Centrum voor Wiskunde en Informatica
Centre for Mathematics and Computer Science

J.L. van den Berg

Queueing analysis of a virtual circuit in a computer communication network with window flow control

Department of Operations Research and System Theory Note OS-N8602 March

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

Queueing Analysis of a Virtual Circuit in a Computer Communication Network with Window Flow Control

J.L. van den Berg
Centre for Mathematics and Computer Science
P.O. Box 4079, 1009 AB Amsterdam, The Netherlands

This note studies a virtual circuit with finite buffer space in a computer communication network with window flow control. An approximation method is derived for the throughput in this circuit. The method can also be used to approximate other important performance measures. For a special case an exact analysis is presented.

1980 Mathematics Subject Classification: 60K25, 68M20. 69C40
Key Words & Phrases: flow control, virtual circuit, overflow, closed queueing model, throughput, end-to-end delay.
Note: This report describes research done by the author as assistant at the C W I in the period August 1985 - January 1986. It also forms the author's Master Thesis at the Mathematical Institute, University of Utrecht.

Note OS-N8602
Centre for Mathematics and Computer Science
P.O. Box 4079, 1009 AB Amsterdam, The Netherlands

TABLE OF CONTENTS page

NOTE ON NOTATIONS AND REFERENCING

Throughout this report, a bold letter denotes a random variable.
The symbol ":=" stands for the defining equality sign.
The name of an author, followed by a number between brackets, refers to the list of references.

ACKNOWLEDGMENT

The author wishes to thank Professor J.W. Cohen and Dr. O.J. Boxma for their support and encouragement during the writing of this report.

Chapter 1

Introduction

In a computer communication network "flow control" procedures are needed to prevent the system from becoming overloaded if too much users lay a claim on the available resources in the network. Overload usually leads to a strong decrease of the performance of the system (long waiting times, a small throughput). The best known example of a flow control procedure is the so called "window flow control protocol" (WFC).

In this report we develop an approximation method for the throughput in a class of WFC-networks. These WFC-networks can be represented by queueing models. Because of the complexity of the problem we restrict ourself to a relatively simple queueing model of a virtual circuit (see Chapter 3) in the network; a model, which, however, does have the characteristic properties of a more general class of WFC-models. The special feature of the study is that the finite size of buffer spaces and the resulting overflow are explicitly taken into account.

The model consists of $M+1$ ($M \geqslant 1$) queues $Q_0, \ldots, Q_M$, which are successively visited by a fixed number N of customers. Q_M has finite waiting room with capacity L, $L<N$. If overflow occurs at Q_M, then all customers who are present in $Q_1, \ldots, Q_{M-1}$ at that moment, are sent back to Q_0, the customer who caused the overflow included. Starting in Q_0 the successive queues are visited again until another overflow takes place.

The approximation method provides good estimates for the throughput in this model, with relative errors of at most a few percent in a broad range of parameter values. The method also gives considerable insight in the behaviour of the system for varying values of the parameters, and it can be used to approximate other quantities, e.g., the mean end-to-end delay, see Section 3.2.

We now give a survey of the contents of this report.

In Chapter 2 we indicate the necessity of flow control in computer communication networks, and we present a detailed description of the window flow control protocol. We also show how a network with window flow control can be modeled as a queueing system. Because of the many references Chapter 2 can also be considered as a survey of important literature concerning flow control.

In Chapter 3, the network to be analyzed is described, and modeled as a queueing system with finite waiting room. Subsequently the model is extensively described, with a detailed discussion of assumptions about service times, service disciplines, etc.

In Chapter 4 we study a special case of the model; viz., the case that the model consists of only two queues, Q_0 and Q_1. For this case an exact analysis is possible. We derive expressions for the throughput, the mean response times in Q_0 and Q_1, and the mean end-to-end delay. For certain capacities of the waiting room at Q_1 ($L=1$, $L=N-1$) the joint distribution of the response times at

Q_0 and Q_1 is derived.
In Section 4.6 we analyze an interesting variant of the model, in which the customers can pass each other. In particular the joint distribution of the response times of this variant is compared with that of the original model.

In Chapter 5 we develop an approximation method for the throughput in the general model. The approximation results are compared with results obtained by simulation. For this purpose a simulation program has been designed, which is extensively described in Appendix B.
The approximation is shown to give exact results for the special case treated in Chapter 4.
In Section 5.6 we propose a method to approximate the mean end-to-end delay, based on the approