16
	0.2	10.625±0.400	11.773	10.80	3.746	3.748	0.05
0.8	0.5	13.609±0.549	13.986	2.77	5.789	5.388	6.93
	1.0	17.482±0.666	17.778	1.69	7.723	7.962	3.09

Table 8.4 Symmetric Shop with Four Two-Server Centres

u	C_s^2	Mean Throughput Time Simulated	Approx.	% Error	Standard Deviation Simulated	Approx.	% Error
	0.2	3.590±0.174	3.580	-0.28	3.220	3.239	0.59
0.4	0.5	3.708±0.175	3.665	-1.16	3.511	3.451	-1.71
	1.0	3.818±0.184	3.810	-0.21	3.827	3.790	-0.97
	0.2	6.430±0.303	6.409	-0.33	6.037	5.842	-3.23
0.6	0.5	6.806±0.295	6.815	0.13	6.571	6.414	-2.39
	1.0	7.582±0.337	7.500	-1.08	7.684	7.348	-4.38
	0.2	13.540±0.711	12.950	-4.36	13.803	12.036	-12.80
0.8	0.5	15.220±0.862	14.750	-3.09	15.712	14.019	-10.78
	1.0	18.309±1.305	17.779	-2.89	19.669	17.333	-11.88

Table 8.5 Flow Shop with Four Five-Server Centres

u	C_s^2	Mean Throughput Time: Simulated	Mean Throughput Time: Approx.	Mean Throughput Time: % Error	Standard Deviation: Simulated	Standard Deviation: Approx.	Standard Deviation: % Error
	0.2	8.073	8.108	0.45	1.806	1.821	0.83
0.4	0.5	8.123	8.127	0.04	2.857	2.860	0.04
	1.0	8.157	8.159	0.02	4.006	4.036	0.02
	0.2	12.647	12.858	1.67	2.866	2.906	1.10
0.6	0.5	12.962	13.064	0.79	4.478	4.668	0.22
	1.0	13.334	13.417	0.22	6.297	6.281	-0.25
	0.2	19.598	20.986	7.08	5.055	4.986	-3.20
0.8	0.5	21.930	22.423	2.25	7.788	7.513	-3.66
	1.0	25.084	24.866	-0.87	10.957	10.522	-3.97

Table 8.6 Symmetric Shop with Four Five-Server Centres

u	C_s^2	Mean Throughput Time: Simulated	Mean Throughput Time: Approx.	Mean Throughput Time: % Error	Standard Deviation: Simulated	Standard Deviation: Approx.	Standard Deviation: % Error
	0.2	8.075	8.111	0.44	7.205	7.257	0.72
0.4	0.5	8.125	8.129	0.05	7.606	7.599	0.09
	1.0	8.183	8.159	0.29	8.170	8.138	0.39
	0.2	12.841	12.903	0.48	11.524	11.551	0.23
0.6	0.5	13.081	13.095	0.10	12.252	12.193	-0.48
	1.0	13.346	13.417	0.53	13.196	13.208	0.09
	0.2	21.741	21.360	-1.78	20.422	19.274	-5.62
0.8	0.5	23.104	22.670	-1.88	22.345	21.060	-5.75
	1.0	24.544	23.967	1.49	24.544	23.967	-2.35

Table 8.7 Symmetric Shop with Ten Three-Server Centres

u	C_s^2	Mean Throughput Time Simulated	Approx.	% Error	Standard Deviation Simulated	Approx.	% Error
	0.2	12.616	12.617	0.01	12.162	12.115	0.39
0.4	0.5	12.649	12.739	0.71	12.293	12.416	1.00
	1.0	12.926	12.941	0.12	12.837	12.908	0.55
	0.2	21.231	21.301	0.33	20.642	20.483	-0.77
0.6	0.5	21.988	22.057	0.31	21.558	21.447	-0.38
	1.0	23.204	23.321	0.50	23.191	23.112	-0.34
	0.2	39.684	39.541	-0.36	39.673	38.25	-3.58
0.8	0.5	44.392	43.412	-1.68	44.392	42.37	-4.54
	1.0	48.699	49.888	2.44	49.657	49.249	-0.82

8.4 CORRELATION OF THE ARRIVAL STREAMS

In the development of the complete decomposition algorithm it was assumed that all the transition processes of a queueing network were renewal. The simulation results listed in Appendix 3 show that significant correlations arose in the arrival streams of most of the networks. The effect of correlation on the accuracy of the approximation of the coefficient of variation of the arrivals is discussed in Chapter 4 (p.44) and Chapter 5.

The greatest degree of correlation was observed in the arrival streams of flow shops with high utilisations and low coefficients of variation of service times; in the single-server networks, correlations were positive and the coefficients of variation of the arrival streams, and the mean and standard deviation of the throughput times were under-estimated, in the networks with more than one server at the centres, the arrivals were negatively correlated and the algorithm under-estimated the network parameters. In the symmetric shops, the correlations of the arrival streams were generally small and positive; the largest values being observed in networks with low utilisations and numbers of servers. As was noted for flow shops, positive correlations coincided with under-estimation of the coefficients of variation of the arrivals by the algorithm but, due to the more complex nature of the networks, this effect was not always reflected in the estimation of the throughput times.

8.5 CONCLUSIONS

The approximations given by the complete decomposition algorithm for the mean and standard deviation of the throughput times in queueing networks were compared with the values obtained from simulations of the networks. The algorithm was most accurate for the more complex networks; the lowest errors were observed for symmetric shops with large numbers of service centres, or large numbers of servers at the centres. The approximation for the mean throughput time was good in all cases considered. The approximation for the standard deviation of the throughput time was reasonable in most instances, though errors of up to 25% were evident in some of the single-server networks.

A significant degree of correlation was noted in the arrival streams of most of the networks simulated, and its presence in flow shops appeared to have a direct affect on the accuracy of the algorithm. The renewal assumption is therefore questionable and a further study of the correlation in a queueing network could result in an improvement of the accuracy of the complete decomposition algorithm.

9 AN INVESTIGATION OF BERTHING REQUIREMENTS AT MALLAIG HARBOUR

9.1 INTRODUCTION

Developments in the fishing industry including the establishment of Exclusive Economic Zones, the over-fishing of some species such as herring, and the growing demand for others, prompted investigations into the adequacy of existing port facilities. The port of Mallaig is considered the only safe harbour on the west coast north of the Clyde. In view of its importance, Crouch and Hogg [20] examined various possible harbour developments. In September 1979 the White Fish Authority [74] produced a detailed economic evaluation for each development option.

The report included a recording of the port operations on a daily basis to appraise existing facilities and to forecast the berthing requirements for 1988. Approximations given by the complete decomposition algorithm for the mean and standard deviation of the times spent by ships in the harbour were compared with the times observed by the White Fish Authority. The algorithm was used to examine how the harbour would cope with increased landings of white fish, and to predict the number of berths needed in 1988 to keep mean waiting times below five minutes.

9.2 BERTHING TIMES IN JUNE 1979

In 1979 the main activity in Mallaig Harbour was the unloading of white fish and shellfish. Most vessels landed their catch between 2 p.m. and 9 p.m. on Tuesdays and Thursdays in order to meet the requirements of the white fish/shellfish auction.

Ships unloaded and took on boxes, fuel and water at the working berths. The number of working berths available was a major determinant of the port's capacity. Although the port had twelve working berths, lack of shelter could mean that as few as six were usable in bad weather. A more critical factor was the congestion in the harbour which caused some berths to be blocked off by parked ships. Even on a quite day, as ships arrived the number of accessible working berths was rapidly reduced to eight.

Some vessels needed to take on ice while they were in the harbour. There was only one ice berth, and bad siting meant that access was often obstructed and long queues formed. Ships could visit the ice berth before unloading at the working berths.

The White Fish Authority observed the operations of the harbour on a 'quiet' auction day, Tuesday 19th June 1979, and on Thursday 21st June 1979, a much busier day. The time of arrival of each vessel, its service times at the berths and its throughput time were recorded. Estimates of the means and standard deviations of the throughput times given by the approximation algorithm, for the two days, were compared with those calculated from the observations.

The ships were assumed to have the use of eight working berths and one ice berth, and the harbour was modelled as a network of two queues with the corresponding numbers of servers. On Tuesday 19th June, eighteen ships entered the harbour in a two-hour period. Eleven of the ships needed ice, only one of these took on ice before unloading at a working berth. This information, and the estimated means and standard deviations of the inter-arrival and service times, allowed the computer algorithm to approximate the mean waiting time at each set of berths, and the mean and standard deviation of the throughput time.

Table 9.1 Waiting Times on Tuesday 19th June 1979.

	Mean Wait (mins.)		Throughput Time (mins.)	
	Working Berths	Ice Berth	Mean	S.D.
Observed	0.0	5.8	34.5	9.6
Approximation	0.5	21.2	53.9	23.0

Table 9.1 shows that the approximation algorithm over-estimated the average waiting time at the ice berth, and the mean and standard deviation of the throughput times. The inaccuracy may be caused by the shortness of the period in which the events occurred; the algorithm relies on the assumption that a queueing system is allowed to build up to steady-state conditions. The figures given by the algorithm can be considerd to be an indication of the way waiting times would increase if arrivals continued at the same rate for a prolonged period.

On the afternoon of Thursday 21st June, the details of forty ships arriving in a four hour period were taken. Twenty three ships required ice, and five took on ice before approaching a working berth. Table 9.2 compares the waiting times approximated by the algorithm with those observed by the White Fish Authority.

Table 9.2 Waiting Times on Thursday 21st June 1979

	Mean Wait (mins.)		Throughput Time (mins.)	
	Working Berths	Ice Berth	Mean	S.D.
Observed	3.2	11.1	69.1	29.7
Approximation	5.1	50.7	91.6	62.1

Lack of steady-state conditions could again account for the over-estimation of the queueing times by the algorithm. During the four hours in which ships arrived, waiting for a working berth was only observed in the second half of the period, and the queue for the ice berth gradually increased, this suggests that if the arrivals had continued at a similar rate for a long period the mean waiting times would have increased. The White Fish Authority calculated the utilisation of the ice berth on the Thursday afternoon to be 1.125, if the events had reached a state of equilibrium with this utilisation there would have been an infinite queue for the ice berth.

It can be concluded that, when a comparatively large number of arrivals occur in a short period of time, steady-state models such as the complete decomposition algorithm do not provide accurate estimates of queueing times. A model that considers the transient behaviour of a queueing system would be more useful for this type of situation.

9.3 BERTHING REQUIREMENTS AS LANDINGS INCREASE

The White Fish Authority forecasted that the landing of white fish and shellfish would have almost doubled by 1988. The approximation algorithm was used to establish the critical factors in determining queueing times in the harbour as the number of arrivals per hour were increased, assuming that arrivals continued for a long enough period of time for steady-state conditions to arise.

It was observed that the average service rates at the working berths and the ice berth were longer on the Thursday than the Tuesday; the White Fish Authority conjectured that this was due to the increased congestion in the harbour on the busier day. In the investigation the estimates of the mean and variance of the service times, the variance of the arrival rates, and the proportions of ships requiring ice were averaged over the two days. Two types of flow through the harbour were considered: in one instance all ships requiring ice were assumed to take on ice after unloading at the working berths, in the other case, ships needing ice were considered equally likely to approach the working berths or the ice berth on entering the harbour.

The estimates given by the approximation algorithm for the queueing times in the harbour with its 1979 facilities are shown in Table 9.3 for up to ten ships arriving an hour. Varying the order in which ships visited the berths had little effect on the average waiting times, but allowing ships to visit the ice berth before the working berths greatly increased the variance of the throughput time. On the two days observed by the White Fish Authority, a small proportion of ships took on ice before unloading. The average value of the mean throughput times recorded for the two days of 51.8 minutes corresponded to the steady-state approximation for an arrival rate of four ships an hour, and the average standard deviation for the two days of 19.7 minutes lay approximately midway between the values of 8.71 and 36.37 given by the algorithm for the two types of route considered.

If more than seven ships were expected an hour the algorithm calculated that they were likely to have to wait at least twenty minutes to take on ice, and an arrival rate of nine ships an hour would cause the ice berth to become overloaded and very long queues to develop.

Table 9.4 gives the approximate waiting times, under steady-state conditions, if two ice berths were available in the harbour. In this case, the queue for the working berth becomes the dominant factor in determining throughput times as the number of arrivals is increased. The harbour would be able to deal comfortably with an average of nine arrivals per hour, if ten or more ships were expected an hour, queueing times for the working berths would be likely to exceed ten minutes, and overloading of the working berths would occur if twelve ships were expected every hour.

9.4 BERTHING REQUIREMENTS FOR 1988

The White Fish Authority attempted to predict the numbers of ice and working berths necessary for Mallaig Harbour to cope with the seasonal fluctuation in the arrivals of vessels unloading white fish and shellfish in 1988. From the data collected on Thursday 21st June 1979, they conjectured that the average waiting time at the working berths corresponded better to the assumption of an M/M/8 queue rather than a GI/G/8 model, and for the 1988 forecast they treated the ice berths and working berths as two separate M/M/n queues. The mean service times at the berths were considered to be those of the busier of the two days observed so that any error in the estimated waiting times would be on the 'safe' side. 70% of the ships were assumed to require ice, and the system was considered to attain steady-state conditions.

The approximation algorithm was used to estimate the numbers of working berths and ice berths needed to keep the average wait at each set of berths below five minutes, under the assumptions adopted by the White Fish Authority. All ships were assumed to unload before taking on ice; the results of the previous section showed that allowing the order of the operations to vary had little effect on the mean throughput time, though it did increase the standard deviation. Table 9.5 shows the berthing requirements in the harbour for up to seventeen ships arriving per hour. The figures for the working berths corresponded to those given by the White Fish Authority, the algorithm also included the ice berths in the model, and was able to provide approximations for the mean and standard deviation of the throughput times.

The peak arrival rate of 15.33 ships per hour was forecast for June 1988. Table 9.5 indicates that the provision of seventeen working berths and four ice berths would keep the average throughput time below seventy minutes throughout the year, but the high values of the standard deviations indicate that the individual throughput times may vary considerably.

Table 9.3 Approximate Waiting Times in the Harbour with Eight Working Berths and One Ice Berth

All ships visiting the working berths first (waiting times in minutes)

Arrivals/hr.	Working Berths u	Working Berths Mean Wait	Ice Berth u	Ice Berth Mean Wait	Throughput Time Mean	Throughput Time S.D.
2	0.17	0.00	0.22	1.58	49.81	8.53
3	0.26	0.01	0.33	2.84	50.57	8.67
4	0.35	0.04	0.44	4.65	51.68	8.71
5	0.44	0.14	0.55	7.19	53.28	8.63
6	0.53	0.40	0.66	11.62	56.18	9.65
7	0.62	0.97	0.77	20.39	61.96	14.41
8	0.70	2.10	0.88	43.74	76.50	31.32
9	0.79	4.77	0.99	447.88	319.69	367.34
10	0.88	12.34	1.10	-	-	-

Ships visiting either berth first (waiting times in minutes)

Arrivals/hr.	Working Berths u	Working Berths Mean Wait	Ice Berth u	Ice Berth Mean Wait	Throughput Time Mean	Throughput Time S.D.
2	0.17	0.00	0.22	1.50	49.77	34.86
3	0.26	0.01	0.33	2.67	50.47	35.46
4	0.35	0.04	0.44	4.33	51.49	36.37
5	0.44	0.15	0.55	6.65	52.97	37.76
6	0.53	0.41	0.66	10.97	55.64	40.43
7	0.62	0.99	0.77	18.63	60.93	46.22
8	0.70	2.08	0.88	40.02	74.23	62.94
9	0.79	4.57	0.99	434.63	311.66	432.93
10	0.88	12.23	1.10	-	-	-

Table 9.4 Approximate Waiting Times in the Harbour with Eight Working Berths and Two Ice Berths

All ships visiting the working berths first (waiting times in minutes)

Arrivals/hr.	Working Berths u	Working Berths Mean Wait	Ice Berth u	Ice Berth Mean Wait	Throughput Time Mean	Throughput Time S.D.
2	0.17	0.00	0.11	0.09	48.93	8.82
3	0.26	0.01	0.16	0.19	49.00	8.79
4	0.35	0.04	0.22	0.34	49.12	8.74
5	0.44	0.14	0.27	0.52	49.32	8.67
6	0.53	0.40	0.33	0.78	49.73	8.59
7	0.62	0.97	0.39	1.11	50.51	8.57
8	0.70	2.11	0.44	1.51	51.88	8.79
9	0.79	4.77	0.49	2.05	54.87	10.13
10	0.88	12.34	0.55	2.75	62.85	16.77
11	0.97	60.44	0.60	3.61	111.46	69.89
12	1.06	-	0.66	4.84	-	-

Table 9.5 Berthing Requirements to keep Mean Waits below Five Minutes

All ships visiting the working berths first (waiting times in minutes)

Arrivals/hr	Working Berths No.	Working Berths Mean Wait	Ice Berth No.	Ice Berth Mean Wait	Throughput Time Mean	Throughput Time S.D.
2	8	0.00	2	0.57	59.12	48.96
3	8	0.04	2	0.81	59.41	48.98
4	8	0.26	2	1.21	60.02	49.02
5	8	0.95	2	1.97	61.25	49.14
6	8	2.86	2	3.04	63.88	49.53
7	9	2.76	2	4.54	64.86	49.45
8	10	2.66	3	0.93	62.23	49.36
9	11	2.55	3	1.33	62.41	49.27
10	12	2.44	3	1.86	62.67	49.19
11	13	2.33	3	2.33	63.07	49.12
12	13	4.85	3	3.46	66.19	49.65
13	14	4.49	3	4.67	66.68	49.50
14	15	4.15	4	1.22	63.93	49.52
15	16	3.86	4	1.60	63.90	49.40
16	17	3.59	4	2.08	63.96	49.29

10 AN APPROXIMATE ANALYSIS OF QUEUEING TIMES IN A MANUFACTURING JOB SHOP

10.1 INTRODUCTION

The Fenner Group of Companies provides power transmission and mining equipment, conveyor systems and oil seals for industry. The Motor Gear and Engineering Company Ltd. based in Chadwell Heath, Essex is a member of the Power Transmission Division and manufactures a wide range of geared motors, couplings and Universal Joints. In December 1973 a computer operated scheduling system known as "MOSQUE" (Motor Operation Scheduler and Queuetime/Urgency Evaluator) was introduced in the Essex works. The main aim of MOSQUE was to determine the despatch date of each production batch by evaluating the time needed for each of its required operations. The queue position, and hence queueing time, of a batch awaiting an operation at a work centre was determined by an algorithm that considered the relative urgency (numbers of days late) of the batch. Standard queueing times were calculated for the work centres as a result of an analysis of the actual queueing times of batches with various degrees of urgency. MOSQUE produced regular print-outs identified by Tab numbers. Two of the most important were:-

Tab 8: A weekly loading summary was tabulated for each work centre and group of similar work centres. This presented an historical record for the previous twenty six weeks of the actual weekly production relative to the available load and planned production, and gave average queueing times for the operations. The load due for the following twenty six weeks was included and compared to the planned production.

Tab 13 At the end of each month a full progress report was printed for each batch in the factory. this included the service and queueing times at work centres where operations had been completed, and the forward schedule of outstanding operations.

10.2 UNIVERSAL JOINT PRODUCTION

The manufacturing job shop scheduled by MOSQUE could be modelled by a multiserver queueing network. Data provided by Tab 8 and Tab 13 issued on day 086 of the Motor Gear Calendar allowed the approximation algorithm to be tested in a practical situation.

To reduce the scale of the analysis, only those work centres needed to produce Universal Joints were considered. With the exception of one form of heat treatment, this group of work centres was involved exclusively in Universal Joint production and could be regarded in isolation from the other activities of the job shop. Tab 13 gave the scheduling details of the 265 batches of Universal Joints in the factory on day 086. Twenty eight work centres were employed in the manufacture of standard Universal Joints. Other work centres were occasionally required for non-standard batches, these were not included in the analysis.

The work centres were grouped according to the type of service they provided (coded by their first letter and number). Table 10.1 records the component work centres of the service groups involved in the production of standard Universal Joints, and the type of operation they performed.

Table 10.1 The Service Groups Employed in Universal Joint Production

Service Group	Component Work Centres	Operation
PO	POAL2 POAL3	Bar-auto turning
1P	1PCL1	Centre turning
2P	2PTL	Turret turning
4P	4PCP1 4PCP2 4PCP3	Capstan lathe turning
P5	P5SH	Shaping
P6	P6CG1 P6CG3 P6CG6 P6CG7 P6CG8	Centreless grinding
P7	P7RMO P7RM2 P7RM8	Routing
Q2	Q2BF1 Q2BF2	Fitting and benchwork
Q3	Q3DR1 Q3DR2 Q3DR3 Q3DR4	Drilling
Q4	Q4LN	Linishing
Q8	Q8HM2 Q8HM3	Milling
E1	E1HT	Heat treatment
E3	E3HR	Induction harden (radyne)
Y8	Y8SC	Sub-contracting

10.3 THE INPUT PARAMETERS FOR THE ALGORITHM

The production of some of the work centres given in Tab 8 bore little relation to either the available load or the planned production. However, the combined planned production of the service groups did appear to correspond to the actual group production. This suggested that if one of the work centres of a service group became overloaded, batches were transferred to one of the other work centres performing a similar type of operation. When applying the approximation algorithm, it seemed reasonable to consider each group of work centres as a single service centre with an appropriate number of servers.

The approximation algorithm was developed for service centres operating a first come first served discipline and was unable to model the earliest due date queueing discipline used to schedule the jobs in the factory. However, in many cases standard lead times meant that work was performed on a first come first served basis.

The Arrival and Service Times

The first two moments of the service time distributions at the service groups were estimated from Tab 13. Large variations were observed in the service times of batches of different sizes. The number of items in a batch varied from 1 to 4000.

To calculate the rates of the external arrivals to the service groups, all the jobs that arrived in the period in which an arrival, with the average throughput time, would be recorded as work in progress on day 086, were considered. On arrival at the factory most jobs spent their first five days waiting for bars or other components to be procured from the stores. A further two days elapsed between the final operation and despatch or delivery to the stores. These delays did not vary and were not included in estimating the average throughput time for the jobs of 40.001 days. To be in the factory on day 086, a job of this length must have arrived at its first service group by day 044, and not later than day 091. Hence, the moments of the distributions of the external inter-arrival times were estimated under the assumption that Tab 13 included all arrivals that occurred between days 044 and 091.

The estimated means and standard deviations of the inter-arrival and service times are listed in Table 10.2.

<u>The Transition Probabilities</u>

The transition probabilities given in Table 10.3 were calculated as the proportions of the departures from a service group that had their next operation at each of the other service groups. In most cases a comparatively large proportion of the departures from one service group approached a particular group of work centres, but a fair amount of variation was evident in the routes taken by different batches.

<u>Number of Servers</u>

The difficulty in quantifying the capacity of the work centres has been stressed in the MOSQUE manual [25]: "At any time neither the firm load nor the capacity at a work centre is known with any precision....... Efficiency and overtime working vary, skilled men can transfer between work centres and machine breakdowns and operator absence can occur at random."

Tab 8 provided details of the planned and actual production at each work centre and service group. From the information issued on day 086, it seemed probable that in some instances the work centres of a group pooled their capacity in order to cope (allowing five hours spare capacity) with demand. In the application of the approximation algorithm, each group of work centres was considered as a single service centre, with a number of servers representing the total number of machines or men at the service group.

In a normal week the work centres were in operation for thirty nine hours. However, efficiency, and the motivation of a bonus scheme meant that a production rate of fifty standard hours a week was common. The approximation algorithm was used to estimate the waiting times at the service groups assuming production rates of forty-five, fifty, and fifty-five standard hours per week. The numbers of servers at the groups were assumed to be the minimum number of men or machines, producing at a rate of fifty standard hours a week, needed to cope with the average weekly loads estimated by the algorithm, these are given in Table 4.

The Y8 service group represented sub-contractors. In theory any number of sub-contractors could have been used and there should have been no waiting time. In practice, Tab 13 showed an average wait of four days for sub-contracting. The heat treatment provided by E1 was also required for batches other than Universal Joints. The number of servers assumed for these two service groups were not particularly meaningful, but they served to give reasonable estimates of the waiting times at the groups.

Table 10.2 The Estimated Inter-Arrival and Service Times

Service Group	Arrival Times(hrs.) Mean	Arrival Times(hrs.) S.d.	Service Times(hrs.) Mean	Service Times(hrs.) S.d.
P0	0.84	1.86	21.43	16.90
1P	1.39	1.66	3.20	3.76
2P	5.83	4.31	6.96	3.88
4P	1.57	1.95	11.30	10.15
P5	0	0	2.60	3.61
P6	4.36	6.53	6.34	7.47
P7	0	0	15.22	9.72
Q2	1.95	2.66	6.61	10.39
Q3	5.71	5.94	15.67	21.67
Q4	0	0	5.31	8.89
Q8	0	0	18.20	19.93
E1	0	0	8.77	14.76
E3	0	0	11.07	12.50
Y8	0	0	58.16	45.46

Table 10.3 The Transition Probability Matrix

FROM \ TO	PO	1P	2P	4P	P5	P6	P7	Q2	Q3	Q4	Q8	E1	E3	Y8
PO	0	0	0	0	0	0.45	0	0.35	0.18	0	0	0	0	0
1P	0	0	0.02	0	0.05	0.27	0	0.19	0.21	0	0	0.06	0	0.03
2P	0	0	0	0	0	0.90	0	0	0.05	0.05	0	0	0	0
4P	0	0.04	0	0	0.08	0.50	0	0.12	0.10	0	0	0	0	0
P5	0	0	0	0	0	0	0	0.78	0.22	0	0	0	0	0
P6	0	0.03	0	0.03	0	0	0	0.08	0.41	0	0.01	0.09	0	0.01
P7	0	0	0	0	0	0.04	0	0.36	0	0.24	0	0.28	0	0
Q2	0	0	0	0	0	0.44	0.02	0	0.12	0	0	0.13	0	0.03
Q3	0	0	0	0	0.01	0.11	0	0.18	0	0.14	0.38	0.11	0	0
Q4	0	0.02	0	0.01	0	0.09	0.12	0.01	0.06	0	0.34	0.35	0	0
Q8	0	0	0	0	0	0	0	0.11	0	0.89	0	0	0	0
E1	0	0	0	0	0	0.28	0	0.08	0	0.02	0	0	0.41	0
E3	0	0	0	0	0	0.07	0	0.37	0	0.16	0	0.02	0	0
Y8	0	0.18	0	0	0	0	0	0	0	0	0	0	0	0

Table 10.4 Machine Requirements for the Average Weekly Loads

Service Group	Average Load(hrs.)	Machine Requirements
P0	127	3 m/cs @ 50hrs/week
1P	3	1 m/c @ 50hrs/week
2P	6	1 m/c @ 50hrs/week
4P	43	1 m/c @ 50hrs/week
P5	2	1 m/c @ 50hrs/week
P6	128	3 m/cs @ 50hrs/week
P7	26	1 m/c @ 50hrs/week
Q2	97	2 m/cs @ 50hrs/week
Q3	222	5 m/cs @ 50hrs/week
Q4	63	2 m/cs @ 50hrs/week
Q8	183	4 m/cs @ 50hrs/week
E1	90	2 m/cs @50hrs/week
E3	46	2 m/cs @50hrs/week
Y8	46	2 m/cs @ 50hrs/week

10.4 APPROXIMATION OF THE WAITING TIMES

The means and standard deviations of the waiting times approximated by the algorithm are compared with those estimated from Tab 13 in Tables 10.5 and 10.6. The waiting times estimated from Tab 13 are hereafter referred to as the observed waiting times. A constant of one day was added to the approximations for the mean wait at each group; this was to account for the time spent moving batches between the service groups; the historical waiting times given in Tab 13 included transfer times, and the forward schedules allowed a day for the moving of batches between departments.

The infinite values for the mean waits in Table 10.5 were given for service groups where the approximated utilisation exceeded one, and indicated that, unless the number of men or machines in operation were increased, a production rate of forty-five hours a week would cause the service groups to become overloaded. The differences in the mean waits given by the algorithm, with productions rates of forty-five and fifty-five standard hours per week, show the actual production rate at a service group to be a critical factor in determining the waiting times.

The observed mean waiting times were within the ranges given by the algorithm for all service groups with utilisations that were estimated to be greater than 0.75 when the production rate of fifty hours per week was assumed. At the service groups with utilisations less than 0.75, the observed mean waits were above the higher end of the range given by the algorithm. It is likely that the waiting times at work centres with small loads would exceed those expected as, in practice, men or machines would be transferred to busier centres, or batches would be allowed to accumulate before production was started.

To obtain more realistic estimates for the average waiting times at the service groups with low utilisations, the algorithm was re-applied with the mean service times adjusted so that all the utilisations that were estimated to be below 0.75 in the original application were equal to 0.75. The results of the second application of the algorithm are given in Tables 10.7 and 10.8.

Table 10.7 shows that four of the observed mean waiting times to be outside the ranges approximated by the algorithm; two were below and two above the values given. Therefore, if one utilisation was to be assumed for all service groups with small loads, the balance of under and over-estimations of the observed mean waits by the algorithm, under the assumption of a utilisation of 0.75, indicate this value to be a reasonable choice.

Tables 10.6 and 10.8 show that similar results were obtained for the standard deviations.

10.5 CONCLUSIONS

The results given by the algorithm suggested that most service groups operated at production rates between forty-five and fifty-five hours a week. The production rate of service groups with low utilisations appeared to be slower, and the assumption of utilisations of 0.75 for these groups provided reasonable estimates of the waiting times.

Lack of information on the actual production rates at the service groups prevented any accurate approximation of the waiting time distributions at the individual groups. However, the algorithm could provide useful estimates of the production rates, and numbers of men and machines, necessary to prevent the development of long queues at the service groups.

Figures 10.9 and 10.10 show that a production rate of fifty standard hours a week may have been close to the average value for the service groups. When this production rate was assumed for all the service groups, some of the approximations given by the algorithm for the mean waiting times at the individual groups were inaccurate but Table 10.7 shows that the approximation for the mean throughput time was a reasonable estimate of the observed time.

Table 10.5 Approximations for the Mean Waiting Times

Service Group	Utilisation @50hrs/wk	Mean Waiting Time (days) @55hrs/wk	@50hrs/wk	@45hrs/wk	Observed
P0	0.85	6.42	11.40	34.42	12.38
1P	0.31	1.17	1.22	1.28	4.23
2P	0.13	1.03	1.04	1.05	6.05
4P	0.89	6.09	11.51	208.41	7.79
P5	0.04	1.01	1.01	1.02	3.11
P6	0.87	1.80	2.56	8.52	4.95
P7	0.54	1.96	2.29	2.85	5.92
Q2	0.67	1.35	1.55	1.94	5.05
Q3	0.92	2.49	5.38	∞	10.28
Q4	0.65	1.49	1.74	2.20	3.81
Q8	0.90	2.63	5.20	∞	5.44
E1	0.93	4.91	12.16	∞	5.90
E3	0.48	1.34	1.49	1.74	2.17
Y8	0.47	1.96	2.34	2.95	4.00
Throughput Time		20.43	33.01		40.00

Table 10.6 Approximations for the Standard Deviations of Waiting Times

Service Group	Utilisation @50hrs/wk	S.D. of Waiting Time (days) @55hrs/wk	@50hrs/wk	@45hrs/wk	Observed
P0	0.85	6.94	11.98	38.00	9.75
1P	0.31	0.52	0.60	0.70	5.48
2P	0.13	0.36	0.40	0.44	10.19
4P	0.89	5.55	10.50	186.71	7.01
P5	0.04	0.38	0.42	0.46	3.35
P6	0.87	1.00	1.91	8.17	8.03
P7	0.54	1.85	2.25	2.86	7.81
Q2	0.67	0.66	0.93	1.40	7.68
Q3	0.92	2.27	5.58	∞	16.73
Q4	0.65	0.88	1.19	1.74	3.72
Q8	0.90	2.29	5.18	∞	8.00
E1	0.93	4.75	12.40	∞	11.14
E3	0.48	0.86	1.08	1.42	1.43
Y8	0.47	2.80	3.42	4.33	6.08
Throughput Time		19.70	36.51		32.69

Table 10.7 Approximations for the Mean Waiting Times

(assuming all utilisations ≥ 0.75)

Service Group	Utilisation @50hrs/wk	Mean Waiting Time (days) @55hrs/wk	@50hrs/wk	@45hrs/wk	Observed
P0	0.85	6.42	11.40	34.42	12.38
1P	0.75	2.23	2.89	4.57	4.23
2P	0.75	3.29	4.52	7.47	6.05
4P	0.89	6.09	11.51	208.41	7.79
P5	0.75	6.19	8.98	15.75	3.11
P6	0.87	1.80	2.56	8.32	4.95
P7	0.75	3.69	5.17	8.99	5.92
Q2	0.75	1.51	1.87	2.72	5.05
Q3	0.92	2.49	5.38		10.28
Q4	0.75	1.73	2.21	3.41	3.81
Q8	0.90	2.63	5.20		5.44
E1	0.93	4.91	12.16		5.90
E3	0.75	2.26	3.12	5.23	2.17
Y8	0.75	5.74	8.71	16.19	4.00
Throughput Time		22.69	36.31		40.00

Table 10.8 Approximations for the Standard Deviations of Waiting Times

(assuming all utilisations ≥ 0.75)

Service Group	Utilisation @50hrs/wk	S.D. of Waiting Time (days) @55hrs/wk	@50hrs/wk	@45hrs/wk	Observed
P0	0.85	6.94	11.98	38.00	9.75
1P	0.75	1.62	2.25	3.78	5.48
2P	0.75	3.08	4.25	16.93	10.19
4P	0.89	5.55	10.50	186.71	7.01
P5	0.75	6.95	9.59	15.71	3.35
P6	0.87	1.00	1.91	8.17	8.03
P7	0.75	3.56	4.96	8.45	7.81
Q2	0.75	0.85	1.29	2.27	7.68
Q3	0.92	2.27	5.58	∞	16.73
Q4	0.75	1.13	1.68	2.98	3.72
Q8	0.90	2.29	5.18	∞	8.00
E1	0.93	4.75	12.40	∞	11.14
E3	0.75	1.91	2.85	5.03	1.43
Y8	0.75	7.28	10.48	18.07	6.08
Throughput Time		26.80	46.60		32.69

Figure 10.9
The Mean Waiting Times

assuming 50 standard hrs/week
and all utilisations ⩾ 0.75

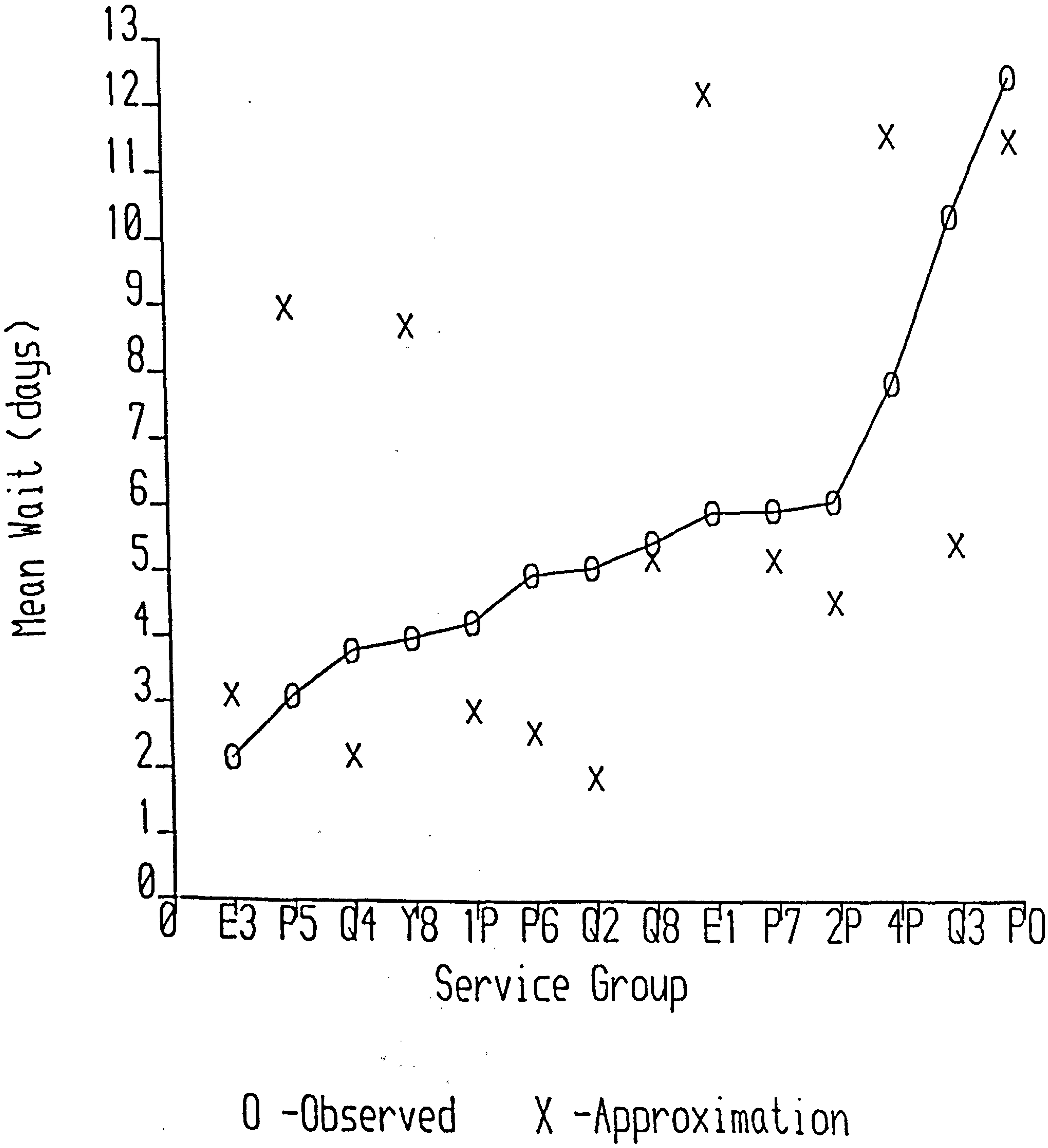
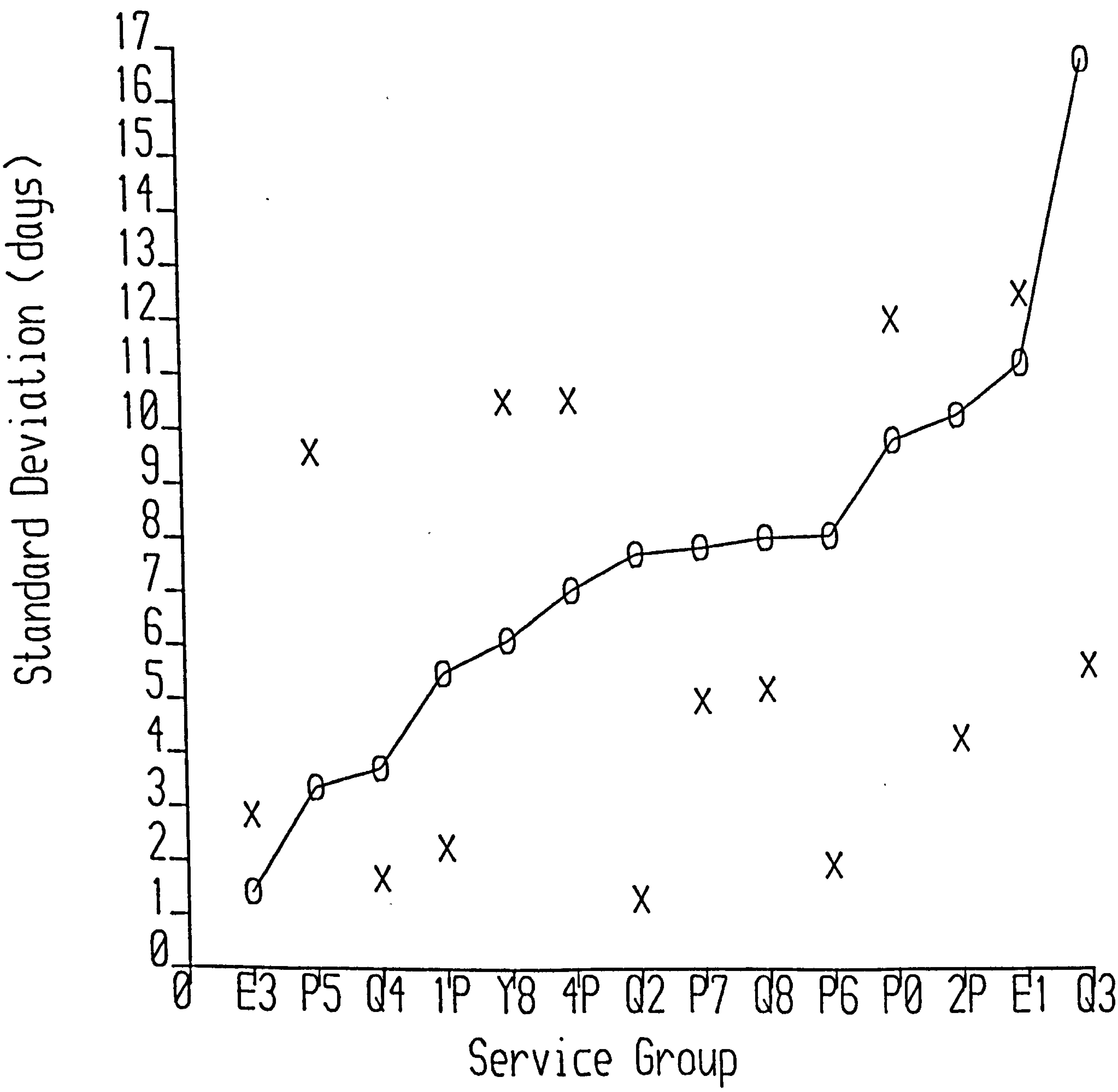


Figure 10.10
Standard Deviations of the Waiting Times

assuming 50 standard hrs/week
and all utilisations ⩾ 0.75



0 -Observed   X -Approximation

APPENDICES

A1 THE SIMULATION PROGRAMS

A1.1 SIMULATION OF A MULTUSERVER QUEUE

The program listed below simulates a multiserver queue with gamma distributed inter-arrival and service times and first come first served queueing discipline. The number of departures to be simulated is determined by a sequential stopping rule and an upper limit can be imposed. The program was written in FORTRAN77 and implemented on the Harris S135.

The input data required is:

N	the number of servers
UTIL	the utilisation
ARCSVQ	the squared coefficient of variation (SCV) of the inter-arrival times
SECVSQ	the SCV of the service times
AVIN	the mean inter-arrival time
RUNT	the maximum number of departures to be simulated.

The output produced is:

AV	the mean inter-departure time
DELI	half the width of the 95% confidence interval for AV
VAR	the variance of the inter-departure times
NUDEP	the number of departures simulated
AWT	the average waiting time
VWT	the variance of the waiting times
AUT(K)	the lag K autocorrelation of the inter-departure times K=1,2,....,10
AQL	the average queue length
VNSYS	the variance of the number in the system
PNW	the probability that an arriving customer does not have to wait
PAB	the probability that all the servers are busy.

```
      DIMENSION AUT(10),TFREE(20),TDEP(100),SUM(10),CORR(10)
   10 CONTINUE
      CALL G05CCF
      READ (15,1)N,UTIL,ARCVSQ,SECVSQ,AVIN
      READ (15,13) RUNT
    1 FORMAT (I4,4F6.3)
   13 FORMAT (F10.2)
      WRITE(16,12) N,UTIL,ARCVSQ,SECVSQ
   12 FORMAT(/'N=',I2,'  U=',F4.2,'  ARCV=',F5.3,'  SECV=',F5.3)
      IF(N.LT.0) STOP

C INITIALISE SYSTEM PARAMETERS
      AVSER=N*UTIL*AVIN
      ARNEX=0.0
      DO 2 I=1,N
    2 TFREE(I)=0.0
      WT=0.0
      VWT=0.0
      DO 33 K=1,10
      SUM(K)=0.0
      CORR(K)=0.0
   33 CONTINUE
      NOCC=0
      DEPREV=0.0
      NUDEP=0
      SDEP=0.0
      SSDEP=0.0
      NDEP=0
      CLOCK=0.0
      TSMALL=0.0
      PNW=0
      TSNSYS=0.0
      TNSYS=0.0
      TQL=0.0
      TAB=0.
      DO 15 I=1,100
   15 TDEP(I)=(AVIN+1)*RUNT

C INITIALISE PARAMETERS FOR THE ARRIVAL AND SERVICE TIME GENERATOR
      PA=1/ARCVSQ
      PS=1/SECVSQ
      TA=PA-1
      TS=PS-1
      CCA=TA*ALOG(2.0)
      CCS=TS*ALOG(2.0)
      UMA=(SQRT(1+4*TA)-1)/(2*TA)
      UMS=(SQRT(1+4*TS)-1)/(2*TS)
      CA=TA*((1-UMA)**TA)+EXP(-UMA*TA)/UMA
      CS=TS*((1-UMS)**TS)+EXP(-UMS*TS)/UMS
      RA=TA*((1-UMA)**TA)/CA
      RS=TS*((1-UMS)**TS)/CS
```

```
C INITIALISE PARAMETERS FOR THE SEQUENTIAL STOPPING RULE
      IC=0
      REF1=0.0
      REF2=0.0
      SUMX=0.0
      SUMY=0.0
      SUMX2=0.0
      SUMY2=0.0
      SUMXY=0.0
      IK=10
      GAM=0.1
C SYSTEM EMPTY SO NEXT EVENT IS AN ARRIVAL
   35 ITYPE=1
C THE START OF A REGENERATIVE CYCLE
C TEST FOR END OF RUN USING SEQUENTIAL STOPPING RULE
      IC=IC+1
      Y=WT-REF1
      X=NUDEP-REF2
      SUMX=SUMX+X
      SUMY=SUMY+Y
      SUMX2=SUMX2+X*X
      SUMY2=SUMY2+Y*Y
      SUMXY=SUMXY+X*Y
      REF1=WT
      REF2=NUDEP
      IF (IC.NE.IK) GOTO 9
      IKM=IK-1
      XN=SUMX/IK
      YN=SUMY/IK
      RN=YN/XN
      S2YN=(SUMY2-IK*(YN*YN))/IKM
      S2XN=(SUMX2-IK*XN*XN)/IKM
      SXYN=(SUMXY-IK*XN*YN)/IKM
      SN=S2YN-2*RN*SXYN+RN*RN*S2XN
      IF (SN.LE.0.0) SN=0.0
      SN=SQRT(SN)
      PIK=IK
      DEL=1.64*SN/(XN*SQRT(PIK))
      IK=IK+10
      TEST=GAM*RN/2.0
      IF (DEL.LT.TEST) GOTO 36
```

```
    9 CONTINUE
C FIND THE NEXT EVENT
      TSMALL=RUNT*(AVIN+1)
      IF(NDEP.LE.0) GOTO 19
      DO 3 I=1,NDEP
      IF(TDEP(I).GT.TSMALL) GOTO 3
      TSMALL=TDEP(I)
      ISTORE=I
    3 CONTINUE
   19 CONTINUE
      ITYPE=2
      IF(ARNEX.LT.TSMALL) ITYPE=1
C THE TYPE OF EVENT IS DECIDED
C  ITYPE=1 ARRIVAL
C  ITYPE=2 DEPARTURE
    8 IF(ITYPE.EQ.2) GOTO 4

C THE NEXT EVENT IS AN ARRIVAL
      ARRIN=PETRE(CCA,UMA,CA,RA,TA)*ARCVSQ*AVIN
      TQL=TQL+(NDEP-NOCC)*(ARNEX-CLOCK)
      TNSYS=TNSYS+NDEP*(ARNEX-CLOCK)
      TSNSYS=TSNSYS+NDEP*NDEP*(ARNEX-CLOCK)
      IF (NDEP.GE.N)   TAB=TAB+ARNEX-CLOCK

C FIND A SERVER
      MENEX=1
      START=TFREE(1)
      IF(N.EQ.1) GOTO 6
      DO 5 I=1,N
      IF(TFREE(I).GT.START) GOTO 5
      START=TFREE(I)
      MENEX=I
    5 CONTINUE
    6 CONTINUE

C FIND THE TIME THE ARRIVAL WILL LEAVE
      TEMP=AMAX1(START,ARNEX)
      W=TEMP-ARNEX
      IF (W.EQ.0.0) PNW=PNW+1
      WT=WT+W
      VWT=VWT+W*W
      TFREE(MENEX)=TEMP+PETRE(CCS,UMS,CS,RS,TS)*AVSER*SECVSQ
      NDEP=NDEP+1
      IF (NDEP.GE.N) THEN
        NOCC=N
      ELSE
        NOCC=NDEP
      ENDIF
      TDEP(NDEP)=TFREE(MENEX)
      CLOCK=ARNEX
      ARNEX=ARNEX+ARRIN
      GOTO 9
    4 CONTINUE
```

```
C THE NEXT EVENT IS A DEPARTURE
      DEPINT=TSMALL-DEPREV
      TQL=TQL+(NDEP-NOCC)*(TSMALL-CLOCK)
      TNSYS=TNSYS+NDEP*(TSMALL-CLOCK)
      TSNSYS=TSNSYS+NDEP*NDEP*(TSMALL-CLOCK)
      IF (NDEP.GE.N)      TAB=TAB+TSMALL-CLOCK
      IF(ISTORE.EQ.NDEP) GOTO 17
      DO 18 I=ISTORE,NDEP
   18 TDEP(I)=TDEP(I+1)
   17 CONTINUE
      IF(ISTORE.EQ.NDEP) TDEP(ISTORE)=2*RUNT
      NDEP=NDEP-1
      IF (NDEP.LT.N) NOCC=NDEP
      NUDEP=NUDEP+1
      DO 29 K=1,10
        SUM(K)=SUM(K)+CORR(K)*DEPINT
   29 CONTINUE
      DO 30 K=1,9
        J=10-K
        M=J+1
        CORR(M)=CORR(J)
   30 CONTINUE
      CORR(1)=DEPINT
      SDEP=SDEP+DEPINT
      SSDEP=SSDEP+DEPINT*DEPINT
      CLOCK=TSMALL
    7 DEPREV=TSMALL
      IF (NUDEP.GE.RUNT) GOTO 36
      IF (NDEP.EQ.0) GOTO 35
      GOTO 9

C END OF RUN
   36 AWT=WT/NUDEP
C ESTIMATE CONFIDENCE INTERVAL FOR THE MEAN WAITING TIME
      ICM1=IC-1
      IF (IC.LT.10) THEN
        DELI=-1
        GOTO 40
      ENDIF
      XN1=SUMX/IC
      YN1=SUMY/IC
      RN1=YN1/XN1
      S2YN1=(SUMY2-IC*(YN1*YN1))/ICM1
      S2XN1=(SUMX2-IC*(XN1*XN1))/ICM1
      SXYN1=(SUMXY-IC*XN1*YN1)/ICM1
      SN1=S2YN1-2*RN1*SXYN1+(RN1**2)*S2XN1
      SN1=SQRT(SN1)
      PIC=IC
      DELI=1.96*SN1/(XN1*SQRT(PIC))
```

```
C CALCULATE THE SYSTEM STATISTICS
   40 VWT=(VWT-AWT*WT)/NUDEP
      AV=SDEP/NUDEP
      VAR=(SSDEP-AV*SDEP)/NUDEP
      ANSYS=TNSYS/CLOCK
      SNSYS=TSNSYS/CLOCK
      VNSYS=SNSYS-ANSYS*ANSYS
      PNW=PNW/NUDEP
      PAB=TAB/CLOCK
      AQL=TQL/CLOCK
      DO 31 K=1,10
        AUT(K)=(SUM(K)/NUDEP-AV**2)/VAR
   31 CONTINUE

C PRINT OUT THE SYSTEM STATISTICS
      WRITE(16,11) AV,VAR,NUDEP,AWT,DELI,VWT
      WRITE (16,41) AQL,VNSYS,PNW,PAB
      WRITE (16,32) (AUT(K), K=1,10)
   11 FORMAT(' AD=',F8.5,' VD=',F8.5,' N=',I8,' AW=',F8.5,' +',F8.5,
     *' VW=',F8.5)
   41 FORMAT(' AQL*',F8.5,' VNSYS=',F8.5,' PNW=',F8.5,' PAB=',F8.5)
   32 FORMAT(' CORR=',,5F9.4)
      GOTO 10
      END

C GENERATION OF THE GAMMA DISTRIBUTED INTER-ARRIVAL AND SERVICE TIMES
      FUNCTION PETRE(CC,UM,C,R,T)
      U=G05CAF(XX)
      E=-ALOG(G05CAF(XX))
      IF (U.GT.R) GOTO 3
    2 X=U*T/R
      IF ((CC-X).GT.E) GOTO 1
      TEMP=T*ALOG(T/X)-T+X
      IF (TEMP.LE.E) GOTO 4
      GOTO 1
    3 X=-ALOG(C*UM*(1-U))/UM
      TEMP*T*ALOG(T/((1-UM)*X))+(1-UM)*X-T
      IF (TEMP.LE.E) GOTO 4
      GOTO 1
    4 PETRE=X
      RETURN
      END
```

A1.2 SIMULATION OF A NETWORK OF MULTISERVER QUEUES

The program listed below simulates a network of multiserver queues with gamma distributed inter-arrival and service times and first come first served queueing disciplines. Routes through the network are determined by a transition probability matrix. The number of departures to be simulated is determined by a sequential stopping rule and an upper limit can be imposed. The program was written in FORTRAN77 and implemented on the Harris S135.

The input data required was:

M	the number of service centres
CE	the SCV of the external inter-arrival times
EVE	the mean external inter-arrival time
RUNT	the maximum number of departures to be simulated
N(I)	the number of servers at centre I I=1,2,....,M
CS(I)	the SCV of the service times at centre I I=1,2,....,M
RHO(I)	the utilisation of centre I I=1,2,....,M
AVE(I)	the arrival rate to centre I I=1,2,....,M
TP	the transition probability matrix.

The output produced was:

ATT	the average throughput time
DELI	half the width of the 95% confidence interval for ATT
SDT	the standard deviation of the throughput times
TNDEP	the number of departures simulated
UTIL(I)	the utilisation of centre I I=1,2,....,M
CVA(I)	the SCV of the arrivals to centre I I=1,2,....,M
AUT(I)	the lag one autocorrelation of the arrivals to centre I I=1,2,....,M

AWK(I) the average waiting time at centre I I=1,2,....,M

SD(I) the standard deviation of the waiting time at centre I. I=1,2,....,M.

```
      DIMENSION N(11),CS(11),RHO(11),TP(11,11),TFREE(11,20),NDEP(11)
     *TDEP(11,100),PS(11),HS(11),CCS(11),BMS(11),GCS(11)
     *AVSER(11),RS(11),W(500), XT(500),IND(11,100),LEFT(500)
     *TAT(11),VTA(11),TNAK(11),OAT(11),CVA(11),TIA(11)
     *AUT(11),CORR(11),AWK(11),TI(11),SD(11),UTIL(11),AVE(11)
   10 CONTINUE
      CALL G05CCF
C READ INPUT DATA
      READ(15,100) M,CE,EVE,RUNT
      IF (M.LE.0) STOP
      WRITE (16,101) M,CE
      DO 1 I=1,M
        READ (15,102) N(I),CS(I),RHO(I),AVE(I)
        WRITE (16,103) N(I), CS(I),RHO(I)
    1 CONTINUE
      MP1=M+1
      DO 2 I=1,MP1
        READ (15,104) (TP(I,J), J=1,MP1)
        WRITE (16,104) (TP(I,J), J=1,MP1)
    2 CONTINUE

C INITIALISE THE SYSTEM PARAMETERS
      DO 27 K=1,M
      AVSER(K)=N(K)*RHO(K)/AVE(K)
      TAT(K)=0.0
      VTA(K)=0.0
      OAT(K)=0.0
      TNAK(K)=0
      AUT(K)=0.0
      CORR(K)=0.0
      AWK(K)=0.0
      SD(K)=0.0
      TI(K)=0.0
      NDEP(K)=0
   27 CONTINUE
      DO 24 I=1,500
        W(I)=0.0
        XT(I)=0.0
   24 CONTINUE
      DO 40 I=1,MP1
        DO 4 J=1,100
          TDEP(I,J)=2*RUNT
    4   CONTINUE
        DO 5 J=1,20
          TFREE(I,J)=0.0
    5   CONTINUE
   40 CONTINUE
    3 CONTINUE
      ARNEX=0.0
      CLOCK=0.0
      EVT=0.0
      WT=0.0
      TT=0.0
      VWT=0.0
      NSYS=0.0
      TNSYS=0.0
      TNDEP=0.0
      ICT=0
```

```
C INITIALISE PARAMETERS FOR THE ARRIVAL AND SERVICE TIME GENERATOR
      CS(MP1)=CE
      DO 3 I=1,MP1
        PS(I)=1/CS(I)
        HS(I)=PS(I)-1
        CCS(I)=HS(I)*ALOG(2.0)
        BMS(I)=(SQRT(1+4*HS(I))-1)/(2*HS(I))
        GCS(I)=HS(I)*((1-BMS(I))**HS(I))+EXP(-BMS(I)*HS(I))/BMS(I)
        RS(I)=HS(I)*((1-BMS(I))**HS(I))/GCS(I)
    3 CONTINUE

C INITITIALISE PARAMETERS FOR THE SEQUENTIAL STOPPING RULE
      IC=0
      REF1=0.0
      REF2=0.0
      SUMX=0.0
      SUMY=0.0
      SUMX2=0.0
      SUMY2=0.0
      SUMXY=0.0
      IK=10
      GAM=0.1

C SYSTEM IS EMPTY SO THE NEXT EVENT IS AN ARRIVAL
C START OF A REGENERATIVE CYCLE
C TEST FOR END OF RUN USING SEQUENTIAL STOPPING RULE
    6 IC=IC+1
      Y=TT-REF1
      X=TNDEP-REF2
      SUMX=SUMX+X
      SUMY=SUMY+Y
      SUMX2=SUMX2+X*X
      SUMY2=SUMY2+Y*Y
      SUMXY=SUMXY+X*Y
      REF1=TT
      REF2=TNDEP
      IF (IC.NE.IK) GOTO 16
      IKM=IK-1
      XN=SUMX/IK
      YN=SUMY/IK
      RN=YN/XN
      S2YN=(SUMY2-IK*(YN*YN))/IKM
      S2XN=(SUMX2-IK*XN*XN)/IKM
      SXYN=(SUMXY-IK*XN*YN)/IKM
      SN=S2YN-2*RN*SXYN+RN*RN*S2XN
      IF (SN.LE.0.0) SN=0.0
      SN=SQRT(SN)
      PIK=IK
      DEL=1.96*SN/(XN*SQRT(PIK))
      IK=IK+10
      TEST=GAM*RN/2.0
      IF (DEL.LT.TEST) GOTO 17

C TAG ARRIVAL WITH A JOB NUMBER
      JC=0
   25 JC=JC+1
      ID=0
      JN=JC
```

```
C EXTERNAL ARRIVAL TO THE SYSTEM
    7 J=MP1
      TNSYS=TNSYS+NSYS*(ARNEX-CLOCK)
      NSYS=NSYS+1
C DETERMINE WHERE THE ARRIVAL GOES TO
      CALL ASSIGN(K,J,MP1,TP,ICT)
      X=PETRE(CCS,BMS,GCS,RS,HS,MP1,ICT)
      ARRIN=X*CE*EVE
      CLOCK=ARNEX
      EVT=ARNEX
      ARNEX=ARNEX+ARRIN

C ARRIVAL TO CENTRE K
C FIND A SERVER
    8 MENEX=1
      START=TFREE(K,1)
      IF (N(K).EQ.1) GOTO 9
      DO 11 I=1,N(K)
        IF (TFREE(K,I).GT.START) GOTO 11
        START=TFREE(K,I)
        MENEX=I
   11 CONTINUE
    9 CONTINUE

C UPDATE SYSTEM PARAMETERS AND FIND TIME ARRIVAL WILL LEAVE K
      X=EVT-OAT(K)
      TAT(K)=TAT(K)+X
      VTA(K)=VTA(K)+X*X
      AUT(K)=AUT(K)+X*CORR(K)
      CORR(K)=X
      TI(K)=TI(K)+DIM(EVT,START)
      TEMP=AMAX1(START,EVT)
      X=PETRE(CCS,BMS,GCS,RS,HS,K,ICT)
      TFREE(K,MENEX)=TEMP+X*AVSER(K)*CS(K)
      Y=TEMP-EVT
      AWK(K)=AWK(K)+Y
        SD(K)=SD(K)+Y*Y
      W(JN)=W(JN)+Y
        XT(JN)=XT(JN)+(TFREE(K,MENEX)-TEMP)+Y
      NDEP(K)=NDEP(K)+1
      TDEP(K,NDEP(K))=TFREE(K,MENEX)
      IND(K,NDEP(K))=JN
      OAT(K)=EVT
      TNAK(K)=TNAK(K)+1

C FIND THE NEXT EVENT
   16 EVT=RUNT*(EVE+1)
      DO 12 K=1,M
        DO 13 I=1,NDEP(K)
          IF (TDEP(K,I).GT.EVT) GOTO 13
          EVT=TDEP(K,I)
          JN=IND(K,I)
          ISTORE=I
          KSTORE=K
   13   CONTINUE
   12 CONTINUE
        IF (ARNEX.LT.EVT) THEN
          IF (ID.EQ.0) GOTO 25
          JN=LEFT(ID)
          ID=ID-1
          GOTO 7
        ENDIF
```

```
C SERVICE COMPLETION AT CENTRE L
      L=KSTORE
      IF (ISTORE.EQ.NDEP(L)) GOTO 14
      DO 15 I=ISTORE,NDEP(L)
        TDEP(L,I)=TDEP(L,I+1)
        IND(L,I)=IND(L,I+1)
   15 CONTINUE
   14 IF (ISTORE.EQ.NDEP(L)) TDEP(L,ISTORE)=2*RUNT
      NDEP(L)=NDEP(L)-1

C DETERMINE WHERE DEPARTURE GOES AND UPDATE THE SYSTEM PARAMETERS
      CALL ASSIGN(K,L,MP1,TP,ICT)
      IF (K.LT.MP1) GOTO 8
      TNSYS=TNSYS+NSYS*(EVT-CLOCK)
      WT=WT+W(JN)
        TT=TT+XT(JN)
      VWT=VWT+XT(JN)*XT(JN)
      NSYS=NSYS-1
      W(JN)=0.0
      XT(JN)=0.0
      ID=ID+1
      LEFT(ID)=JN
      TNDEP=TNDEP+1
      IF (TNDEP.GT.RUNT) GOTO 17
      CLOCK=EVT
      IF (NSYS.EQ.0) GOTO 6
      GOTO 16

C END OF RUN
   17 ATT=TT/TNDEP
C ESTIMATE THE CONFIDENCE INTERVAL FOR THE AVERAGE THROUGHPUT TIME
      ICM1=IC-1
      IF (IC.LT.10) THEN
        DELI=-1.0
        GOTO 30
      ENDIF
      XN1=SUMX/IC
      YN1=SUMY/IC
      RN1=YN1/XN1
      S2YN1=(SUMY2-IC*(YN1*YN1))/ICM1
      S2XN1=(SUMX2-IC*XN1*XN1)/ICM1
      SXYN1=(SUMXY-IC*XN1*YN1)/ICM1
      SN1=S2YN1-2*RN1*SXYN1+(RN1**2)*S2XN1
      SN1=SQRT(SN1)
      PIC=IC
      DELI=1.96*SN1/(XN1*SQRT(PIC))

C CALCULATE THE SYSTEM STATISTICS
   30 VWT=(VWT-ATT*TT)/TNDEP
      SDT=SQRT(VWT)
      DO 26 K=1,M
        UTIL(K)=1-TI(K)/(CLOCK*N(K))
        AWK(K)=AWK(K)/TNAK(K)
        TIA(K)=TAT(K)/TNAK(K)
        VTA(K)=(VTA(K)-TIA(K)*TAT(K))/TNAK(K)
        CVA(K)=VTA(K)/(TIA(K)*TIA(K))
        AUT(K)=(AUT(K)/TNAK(K) -TIA(K)*TIA(K))/VTA(K)
        SD(K)=(SD(K)-TNAK(K)*AWK(K)*AWK(K))/TNAK(K)
        SD(K)=SQRT(SD(K))
   26 CONTINUE
```

```
C PRINT OUT THE SYSTEM STATISTICS
      WRITE (16,105) ATT,DELI,SDT,TNDEP
      WRITE (16,106) (UTIL(K), K=1,M)
      WRITE (16,107) (CVA(K), K=1,M)
      WRITE (16,108) (AUT(K), K=1,M)
      WRITE (16,109) (AWK(K),K=1,M)
      WRITE (16,110) (SD(K), K=1,M)
      WRITE (16,111) ICT
  100 FORMAT (I2,2F6.3,F9.2)
  101 FORMAT (/,' M=',I2,' CE=',F6.3)
  102 FORMAT (I2,3F6.3)
  103 FORMAT (' N=',I2,' CS=',F6.3,'  RHO=',F6.3)
  104 FORMAT (11F6.3)
  105 FORMAT (' AT=',F8.5,' +',F8.5,' SD=',F8.5,' TN=',F10.1)
  106 FORMAT ('  UTIL=:',11F7.4)
  107 FORMAT (' CVA=:',11F7.4)
  108 FORMAT (' CORR=:',11F7.4)
  109 FORMAT (' AWK=:',11F7.4)
  110 FORMAT (' SDK=:',11F7.4)
  111 FORMAT ('NO. OF R.V.S USED=' I10)
      GOTO 10
      END

C GENERATION OF THE GAMMA DISTRIBUTED INTER-ARRIVAL AND SERVICE TIMES
      FUNCTION PETRE(CC,BM,C,R,H,K,ICT)
      DIMENSION CC(10),BM(10),C(10),R(10),H(10)
   18 U=G05CAF(XX)
      ICT=ICT+1
   28 V=G05CAF(XX)
      ICT=ICT+1
      E=-ALOG(V)
      IF (U.GT.R(K)) GOTO 20
   19 X=U*H(K)/R(K)
      IF ((CC(K)-X).GT.E) GOTO 18
      Y=H(K)/X
      FEMP=H(K)*ALOG(H(K)/X)-H(K)+X
      IF (FEMP.LE.E) GOTO 21
      GOTO 18
   20 Y=C(K)*BM(K)*(1-U)
      X=-ALOG(Y)/BM(K)
      Y=H(K)/((1-BM(K))*X)
      FEMP=H(K)*ALOG(H(K)/((1-BM(K))*X))+(1-BM(K))*X-H(K)
      IF (FEMP.LE.E) GOTO 21
      GOTO 18
   21 PETRE=X
      RETURN
      END

C ASSIGN THE NEXT SERVICE CENTRE TO BE VISITED
      SUBROUTINE ASSIGN(K,J,MP1,TP,ICT)
      DIMENSION TP(11,11)
      U=G05CAF(XX)
      ICT=ICT+1
      DO 22 I=1,MP1
        U=U-TP(J,I)
        IF (U.LE.0.0) GOTO 23
   22 CONTINUE
   23 K=I
      RETURN
      END
```

A2 THE APPROXIMATION PROGRAM

The program listed below performs an approximate analysis of a multiserver queueing network. The program was written in BASIC to be used interactively on the Harris S135.

The input data required is:

N	the number of service centres
CUST(I)	the rate of external arrivals to centre I I=1,2,....,M
CE(I)	the SCV of external arrival to centre I I=1,2,....,M
T	the matrix of transition probabilities
AVST(I)	the average service time at centre I I=1,2,....,M
SDST(I)	the standard deviation of the service times at centre I I=1,2,....,M
NS(I)	the number of servers at centre I I=1,2,....,M.

The output produced is:

A(I)	the internal arrival rate to centre I I=1,2,....,M
RHO(I)	the utilisation of centre I I=1,2,....,M
CA(I)	the SCV of the arrivals to centre I I=1,2,....,M
AVW(I)	the average wait at centre I I-1,2,....,M
SDW(I)	the standard deviation of the wait at centre I I=1,2,....,M
XX	the average waiting time in the network
THRU	the average throughput time
TSD	the standard deviation of the throughput times.

```
110 OPTION BASE=1
200 DIM P(15,15),W(15,15),A(15,1),CUST(15,1),T(15),NS(15),AVST(15)
210 DIM SDST(15),AVW(15),U(15),Q(15,15),RHO(15),CA(15,1),CS(15)
220 DIM B(15,1),SDW(15),EN(15),VAR(15),COV(15,15),ATT(15),PT(15,15)
230 DIM XSCA(15),CE(15),RCA(15),RCS(15),E(15),F(15),G(15)
300 A$="####.###"
310 B$="Y"
320 E$="###"
330 F$=" ENTER THE FLOWS FROM CENTRE ### AS PROPORTIONS"
340 G$=" CENTRE ### NEEDS AT LEAST ### SERVERS"
350 H$=" AVERAGE OVERALL WAIT = ####.####"
351 J$="\                          \####.####"
360 I$=" AVERAGE THROUGHPUT TIME = ####.####"

370 REM READ THE INPUT DATA
400 PRINT " ENTER THE NUMBER OF SERVICE CENTRES IN YOUR SYSTEM"
410 INPUT " NUMBER OF CENTRES =",N
600 PRINT " ENTER THE ARRIVAL RATES OF EXTERNAL CUSTOMERS AT EACH CENTRE"
610 PRINT " PUT EACH NO. ON A SEPERATE LINE "
620 FOR I=1 TO N
621 INPUT CUST(I,1)
623 NEXT I
630 PRINT " THE ARRIVAL RATES ENTERED ARE :-"
640 FOR I=1 TO N
641 PRINT USING A$, CUST(I,1);
642 NEXT I
650 INPUT " DO YOU WISH TO AMEND THESE VALUES? ENTER Y/N",C$
660 IF C$=B$ GOTO 600
670 PRINT " ENTER THE SCV OF EXTERNAL ARRIVALS TO EACH CENTRE"
680 FOR I=1 TO N
681 INPUT CE(I)
682 NEXT I
690 PRINT " THE SCVS ENTERED ARE :- "
700 FOR I=1 TO N
701 PRINT USING A$, CE(I);
702 NEXT I
710 INPUT " DO YOU WANT TO CHANGE THESE VALUES? ENTER Y/N",C$
720 IF C$=B$ GOTO 670
730 FOR I=1 TO N
740 PRINT USING F$,I
750 FOR J=1 TO N
751 INPUT T(J)
752 NEXT J
760 SP=0.0
770 FOR J=1 TO N
780 SP=SP+T(J)
790 NEXT J
800 IF SP>1.0 THEN DO: PRINT " THE PROPORTIONS SUM TO MORE THAN ONE"
810 PRINT " PLEASE RE-ENTER THIS LINE OF DATA"
820 GOTO 740
830 DOEND
840 PRINT " YOU HAVE ENTERED THE FLOWS:-"
850 FOR J=1 TO N
851 PRINT USING A$, T(J);
852 NEXT J
860 INPUT " DO YOU WISH TO AMEND THESE VALUES? ENTER Y/N",C$
870 IF C$=B$ GOTO 740
```

```
880 FOR J=1 TO N
890 P(I,J)=T(J)
900 NEXT J
910 NEXT I
1000 MAT W=IDN
1010 MAT PT=TRN(P)
1020 MAT W=W-PT
1030 MAT W=INV(W)
1040 MAT A=W*CUST
1050 PRINT " THE OVERALL FLOW, INTERNAL AND EXTERNAL, THROUGH EACH CENTRE"
1060 PRINT " OF YOUR SYSTEM IS AS INPUT IN SEQUENCE:-"
1070 FOR I=1 TO N
1071 PRINT USING A$, A(I,1)
1072 NEXT I
1200 PRINT " ENTER THE AVERAGE SERVICE TIME AT EACH CENTRE"
1210 FOR I=1 TO N
1211 INPUT AVST(I)
1212 NEXT I
1213 FOR I=1 TO N
1214 PRINT USING A$, AVST(I);
1215 NEXT I
1230 INPUT " DO YOU WISH TO CHANGE THESE VALUES? ENTER Y/N",C$
1240 IF C$=B$ GOTO 1200
1400 PRINT " ENTER THE S.D. OF SERVICE TIME AT EACH CENTRE"
1410 FOR I=1 TO N
1411 INPUT SDST(I)
1412 NEXT I
1420 PRINT " THE S.D.'S ENTERED ARE:-"
1430 FOR I=1 TO N
1431 PRINT USING A$, SDST(I);
1432 NEXT I
1440 INPUT " DO YOU WISH TO ALTER THESE VALUES? ENTER Y/N",C$
1450 IF C$=B$ GOTO 1400
1600 FOR I=1 TO N
1610 M=INT(A(I,1)*AVST(I))+1
1620 PRINT USING G$,I,M
1630 NEXT I
1640 REM THE MINIMUM NUMBER OF SERVERS IS CALCULATED AS A GUIDE
1800 PRINT " ENTER THE NUMBER OF SERVERS IN EACH CENTRE"
1810 FOR I=1 TO N
1811 INPUT NS(I)
1812 NEXT I
1820 PRINT " THE NUMBER OF SERVERS ENTERED ARE:-"
1830 FOR I=1 TO N
1831 PRINT USING E$,NS(I);
1832 NEXT I
1840 INPUT " DO YOU WISH TO AMEND THESE VALUES? ENTER Y/N".C$
1850 IF C$=B$ GOTO 1800
2000 FOR I=1 TO N
2010 RHO(I)=A(I,1)*AVST(I)/NS(I)
2020 X=SDST(I)/AVST(I)
2030 CS(I)=X*X
2040 NEXT I
2050 PRINT" THE UTILISATIONS OF THE CENTRES ARE:-"
2060 FOR I=1 TO N
2061 PRINT USING A$;RHO(I)
2062 NEXT I
2063 INPUT " TYPE 1 TO CONTINUE",I
```

```
2070 REM ITERATION TO FIND THE COEFFICIENTS OF VARIATION OF THE ARRIVALS
2080 EPS=0.0000001
2090 FOR I=1 TO N
2100 RCS(I)=SQR(CS(I))
2110 E(I)=(0.04*NS(I)-1.37)*(1-RHO(I))
2120 F(I)=0.65*(1-RHO(I))*RCS(I)
2130 G(I)=1.13*RHO(I)*RHO(I)*(1-CS(I))/NS(I)
2160 RCA(I)=0.0
2170 CA(I,1)=0.0
2190 NEXT I
2200 K=0
3400 K=K+1
3410 IF K<=1000 GOTO 4250
3420 PRINT " THE ITERATIONS EXCEED 1000"
3430 STOP
4250 FOR I=1 TO N
4260 SUM=0.0
4270 FOR J=1 TO N
4280 SUM=SUM+P(J,I)*A(J,1)*(P(J,I)*(E(J)*(1-CA(J,1))+F(J)*(1-RCA(J))-G(J))+1.0)
4290 NEXT J
4300 SUM=SUM+CUST(I,1)*CE(I)
4310 XSCA(I)=SUM/A(I,1)
4320 NEXT I
4330 M=0
4340 FOR I=1 TO N
4350 IF ABS(XSCA(I)-CA(I,1))<EPS GOTO 4400
4360 M=M+1
4370 CA(I,1)=XSCA(I)
4380 IF CA(I,1)<0.0 CA(I,1)=0.0
4390 RCA(I)=SQR(CA(I,1))
4400 NEXT I
4410 IF M>0.0 GOTO 3400
4411 PRINT " THE COEFFICIENT OF VARIATION OF ARRIVAL TO EACH CENTRE IS:-"
4420 FOR I=1 TO N
4421 PRINT USING A$, CA(I,1)
4422 NEXT I
4423 INPUT " TYPE 1 TO CONTINUE",I

4600 REM ESTIMATION OF THE MEAN WAITNG TIME AT EACH CENTRE
4620 FOR I=1 TO N
4630 P0=1.0
4640 F=1.0
4650 FOR K=1 TO NS(I)-1
4680 F=F*K
4690 P0=P0+(NS(I)*RHO(I))**K/F
4700 NEXT K
4710 F=F*NS(I)
4720 P0=P0+(NS(I)*RHO(I))**NS(I)/(F*(1-RHO(I)))
4730 X=AVST(I)*RHO(I)*(NS(I)*RHO(I))**(NS(I)-1)/(P0*F*(1-RHO(I))**2)
4740 Z=(1-RHO(I))*(NS(I)-1)*(SQR(4+5*NS(I))-2)/(16*RHO(I)*NS(I))
4750 Y=CS(I)*(1-CA(I,1))*(0.5-0.325*(2-RHO(I))*(1-RHO(I)))*(1-4*Z)
4760 Q=0.5*(1-CS(I))
4770 AVW(I)=X*(CS(I)*CA(I,1)+Q*CA(I,1)*(1+Z)+Y)
4780 IF RHO(I)>=1.0 AVW=10000.0
4790 NEXT I
4800 PRINT " THE AVERAGE WAIT AT EACH CENTRE IS:-"
4801 FOR I=1 TO N
4802 PRINT USING A$,AVW(I)
4803 NEXT I
4804 INPUT "TYPE 1 TO CONTINUE",I
```

```
4805 REM ESTIMATION OF THE STANDARD DEVIATIONS OF THE WAITING TIMES
4813 FOR I=1 TO N
4814 X=RHO(I)**SQR(3.0*NS(I))
4815 Y=RHO(I)**SQR(7.0*NS(I))
4816 Z=0.3850*X*CA(I,1)*CS(I)+0.1877*X*CA(I,1)*CA(I,1)+0.2659*Y*CS(I)*CS(I)
4817 VAR(I)=(NS(I)**0.25)*Z/((1-RHO(I))*A(I.1))**2
4818 SDW(I)=SQR(VAR(I))
4819 NEXT I
4820 PRINT " THE S.D. OF WAITING TIME AT EACH CENTRE IS:-"
4821 FOR I=1 TO N
4822 PRINT USING A$, SDW(I)
4823 NEXT I
4824 INPUT "TYPE 1 TO CONTINUE",I

4825 REM ESTIMATION OF THE AVERAGE THROUGHPUT TIME
5000 TW=0.0:TS=0.0:TC=0.0
5010 FOR I=1 TO N
5020 TW=TW+A(I,1)*AVW(I)
5030 TS=TS+A(I,1)*AVST(I)
5040 TC=TC+CUST(I,1)
5041 ATT(I)=AVW(I)+AVST(I)
5050 NEXT I
5060 XX=TW/TC
5070 THRU=(TW+TS)/TC
5080 PRINT USING H$,XX
5081 PRINT USING I$,THRU

5082 REM ESTIMATION OF THE STANDARD DEVIATION OF THE THROUGHPUT TIME
5083 FOR I=1 TO N
5084 EN(I)=A(I,1)/TC
5085 VAR(I)=VAR(I)+CS(I)*(AVST(I))**2
5093 NEXT I
5094 OVER=0
5095 SUMJ=0
5096 FOR I=1 TO N
5097 FOR J=1 TO N
5098 IF J=I THEN DO
5099 COV(I,J)=EN(J)*(2*W(J,J)-EN(J)-1)*ATT(J)
5100 ELSE
5101 COV(I,J)=EN(J)*W(J,I)+EN(I)*W(I,J)-EN(I)*EN(J)
5102 COV(I,J)=COV(I,J)*ATT(J)
5103 DOEND
5104 SUMJ=SUMJ+COV(I,J)
5105 NEXT J
5106 OVER=OVER+EN(I)*VAR(I)+SUMJ*ATT(I)
5107 SUMJ=0
5108 NEXT I
5109 TSD=SQR(OVER)
5113 PRINT USING J$," S.D. OF THROUGHPUT TIME=";TSD
```

```
5114 REM END OF RUN
5200 PRINT " IF YOU WISH TO RUN THE PROGRAM AGAIN MODIFYING ONLY "
5210 INPUT " PART OF THE DATA ENTER Y",C$
5220 IF C$=B$ GOTO 5240
5230 STOP
5240 PRINT " IF THE FIRST DATA TO BE CHANGED IS THE FLOW PROPORTIONS TYPE 1"
5250 PRINT " IF THE FIRST DATA TO BE CHANGED IS THE SERVICE TIMES TYPE 2"
5260 PRINT " IF THE FIRST DATA TO BE CHANGED IS THE S.D.'S TYPE 3"
5270 PRINT " IF THE FIRST DATA TO BE CHANGED IS THE NUMBER OF SERVERS TYPE 4"
5280 PRINT " IF THE FIRST DATA TO BE CHANGED IS THE RATES OF ARRIVALS TYPE 5 "
5290 INPUT "TYPE YOUR OPTION:-",I
5300 IF I=1 GOTO 730
5310 IF I=2 GOTO 1200
5320 IF I=3 GOTO 1400
5330 IF I=4 GOTO 1800
5340 IF I=5 GOTO 600
5350 STOP
5360 END
```

A3 SIMULATION AND APPROXIMATION RESULTS

Tables A3.1-A3.10 give the simulation and approximation results for the networks considered in Chapter 8. The value of the coefficient of variation of the arrivals (C_a^2), the correlation of the arrival stream, and the mean, and standard deviation of the waiting time at the M^{th} centre is given for each of the M-centre networks. Coefficients of variation of the service times (C_s^2) of 0.2, 0.5 and 1.0, and utilisations (u) of the service centres of 0.4, 0.6 and 0.8 are included.

The approximation algorithm is based on the assumption that all the transition processes in a queueing network are renewal, hence the approximations for the correlation of the arrival streams are shown as zero. Significant correlations (>1.96/ N, where N is the number of observations [6]) arose in many of the simulated arrival streams, these are indicated by asterisks.

Table A3.1 Networks of Four Single-Server Centres with u=0.4

c_s^2		Flow Shop Simulation	Flow Shop Approx.	Symmetric Shop Simulation	Symmetric Shop Approx.
0.2	c_a^2	0.851	0.681	1.274	0.969
	Correlation	0.045*	0.000	0.086*	0.000
	Mean Waiting Time	0.075	0.112	0.157	0.155
	S.D. Waiting Time	0.171	0.286	0.288	0.381
	Mean Throughput Time	2.024±0.105	2.134	2.220±0.111	2.221
	S.D. Throughput Time	0.581	0.760	2.085	2.100
0.5	c_a^2	0.866	0.812	1.279	0.981
	Correlation	0.008	0.000	0.075*	0.000
	Mean Waiting Time	0.146	0.167	0.210	0.197
	S.D. Waiting Time	0.312	0.419	0.400	0.476
	Mean Throughput Time	2.281±0.114	2.325	2.427±0.117	2.386
	S.D. Throughput Time	0.927	1.057	2.349	2.344
1.0	c_a^2	0.983	1.000	1.277	1.000
	Correlation	0.012	0.000	0.055*	0.000
	Mean Waiting Time	0.248	0.267	0.257	0.267
	S.D. Waiting Time	0.488	0.625	0.512	0.625
	Mean Throughput Time	2.653±0.128	2.677	2.647±0.132	2.667
	S.D. Throughput Time	1.318	1.484	2.743	2.745

Table A3.2 Networks of Four Single-Server Centres with u=0.6

c_s^2		Flow Shop Simulation	Flow Shop Approx.	Symmetric Shop Simulation	Symmetric Shop Approx.
0.2	c_a^2	0.594	0.453	1.053	0.933
	Correlation	0.066*	0.000	0.064*	0.000
	Mean Waiting Time	0.231	0.276	0.541	0.508
	S.D. Waiting Time	0.420	0.454	0.771	0.790
	Mean Throughput Time	3.776±0.181	3.908	4.502±0.225	4.431
	S.D. Throughput Time	1.313	1.349	4.450	4.184
0.5	c_a^2	0.728	0.671	1.043	0.958
	Correlation	0.029*	0.000	0.036*	0.000
	Mean Waiting Time	0.490	0.500	0.728	0.653
	S.D. Waiting Time	0.799	0.811	1.103	1.014
	Mean Throughput Time	4.641±0.232	4.662	5.215±0.250	5.012
	S.d. Throughput Time	2.062	1.995	5.408	4.865
1.0	c_a^2	0.978	1.000	1.086	1.000
	Correlation	-0.003	0.000	0.026*	0.000
	Mean Waiting Time	0.865	0.900	0.837	0.900
	S.D. Waiting Time	1.303	1.381	1.270	1.381
	Mean Throughput Time	5.990±0.290	6.000	5.848±0.273	6.000
	S.D. Throughput Time	3.004	3.012	6.103	6.006

Table A3.3 Networks of Four Single-Server Centres with u=0.8

c_s^2		Flow Shop Simulation	Flow Shop Approx.	Symmetric Shop Simulation	Symmetric Shop Approx.
0.2	c_a^2	0.383	0.259	0.918	0.887
	Correlation	-0.056*	0.000	0.011*	0.000
	Mean Waiting Time	0.800	0.697	1.913	1.733
	S.d Waiting Time	1.134	0.837	2.348	1.953
	Mean Throughput Time	7.847±0.383	7.528	10.788±0.555	10.131
	S.D. Throughput Time	3.845	2.812	11.689	9.630
0.5	c_a^2	0.598	0.545	0.950	0.929
	Correlation	-0.005	0.000	0.008*	0.000
	Mean Waiting Time	1.848	1.615	2.406	2.278
	S.D. Waiting Time	2.473	1.910	2.931	2.591
	Mean Throughput Time	11.149±0.495	10.634	12.766±0.637	12.312
	S.D. Throughput Time	5.434	4.514	13.844	11.909
1.0	c_a^2	0.995	1.000	1.017	1.000
	Correlation	-0.002	0.000	0.011*	0.000
	Mean Waiting Time	3.118	3.200	3.138	3.200
	S.D. Waiting Time	3.118	3.662	3.843	3.662
	Mean Throughput Time	15.675±0.776	16.000	16.015±0.774	16.000
	S.D. Throughput Time	8.031	7.497	17.509	15.755

Table A3.4 Networks of Four Two-Server Centres with u=0.4

c_s^2		Flow Shop Simulation	Flow Shop Approx.	Symmetric Shop Simulation	Symmetric Shop Approx.
0.2	c_a^2	0.920	0.844	1.140	0.984
	Correlation	-0.016*	0.000	0.088*	0.000
	Mean Waiting Time	0.055	0.082	0.095	0.095
	S.D. Waiting Time	0.163	0.266	0.248	0.302
	Mean Throughput Time	3.488±0.170	3.553	3.590±0.174	3.580
	S.D. Throughput Time	0.852	0.914	3.220	3.239
0.5	c_a^2	0.926	0.908	1.169	0.990
	Correlation	-0.005	0.000	0.087*	0.000
	Mean Waiting Time	0.093	0.108	0.124	0.116
	S.D. Waiting Time	0.262	0.350	0.330	0.372
	Mean Throughput Time	3.588±0.175	3.647	3.708±0.175	3.665
	S.D. Throughput Time	1.260	1.342	3.511	3.451
1.0	c_a^2	0.981	1.000	1.179	1.000
	Correlation	-0.010	0.000	0.078*	0.000
	Mean Waiting time	0.149	0.152	0.159	0.152
	S.D. Waiting Time	0.422	0.479	0.442	0.479
	Mean Throughput Time	3.791±0.199	3.810	3.818±0.184	3.810
	S.D. Throughput Time	1.803	1.864	3.827	3.790

Table A3.5 Network of Four Two-Server Centres with u=0.6

c_s^2		Flow Shop Simulation	Flow Shop Approx.	Symmetric Shop Simulation	Symmetric Shop Approx.
0.2	c_a	0.758	0.730	0.998	0.967
	Correlation	-0.082*	0.000	-0.041*	0.000
	Mean Waiting Time	0.214	0.313	0.395	0.402
	S.D. Waiting Time	0.441	0.577	0.689	0.737
	Mean Throughput Time	5.929±0.278	6.206	6.430±0.303	6.409
	S.D. Throughput Time	1.625	1.694	6.037	5.842
0.5	c_a^2	0.861	0.838	1.006	0.979
	Correlation	-0.030*	0.000	0.034*	0.000
	Mean Waiting Time	0.382	0.444	0.502	0.504
	S.D. Waiting Time	0.745	0.835	0.955	0.926
	Mean Throughput Time	6.508±0.279	6.678	6.806±0.295	6.815
	S.D. Throughput Time	2.431	2.436	6.571	6.414
1.0	c_a^2	0.992	1.000	1.062	1.000
	Correlation	-0.000	0.000	0.035*	0.000
	Mean Waiting Time	0.683	0.675	0.722	0.675
	S.D. Waiting Time	1.247	1.229	1.397	1.229
	Mean Throughput Time	7.502±0.347	7.500	7.582±0.337	7.500
	S.D. Throughput Time	3.511	3.435	7.684	7.348

Table A3.6 Networks of Four Two-Server Centres with u=0.8

c_s^2		Flow Shop Simulation	Flow Shop Approx.	Symmetric Shop Simulation	Symmetric Shop Approx.
0.2	c_a^2	0.588	0.631	0.931	0.943
	Correlation	-0.174*	0.000	-0.003	0.000
	Mean Waiting Time	0.807	1.170	1.827	1.637
	S.D. Waiting Time	1.210	1.503	2.431	2.064
	Mean Throughput Time	10.625±0.400	11.773	13.540±0.862	12.950
	S.D. Throughput Time	3.746	3.748	13.803	12.036
0.5	c_a^2	0.726	0.773	0.957	0.965
	Correlation	-0.063*	0.000	0.000	0.000
	Mean Waiting Time	1.603	1.786	2.185	2.087
	S.D. Waiting Time	2.358	2.314	2.873	2.657
	Mean Throughput Time	13.609±0.549	13.986	15.220±0.862	14.750
	S.D. Throughput Time	5.789	5.388	15.712	14.019
1.0	c_a^2	1.004	1.000	1.005	1.000
	Correlation	-0.003	0.000	0.006	0.000
	Mean Waiting Time	2.898	2.844	2.832	2.844
	S.D. Waiting Time	3.717	3.645	3.870	3.645
	Mean Throughput Time	17.482±0.666	17.778	18.309±1.305	17.779
	S.D. Throughput Time	7.723	7.962	19.669	17.333

Table A3.7 Networks of Four Five-Server Centres with u=0.4

c_s^2		Flow Shop Simulation	Flow Shop Approx.	Symmetric Shop Simulation	Symmetric Shop Approx.
0.2	c_a^2	0.985	0.942	1.013	0.994
	Correlation	-0.009*	0.000	0.020*	0.000
	Mean Waiting Time	0.018	0.026	0.028	0.028
	S.D Waiting Time	0.112	0.170	0.152	0.177
	Mean Throughput Time	8.073	8.108	8.075	8.111
	S.D. Throughput Time	1.806	1.821	7.205	7.257
0.5	c_a^2	0.982	0.966	1.026	0.996
	Correlation	0.005	0.000	0.035*	0.000
	Mean Waiting Time	0.029	0.031	0.033	0.032
	S.D. Waiting Time	0.164	0.211	0.186	0.215
	Mean Throughput Time	8.123	8.127	8.125	8.127
	S.D. Throughput Time	2.857	2.860	7.606	7.599
1.0	c_a^2	0.997	1.000	1.063	1.000
	Correlation	-0.001	0.000	0.043*	0.000
	Mean Waiting Time	0.042	0.040	0.043	0.046
	S.D. Waiting Time	0.239	0.271	0.238	0.271
	Mean Throughput Time	8.157	8.159	8.183	8.159
	S.D. Throughput Time	4.006	4.036	8.170	8.138

Table A3.8 Networks of Four Five-Server Centres with u=0.6

C_s^2		Flow Shop Simulation	Flow Shop Approx.	Symmetric Shop Simulation	Symmetric Shop Approx.
0.2	C_a	0.924	0.897	0.986	0.987
	Correlation	-0.032*	0.000	0.000	0.000
	Mean Waiting Time	0.134	0.206	0.233	0.226
	S.D. Waiting Time	0.392	0.538	0.611	0.583
	Mean Throughput Time	12.647	12.858	12.841	12.903
	S.D. Throughput Time	2.866	2.906	11.524	11.551
0.5	C_a	0.959	0.938	0.997	0.992
	Correlation	-0.018*	0.000	0.000	0.000
	Mean Waiting Time	0.229	0.261	0.274	0.274
	S.D. Waiting Time	0.625	0.691	0.722	0.718
	Mean Throughput Time	12.962	13.064	13.081	13.095
	S.D. Throughput Time	4.478	4.468	12.252	12.193
1.0	C_a	1.003	1.000	1.013	1.000
	Correlation	-0.002	0.000	0.015*	0.000
	Mean Waiting Time	0.340	0.354	0.349	0.354
	S.D. Waiting Time	0.927	0.928	0.979	0.928
	Mean Throughput Time	13.334	13.417	13.346	13.417
	S.D. Throughput Time	6.297	6.281	13.196	13.208

Table A3.9 Networks of Four Five-Server Centres with u=0.8

c_s^2		Flow Shop Simulation	Flow Shop Approx.	Symmetric Shop Simulation	Symmetric Shop Approx.
0.2	c_a^2	0.813	0.856	0.970	0.977
	Correlation	-0.097*	0.000	-0.013*	0.000
	Mean Waiting Time	0.685	1.193	1.462	1.340
	S.D. Waiting Time	1.160	1.819	2.375	2.029
	Mean Throughput Time	19.598	20.986	21.741	21.360
	S.D. Throughput Time	5.055	5.217	20.422	19.274
0.5	c_a^2	0.874	0.912	0.979	0.986
	Correlation	-0.040*	0.000	-0.011*	0.000
	Mean Waiting Time	1.334	1.571	1.608	1.667
	S.D. Waiting Time	2.303	2.426	2.555	2.555
	Mean Throughput Time	21.930	22.423	23.104	22.670
	S.D. Throughput Time	7.788	7.513	22.345	21.060
1.0	c_a^2	1.006	1.000	1.007	1.000
	Correlation	-0.004	0.000	0.001	0.000
	Mean Waiting Time	2.330	2.216	2.058	2.216
	S.D. Waiting Time	3.805	3.417	3.398	3.417
	Mean Throughput Time	25.084	24.866	24.500	24.866
	S.D. Throughput Time	10.957	10.522	24.544	23.967

TABLE A3.10 Symmetric Shops with Ten Three-Server Centres

c_s^2		u=0.4 Sim.	u=0.4 Approx.	u=0.6 Sim.	u=0.6 Approx.	u=0.8 Sim.	u=0.8 Approx.
0.2	c_a^2	1.010	0.995	1.003	0.990	0.969	0.982
	Correlation	0.030*	0.000	-0.001	0.000	-0.006	0.000
	Mean Waiting Time	0.064	0.062	0.331	0.330	1.623	1.554
	S.D. Waiting Time	0.219	0.249	0.682	0.686	2.306	2.110
	Mean Throughput Time	12.616	12.617	21.231	21.301	39.684	39.541
	S.D. Throughput Time	12.162	12.115	20.642	20.483	39.673	38.251
0.5	c_a^2	1.034	0.997	1.009	0.994	0.983	0.989
	Correlation	0.039*	0.000	0.015*	0.000	-0.006	0.000
	Mean Waiting Time	0.071	0.074	0.416	0.406	2.018	1.941
	S.D. Waiting Time	0.250	0.304	0.867	0.850	2.857	2.662
	Mean Throughput Time	12.649	12.739	21.988	22.057	44.152	43.412
	S.D. Throughput Time	12.293	12.416	21.558	21.477	44.392	42.377
1.0	c_a^2	1.046	1.000	1.021	1.000	1.010	1.000
	Correlation	0.038*	0.000	0.012*	0.000	0.009*	0.000
	Mean Waiting Time	0.096	0.094	0.500	0.532	2.473	2.589
	S.D. Waiting Time	0.342	0.385	1.071	1.108	3.549	3.577
	Mean Throughput Time	12.926	12.941	23.204	23.321	48.699	49.888
	S.D. Throughput Time	12.837	12.908	23.191	23.112	49.657	49.249

BIBLIOGRAPHY

1. A. O. Allen (1978), Probability, Statistics and Queueing Theory. Academic Press, Inc., New York.

2. A. C. Atkinson (1977), An Easily Programmable Algorithm for Generating Gamma Random Varaibles. J.R.S.S. 140, Series A Part 2, 232-234.

3. A. D. Barbour (1976), Networks of Queues and the Method of Stages. Adv. Appl. Prob. 8, 584-591.

4. F. Baskett, K. M. Chandy, R. R. Muntz and F.G. Palacios (1975), Open, Closed and Mixed Networks of Queues with Different Classes of Customers. J. ACM 22, 248-260.

5. F. J. Beutler and B. Melamed (1978), Decomposition and Customer Streams of Feedback Networks of Queues in Equilibrium. O.R. 26, p.1059.

6. G. E. P. Box and G. M. Jenkins (1976), Time Series Analysis Forecasting and Control. Revised Edition, Holden-Day, London.

7. O. J. Boxma, J. W. Cohen and N. Huffels (1979), Approximations of the Mean Waiting Time in an M/G/s System. O.R. 27, 1115-1127.

8. S. C. Bruell and G. Balbo (1980), Computational Algorithms for Closed Queueing Networks. Operating and Programming Series 7, North Holland, New York.

9. S. L. Brumelle (1971), Some Inequalities for Parallel Server Queues. O.R. 19, 402-413.

10. P. J. Burke (1956), The Output of a Queueing System. O.R. 4, p.699.

11. K. M. Chandy, U. Herzog and L. Woo (1975), Parametric Analysis of Queueing Networks. IBM J. Res. and Dev. 19, 36-42.

12. K. M. Chandy, U. Herzog and L. Woo (1975), Approximate Analysis of General Queueing Networks. IBM J. Res. and Dev. 19, 43-49.

13. K. M. Chandy and C. H. Sauer (1978), Approximate Methods for Analysing Queueing Network Models of Computing Systems. Computing Surveys 10, 281-317.

14. G. P. Cosmetatoes (1975), Approximate Explicit Formulae for the Average Queueing Time in the Processes M/D/r and D/M/R. INFOR 13, 328-332.

15. D. R. Cox and W. L. Smith (1961), Queues. Methuen and Co. Ltd., London.

16. D. R. Cox (1962), Renewal Theory. Methuen and Co. Ltd., London.

17. D.R. Cox (1964), Some Problems of Statistical Analysis connected with Congestion. Proceedings of the Symposium on Congestion Theory, University of N. Carolina, 289-310.

18. M. A. Crane and D. L. Iglehart (1974), Simulating Stable Stochastic Systems I: General Multiserver Queues. J. ACM 21, 103-113.

19. C. D. Crommelin (1934), Delay Probability Formulae. P.O. Elec. Engrs. J. 26, 266-274.

20. Crouch and Hogg (1978), Mallaig Harbour Authority Development: Report on Outcome of Wave Model Tests and Resulting Development Schemes. Messrs. Crouch and Hogg, Consulting Engineers, Glasgow.

21. D. J. Daley (1976), Queueing Output Processes. Adv. Appl. Prob. 8, 395-415.

22. R. L. Disney, R. L. Farell and P. R. De Morais (1973), A Characterisation of M/G/1 Queues with Renewal Departure Processes. Mgmt. Sci. 19, p.1222.

23. J. R. Fraker (1971), Approximate Techniques for the Analysis of Tandem Queueing Systems. Ph.D. Thesis, Clemenson University.

24. E. Gelenbe (1975), On Approximate Computer System Models. J. ACM 22, 261-269.

25. A. S. Gibbs. (1973), Mosque Users Manual. Motor Gear and Engineering Co. Ltd., Chadwell Heath, Essex.

26. W. J. Gordon and G. F. Newell (1967), Closed Queueing Systems with Exponential Queues. O.R. 15, 254-265.

27. F. S. Hillier and O. S. Yu (1981), Queueing Tables and Graphs. North Holland, New York.

28. P. Hokstad (1978), Approximations for the M/G/m Queue. O.R. 26, 510-523.

29. R. R. P. Jackson (1956), Random Queueing Processes with Phase Type Service. J.R.S.S. Series B 18, p.131.

30. J. R. Jackson (1957), Networks of Waiting Lines. O.R. 5, 518-521.

31. J. R. Jackson (1963), Job Shop-Like Queueing Systems. Mgmt. Sci. 10, 131-142.

32. J.H. Jenkins (1966), On the Correlation Structure of the Departure Process of the $M/E_\lambda/1$ Queue. J.R.S.S. Series B 28, 336-344.

33. F. P. Kelly (1975), Networks of Queues with Customers of Different Classes. J. Appl. Prob. 12, 542-554.

34. F. P. Kelly (1976), Networks of Queues. Adv. Appl. Prob. 8, 416-432.

35. D. G. Kendall (1953), Stochastic Processes Occurring in the Theory of Queues and their Analysis by the Method of the Imbedded Markov Chain. Ann. Math. Stat. 24, 338-354.

36. J. F. C. Kingman (1962), Some inequalities for the GI/G/1 Queue. Biometrica 49, 315-324.

37. J. F. C. Kingman (1964), The Heavy Traffic Approximation in the Theory of Queues. Proceedings of the Symposium on Congestion Theory, University of N. Carolina.

38. J. F. C. Kingman (1970), Inequalities in the Theory of Queues. J.R.S.S. Series B 32, 102-110.

39. L. Kleinrock (1976), Queueing Systems. Volumes 1 & 2, Wiley Interscience, London.

40. H. Kobayashi (1974), Application of the Diffusion Approximation to Queueing Networks: Part 1 Equilibrium Queue Distributions. J. ACM 21, 316-328.

41. J. Kollerstrom (1974), Heavy Traffic Theory for Queues with Several Servers 1. J. Appl. Prob. 11, 544-552.

42. W. Kramer and M. L. Lagenbach-Belz (1976), Approximate Formulae for the Delay in the Queueing System GI/G/1. 8th International Teletraffic Congress, Melbourne, 235-1 to 235-8.

43. P. Kuhn (1976), Analysis of Complex Queueing Networks by Decomposition. 8th International Teletraffic Congress, Melbourne, 236-1 to 236-8.

44. S. S. Lavenberg and C. H. Sauer (1977), Sequential Stopping Rules for the Regenerative Method of Simulation. IBM J. Res. and Dev. 21, 545-558.

45. A. J. Lemoine (1977), Networks of Queues - A Survey of Equilibrium Analysis. Mgmt. Sci. 24, 464-481.

46. D. V. Lindley (1952), The Theory of Queues with a Single Server. Proceedings of the Cambridge Philosophical Society 48, 277-289

47. J. D. C. Little (1961), A Proof of the Queueing Formula $L=\lambda W$. O.R. 9, 383-387.

48. K. T. Marshall (1968), Some Inequalities in Queueing. O.R. 16, 651-658.

49. J. Martin (1972), Systems Analysis for Data Transmission. Prentice-Hall, Englewood Cliffs, New Jersey.

50. J. O. Mayhugh and R. E. McCormick (1968), Steady State Solution of the Queue $M/E_k/r$. Mgmt. Sci. 14, 692-712.

51. N. M. Mirasol (1963) The Output of an $M/G/\infty$ Queueing System is Poisson. O.R. 11, p.282.

52. M. Mori (1975), Some Bounds for Queues. J.O.R. Soc. of Japan 18, 152-181.

53. R. R. Muntz (1978), Queueing Networks: A Critique of the State of the Art and Directions for the Future. Computing Surveys 10, 353-359.

54. G. F. Newell (1971), Applications of Queueing Theory. Chapman and Hall, London.

55. S. A. Nozaki and M. R. Sheldon (1978), Approximations in Finite Capacity Multiserver Queues with Poisson Arrivals. J. Appl. Prob. 15, 826-834.

56. E. Page (1972), Queueing Theory in O.R. Butterworths, London.

57. E. Page (1982), Waiting Times in General Queues. J.O.R. Soc. 33, 453-473.

58. C. Poyntz and R. R. P. Jackson (1976), The Steady-State Solution for the Queueing Process $E_k/E_m/r$. O.R.Q. 24, 685-694.

59. N. U. Prabhu (1965), Queues and Inventories. Wiley, London.

60. E. Reich (1957), Waiting Times when Queues are in Tandem. Ann. Math. Stats. 28, 768-773.

61. M. Reiser and H. Kobayashi (1974), Accuracy of the Diffusion Approximation for some Queueing Systems. IBM J. Res. and Dev. 18, 110-123.

62. M. Reiser and H. Kobayashi (1975), Queueing Networks with Multiple Closed Chains: Theory and Computational Algorithms. IBM J. Res. and Dev. 19, 283-294.

63. M. Rosenshine and M. J. Chandra (1975), Approximate Solutions for some Two-Stage Tandem Queues, Part I: Individual Arrivals to the Second Sage. O.R. 23, 1155-1166.

64. H. Sakasegawa (1977), An Approximation Formula $L \simeq \alpha\rho^{\beta}/(1-\rho)$. Ann.Inst. Stat. Maths. 29, 67-75.

65. H. Sakasegawa (1978), Numerical Tables of the Queueing SystemsI: $E_k/E_2/s$. Computer Science Monographs, Inst. Stat. Maths.

66. C. H. Sauer and K. M. Chandy (1975) Approximate Analysis of Central Server Models. IBM J. Res. and Dev. 19, 301-313.

67. K. C. Sevcik, A. I. Levy, S. K. Tripathi and J. L. Zahorjan (1977), Improving Approximations of Aggregated Queueing Network Subsystems. Computer Performance Modelling, Measurement and Evaluation.

68. J. G. Shanthikumar (1979), Approximate Queueing Models of Dynamic Job Shops. Ph.D. Thesis, Division of Eng. and App. Physics, Harvard University, Cambridge, Mass.

69. D. G. Shimshak and G. P. Sphicas (1982), Waiting Time in a Two-Station Series Queueing System: The Effect of Dependent Interarrival Times. J.O.R. Soc. 33, 745-750.

70. A. W. Shum (1976), Queueing Models of Computer Systems with General Service Time Distributions. Ph.D. Thesis, Division of Eng. and Appl. Physics, Harvard University, Cambridge, Mass.

71. L. Takacs (1964), A Single-Server Queue with Feedback. B.S.T.J. 42, 505-519.

72. H. C. Tijms, M. H. Van Hoorn and A. Federgruen (1981), Approximations for the Steady-State Probabilities in the M/G/c Queue. Adv. Appl. Prob 13, 186-206.

73. J. Walrand and P. Varaiya (1981), Flows in Queueing Networks - A Martingale Approach. Maths. of O.R. 6, No. 3.

74. White Fish Authority (1979), An Economic Study of Mallaig Harbour Development Schemes. Report No. M & D 405, White Fish Authority, Industrial Development Unit, Hull.

Notation

Equation (1) on page 15 and equation (3) on page 37 should both read

$$C_{a_i}^2 = \sum_{j=1}^{M} \frac{\lambda_j p_{ji}}{\lambda_i} \left(p_{ji} C_{d_j}^2 + (1-p_{ji})\right) + \frac{\gamma_i}{\lambda_i} C_{e_i}^2$$

A better notation for the decomposition step on page 14 may be

$$C_{a_i}^2 = p_{ji} C_{d_j}^2 + (1-p_{ji})$$

Where $C_{a_i}^2$ is the squared coefficient of variation of an arrival stream created by choosing jobs with probability $p_{ji} > 0$ from an original stream with squared coefficient of variation $C_{d_j}^2$.

Similarly, the composition step on page 15 can be expressed as

$$C_{a_i}^2 = \sum_{j=1}^{M} \frac{\lambda_j p_{ji}}{\lambda_i} C_{d_j}^2$$

Where $C_{a_i}^2$ is the squared coefficient of variation of the composition of M renewal processes, $C_{d_j}^2$ is the squared coefficient of variation of the j^{th} component stream, $\lambda_j p_{ji}$ is the arrival rate of stream j, and $\lambda_i = \sum_{j=1}^{M} \lambda_j p_{ji}$.

The equation on page 26 should read

$$I(\alpha,\gamma) = \left(r(N(\alpha,\gamma)) - \delta(N(\alpha,\gamma),\alpha),\; r(N(\alpha,\gamma)) + \delta(N(\alpha,\gamma),\alpha)\right)$$

AD1.2 Further Comparisons with Simulated Networks

The tables in Chapter 8 compare approximations given by the algorithm for the mean and standard deviations of throughput times with those obtained by the simulation of a number of queueing networks.

The networks are either flow or symmetric shops and, in order to simplify notation, only networks with identical service centres are considered.

A further set of tables are presented here (Tables AD1.1 - AD1.7) to illustrate the performance of the algorithm under less uniform circumstances. All the networks consist of four service centres and are described by their transition probability matrices (as defined in Chapter 8), squared coefficients of variation of service time distributions, and the utilisations and numbers of servers at the service centres.

Good approximations are given for the mean throughput times in the networks. The largest percentage error (7.1%) occurs for the network described in Table AD1.2. As was observed for the networks considered in Chapter 8, the approximations to the standard deviations of throughput times are not as accurate, a 60% error can be calculated for the network in Table AD1.1. Inaccuracies in the approximation of the standard deviations of throughput times arise in networks when the estimates of the standard deviations of the waiting times at individual service centres are poor. This appears to be the case at service centres with low coefficients of variation of arrival or service time distributions and large numbers of servers.

Correlation of the arrival processes in the networks seems to affect the accuracy of the approximations in the same way as observed in Chapter 8. In general, significantly negatively correlated arrival streams coincide with overestimations of the perameters of a service centre, and large positive correlations correspond to underestimations. The overall effect of the correlations on estimated mean throughput times is not obvious because of the more diverse nature of the service centres of the networks.

TABLE AD1.1

Transition Probability Matrix

$$\begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0.5 & 0.5 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0.5 & 0.5 & 0 & 0 & 0 \end{pmatrix}$$

Service Centre	1		2		3		4	
C_s^2	0.2		0.2		0.2		0.2	
Utilisation	0.45		0.90		0.45		0.90	
Number of Servers	5		5		5		5	
	Sim	Approx	Sim	Approx	Sim	Approx	Sim	Approx
C_a^2	0.999	1.000	0.990	0.982	0.833	0.926	0.903	0.924
Correlation	0.005	0.000	-0.003	0.000	-0.056	0.000	-0.052	0.000
Average Wait	0.099	0.101	4.192	4.104	0.039	0.094	2.496	3.896
Standard Deviation	0.440	0.489	5.416	5.134	0.237	0.459	3.023	4.884
Mean Throughput Time							20.253	21.598
Standard Deviation							8.159	13.046

TABLE AD1.2

Transition Probability Matrix

$$\begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0.5 & 0.5 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0.5 & 0.5 & 0 & 0 & 0 \end{pmatrix}$$

Service Centre	1		2		3		4	
C_s^2	0.5		0.5		0.5		0.5	
Utilisation	0.45		0.90		0.45		0.90	
Number of Servers	5		5		5		5	
	Sim	Approx	Sim	Approx	Sim	Approx	Sim	Approx
C_a^2	1.000	1.000	0.997	0.989	0.884	0.954	0.924	0.953
Correlation	-0.002	0.000	-0.003	0.000	-0.021	0.000	-0.029	0.000
Average Wait	0.107	0.119	4.653	5.137	0.071	0.114	3.989	5.007
Standard Deviation	0.491	0.594	5.765	6.518	0.368	0.574	5.434	6.368
Mean Throughput Time							22.176	23.760
Standard Deviation							10.426	15.312

TABLE AD1.3

Transition Probability Matrix

$$\begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0.5 & 0.5 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0.5 & 0.5 & 0 & 0 & 0 \end{pmatrix}$$

Service Centre	1		2		3		4	
C_s^2	1.0		1.0		1.0		1.0	
Utilisation	0.45		0.90		0.45		0.90	
Number of Servers	5		5		5		5	
	Sim	Approx	Sim	Approx	Sim	Approx	Sim	Approx
C_a^2	1.000	1.000	1.003	1.000	1.005	1.000	1.001	1.000
Correlation	0.000	0.000	0.000	0.000	-0.002	0.000	0.000	0.000
Average Wait	0.145	0.149	8.185	6.862	0.154	0.149	6.574	6.862
Standard Deviation	0.661	0.749	10.989	8.847	0.705	0.749	8.729	8.847

	Sim	Approx
Mean Throughput Time	28.402	27.374
Standard Deviation	17.428	18.992

TABLE ADI.4

Transition Probability Matrix

$$\begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Service Centre	1		2		3		4	
C_s^2	0.25		0.50		0.75		1.00	
Utilisation	0.4		0.8		0.4		0.8	
Number of Servers	1		2		3		4	
	Sim	Approx	Sim	Approx	Sim	Approx	Sim	Approx
C_a^2	1.016	1.000	0.894	0.864	0.784	0.791	0.872	0.865
Correlation	0.004	0.000	0.036	0.000	-0.043	0.000	-0.029	0.000
Average Wait	0.165	0.167	2.332	1.929	0.046	0.068	1.884	2.182
Standard Deviation	0.297	0.407	3.095	2.477	0.205	0.298	2.995	3.259
Mean Throughput Time							10.840	10.745
Standard Deviation							5.656	5.445

TABLE AD1.5

Transition Probability Matrix	(0 0.25 0.25 0.25 0.25) (0.25 0 0.25 0.25 0.25) (0.25 0.25 0 0.25 0.25) (0.25 0.25 0.25 0 0.25) (0.25 0.25 0.25 0.25 0)							
Service Centre	1		2		3		4	
C_s^2	0.25		0.50		0.75		1.00	
Utilisation	0.40		0.80		0.40		0.80	
Number of Servers	1		2		3		4	
	Sim	Approx	Sim	Approx	Sim	Approx	Sim	Approx
C_a^2	1.056	0.987	1.023	0.989	1.021	0.979	0.992	0.978
Correlation	0.035	0.000	0.027	0.000	0.032	0.000	0.020	0.000
Average Wait	0.165	0.165	2.339	2.126	0.077	0.082	2.163	2.352
Standard Deviation	0.304	0.402	3.134	2.701	0.281	0.341	3.378	3.458
Mean Throughput Time							11.073	11.125
Standard Deviation							12.430	11.594

TABLE AD1.6

Transition Probability Matrix

$$\begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Service Centre	1		2		3		4	
C_s^2	0.25		0.50		0.75		1.00	
Utilisation	0.40		0.80		0.40		0.80	
Number of Servers	4		3		2		1	
	Sim	Approx	Sim	Approx	Sim	Approx	Sim	Approx
C_a^2	0.999	1.000	0.986	0.966	0.833	0.873	0.869	0.901
Correlation	-0.000	0.000	-0.007	0.000	-0.050	0.000	-0.039	0.000
Average Wait	0.041	0.042	1.903	1.908	0.089	0.120	3.137	3.017
Standard Deviation	0.183	0.217	2.678	2.621	0.283	0.393	4.377	3.488
Mean Throughput Time							10.763	10.687
Standard Deviation							5.777	4.889

TABLE AD1.7

Transition Probability Matrix	(0 0.25 0.25 0.25 0.25) (0.25 0 0.25 0.25 0.25) (0.25 0.25 0 0.25 0.25) (0.25 0.25 0.25 0 0.25) (0.25 0.25 0.25 0.25 0)							
Service Centre	1		2		3		4	
c_s^2	0.25		0.50		0.75		1.00	
Utilisation	0.40		0.80		0.40		0.80	
Number of Servers	4		3		2		1	
	Sim	Approx	Sim	Approx	Sim	Approx	Sim	Approx
c_a^2	0.998	0.990	1.031	0.996	1.029	0.990	0.995	0.988
Correlation	0.025	0.000	0.017	0.000	0.019	0.000	0.020	0.000
Average Wait	0.042	0.042	2.199	1.951	0.135	0.134	3.386	3.178
Standard Deviation	0.192	0.215	3.037	2.674	0.372	0.425	4.152	4.152
Mean Throughput Time							11.418	10.905
Standard Deviation							12.634	10.937

ADDENDUM

AD1 Further Discussion on the Accuracy of the Algorithm

AD1.1 The Application to Networks of G1/G/n Queries

In this thesis it is asserted that an algorithm is developed to approximate queueing times in networks of queues with generalised arrival and service time distributions. The algorithm estimates the throughput time in a network by a composition of the properties of the individual queues of the network. The composition process requires that the arrival and departure streams of the individual queues form renewal processes. The effect of correlated arrival and departure processes on the accuracy of the algorithm is discussed in Chapter 8 and Addendum 4.

The mean and standard deviation of the waiting time, and the variance of the departures of the individual queues are estimated by approximation formulae. Queues with gamma distributed arrival and service time distributions and with coefficients of variation less than or equal to one are used both to develop and to assess the accuracy of the approximations. The algorithm has not been tested for networks of queues with arrival and service time distributions that are not readily approximated by gamma distributions with coefficients of variation less than one. The results of the algorithm should therefore be viewed with caution if such networks are to be modelled. However, the algorithm proved to be a useful planning tool in the job shop application described in Chapter 10 although, as Table 10.2 shows, the coefficients of variation of the arrival and service time distributions at most of the work centres are greater than one.

AD2 A Re-Estimation of Queueing Times in Mallaig Harbour

AD2.1 Introduction

In Chapter 9 the approximation algorithm is used to model the queues that arise as ships arrive to unload fish and take on ice in Mallaig Harbour. Approximate waiting times given by the algorithm are compared with data collected by the White Fish Authority on two days with differing traffic conditions. On both the days considered most of the ships arrived in the harbour between 2.00 p.m. and 9.00 p.m. The first day observed, Tuesday 19th June 1979, was a 'quiet' day. Only twenty three ships arrived, and eighteen of the arrivals were between 7.00 p.m. and 9.00 p.m. The algorithm is used to approximate the queueing times in this two hour period and overestimates the average waiting times for unloading berths and ice berths (Table 9.2). It is suggested that this is because ships arrive over too short a period for the queueing system to build up to conditions that can be regarded as steady-state.

The same exercise is undertaken for the 'busy' day, Thursday 21st June 1979. The algorithm is used to estimate the average waiting times in the harbour between 2.00 p.m. and 8.00 p.m. when forty of the day's forty three ships arrive. Table 9.2 shows that the queueing times are again overestimated because of the time taken before congestion begins to build up in the harbour.

AD2.2 A Further Application of the Algoritum

Another application of the algorithm to the situation in the harbour on Thursday 21st June is considered here. Only the peak two hours from 4.00 p.m. to 6.00 p.m. are modelled. In this period conditions were fairly steady. Large queues had already begun to develop in the harbour and the arrival rate of ships had not started to decline.

Nine working berths were available for use between 4.00 p.m. and 6.00 p.m. The means of the arrival and service times in the period implied a 90% utilisation of the berths. An 104% utilisation can be estimated for the ice berths. This indicates that, if arrivals had continued at the same rate and the average loading time remained the same for a longer period of time, very long queues would have developed. In fact the data collected by the White Fish Authority shows that queueing times for the ice berths gradually increased between 4.00 p.m. and 6.00 p.m.

The White Fish Authority observed that as congestion grew in the harbour the time taken for ships to load ice increased. In order to get a realistic estimate of the average waiting times at the ice berth, an average service time of 11.6 minutes was input into the algorithm. This is a weighted average of the actual mean service times of 9.54 minutes on the Tuesday and 12.6 minutes on the Thursday.

Table AD2.2 compares observed and approximated queueing times for the two hour period.

Table AD2.2 Queueing Times Between 4.00 p.m. and 6.00 p.m. on Thursday 21st June

	Average Waiting Times		Throughput Times	
	Working Berths	Ice Berth	Mean	Standard Deviation
Observed	8.40	22.0	85.9	34.4
Approx.	9.16	22.4	87.3	24.3

The approximation to the mean waiting time at the working berths is close to the observed time. As the service time at the ice berth was chosen to represent the actual queueing times there, a good approximation is to be expected. An accurate estimate of the mean throughput time is now given. Although the approximation to the standard deviation of throughput times is not as close (29% error), it is of a similar order to the value observed.

AD2.3 Conclusions

The algorithm provides good approximations to the average queueing times in Mallaig Harbour between 4.00 p.m. and 6.00 p.m. on a 'busy' day. Better results are obtained here because a short enough period was modelled to exclude most of the build up and run down of congestion occurring at the beginning and end of the afternoon.

Hence, by identifying periods when a queueing system remains relatively stable, a steady-state queueing model such as the algorithm can be useful in the estimation of queueing times.

AD3 The Use of the Algorithm to Estimate Due Dates in a Job Shop

AD3.1 Introduction

The algorithm developed in this thesis can prove a useful tool in estimating the due dates of jobs processed in a manufacturing job shop. In this section the typical assumptions that may need to be made in order to apply the algoritum are discussed. Particular reference is made to Universal Joint production at Fenner's Motor Gear and Engineering Company Ltd. - the job shop considered in Chapter 10.

AD3.2 The Application of the Algorithm to a Job Shop

In the development of the algorithm a number of assumptions are made about the type of queueing network to be modelled. Some of these assumptions can be translated into the context of a manufacturing job shop as:

1) All jobs enter the shop in their order of arrival and each job has a processing order which can be described on arrival by a series of transition probabilities. Jobs are not allowed priority - all jobs are serviced on a first come first served basis.

2) Machine centres are not interchangeable - each operation required by a job can only be performed at one machine centre. A machine centre consists of a number of machines with identical mean processing times. The processing times of all jobs are independently identically distributed.

3) The production capacity of a machine centre remains constant - the same number of machines are always available for use.

It is unlikely that such conditions will prevail in a practical job shop situation. It will usually be necessary to make some kind of modifications to produce a system that can be modelled by the algorithm. The assumptions that were made in the Fenner's application are discussed here.

Service Discipline

Most orders for Universal Joints arrive at Fenner's job shop with standard lead times. In this instance first come first served is a sensible scheduling discipline to adopt as in general, depending on the number and sequence of operations required, jobs will be despatched in the order they arrive. However, the queueing discipline assumed may not have a great effect on the date that a job is completed. Table AD3.1 shows simulated values for the means and standard deviations of throughput times in queueing networks with four single-server centres under first come first served (FCFS) and shortest processing time (SPT) (Shanthikumar (68)) disciplines. It is only at high utilisations(U = 0.8) that any appreciable differences in throughput times emerge. At utilisations of 0.4 and 0.6 both the means and standard deviations take similar values for the two queueing disciplines.

TABLE AD3.1 SIMULATED THROUGHPUT TIMES FOR TWO QUEUING DISCIPLINES
(In Units of Mean Service Time)

Flow Shop with Four Single-Service Centres

u	C_s^2	SPT Mean	SPT SD	FCFS Mean	FCFS SD
0.4	0.2	4.933	1.600	5.060	1.545
	0.5	5.369	2.240	5.703	2.318
	1.0	6.136	3.191	6.633	3.295
0.6	0.2	5.988	2.784	6.293	2.189
	0.5	6.740	3.469	7.735	3.438
	1.0	7.774	4.369	9.983	5.007
0.8	0.2	8.436	6.992	9.809	4.806
	0.5	9.835	9.016	13.786	6.793
	1.0	11.342	10.656	19.706	10.039

Symmetric Shop with Four Single-Server Centres

u	C_s^2	SPT Mean	SPT SD	FCFS Mean	FCFS SD
0.4	0.2	5.371	5.086	5.550	5.213
	0.5	5.625	5.600	6.068	5.873
	1.0	6.074	6.254	6.618	6.858
0.6	0.2	6.710	8.184	7.503	7.417
	0.5	7.107	8.115	8.692	9.013
	1.0	7.850	9.342	9.747	10.172
0.8	0.2	9.965	15.244	13.485	14.611
	0.5	10.452	16.356	15.958	17.305
	1.0	11.528	16.961	20.019	21.886

Processing Times

It may be the case in a job shop that a required operation can be performed by more than one machine centre, or that men and/or machines can transfer from one centre to another to help work off back logs. Both these situations arise in Fenner's job shop.

In order to apply the algorithm groups of machines that perform similar operations and are capable of some degree of interchangeability were amalgamated and considered to form one work centre. The average processing time of all the machines was estimated and this was assumed to be the mean service time of each of the machines at the work centre. The overall standard deviation of processing times at the work centres was also assumed for each machine, so providing some measure of the variability of the service times. The same principle applied to the amount of processing required by different jobs. The batch sizes of jobs varied from one to four thousand but all jobs were considered to require the same mean processing times with the overall standard deviation representing the possible variation. These assumptions resulted in hyperexponential service time distributions being assigned to many of the work centres (Table 10.2) in the application of the algorithm. Although not tested for hyperexponential distributions, the algorithm provides a good model of the queues arising in the job shop.

Production Capacity

The production capacity of machine centres can be affected by machine breakdowns, shift patterns and bonus schemes. The best estimate given of available production time in Fenner's job shop was between forty five and fifty five hours a week. Tables 10.5 and 10.6 show that varying production from forty five to fifty five hours per week greatly effects queueing times at the machine centres. These results indicated that the assumption of fifty hours production capacity would provide a realistic model of waiting times at most of the work centres. Lower production capacities were assumed for work centres with low utilisations. This was to take account of the practice of allowing jobs to accumulate before starting production.

AD3.3 Conclusions

The algorithm can provide a useful representation of the queueing processes arising in a manufacturing job shop. As it is unlikely that conditions in a job shop will be directly amenable to analysis by the algorithm, various modifications may need to be made. In the application of the algorithm to Fenner's job shop, the grouping together of similar work centres, and the production capacities assumed, allowed the algorithm to estimate the mean throughput time of a job to within 10% of the observed mean.

Once a reasonable model has been established it can be used to investigate the effects of changes in the job shop. The algorithm can assess how throughput times are altered by different routings of jobs through the shop, by varying the number of machines in operation at a work centre, or by increasing the available production time.

AD4 Further Discussion on the Accuracy of the Algoritum and Future Development of the Research

AD4.1 Introduction

It was noted in Chapter 8 that the arrival streams in the queueing networks considered were often correlated and that the presence of correlation appears to affect the accuracy of the algorithm.

Correlation in the arrival processes of a queueing network may be attributable to three factors:

1) The arrivals are made up of the departures from non M/M/N queues. In Chapter 5 the correlation of the departures from $E_j/E_k/n$ queues are discussed. Only M/M/N queues, and M/D/1 and M/G/1 queues with restricted waiting room, have uncorrelated departure processes. Hence, correlations will arise in any networks with arrival or service time distributions that are not exponential.

2) Further correlation may result when the routings through the queueing network cause the arrivals to a service centre to consist of decompositions and compositions of departure streams from other centres. If all the streams are renewal processes then any compositions or decompositions will also produce renewal processes (47). On the other hand, if the departures from one service centre are correlated then all the arrival processes that the departures contribute to may be correlated to some extent.

3) Correlations may arise in the flows of a network due to its routing patterns. Positively correlated arrival streams have been observed in symmetric shops with exponentially distributed service times and low utilisations. In queues with low utilisations interdeparture times are likely to be correlated to the interarrival times. The transition probabilities of a symmetric shop allow a component of the departures from a queue to revisit the same queue after visiting one or more other queues. If the intermediate queues also have low utilisations they will only serve to introduce a lag effect on the correlation with the original arrival process.

AD4.2 Correlation and the Accuracy of the Algoritum

It is pointed out in Chapter 4 that the existence of correlated arrival streams in a queueing network affects the accuracy of the approximations to the squared coefficients of variation of the arrival times. Positive correlations coincide with underestimations of the coefficient of variation, and negative correlations with overestimations. Similarly, the simulation results for the networks in Appendix 3 show that in general positive correlations correspond to underestimations of mean waiting times and negative correlations to overestimates - that this is not always the case may be due to other errors involved both in the approximation of the coefficients of variation of the arrival time distributions and in the approximation of the mean waiting times. The effect of correlation on the estimates of network throughput times may be further masked by the method of composition of the results from the individual queues. However, in the simple example of the flow shop, there does seem to be a relationship between the type of correlation in the arrival streams and under and overestimation of mean throughput times (Tables A3.3 and A3.6).

AD4.3 Conclusions

Correlation is present to some extent in the arrival processes of most queueing networks, either because the departures from individual queues are correlated or due to the pattern of routings through the networks. From the results for networks considered in this thesis it seems evident that correlated arrival streams effect both the variance of the interdeparture times and the mean waiting time of a queue. Because of the number of factors involved, it is difficult to ascertain the effect of correlation on network throughput times, though it does appear to directly affect the accuracy of the algorithm for flow shops with highly utilised service centres. In networks with more complex patterns of flows the effect of correlated arrivals is dampened by the other approximations involved and no appreciable difference is made to the accuracy of the algorithm.

AD4.4 Future Developments of the Research

Shimshak and Sphicas(69) have examined the effect of positively correlated arrivals on the mean waiting times of single-server queues with Erlang service time distributions. They found that the assumption that arrivals were independent caused the mean waiting times to be underestimated. The comparisons of simulated and approximate results for queueing networks in this thesis suggest the more general result - the assumption that positively correlated arrivals to a multiserver queue are independent causes mean waiting times to be underestimated, and assuming negatively correlated arrivals to be independent leads to overestimation of mean waiting times. As the algorithm approximates both the coefficient of variation of the arrivals to a queue and the mean waiting time in a queue, no accurate assessment of the effect of correlation can be made without taking into account the errors of these approximations.

A quantitative evaluation of the effect of correlation on the mean waiting times of queues will allow better approximations to be made both for individual queues and queueing networks.

When the complete decomposition algorithm is applied to networks with complex flow patterns, such as symmetric shops, the composition of the results of the individual queues may also be affected by correlation. The method of composition is based on independence assumptions and no investigations has been made into the influence of correlated flows. However, the accuracy of the algorithm for symmetric shops with positively correlated arrivals suggests that errors in the estimations of waiting times for the individual queues may be cancelled out to some extent by the composition process.

Although it may be desirable to have a better understanding of how correlation affects the composition process in networks with variable routings, it may not greatly improve the accuracy of the algorithm. A more fruitful area for future research may be the development of a better approximation for the standard deviation of the waiting time distribution of a GI/G/n queue. Errors of up to 60% have been observed in the algorithm's estimation of standard deviations of throughput times. Large errors arise when the approximations for the standard deviations of the waiting times at the individual service centres are poor. This is most often the case when the coefficients of variation of the arrival or service time distributions are small. The development of a series of approximation formulae applicable over narrower ranges of the coefficients of variation may result in improved accuracy.

CB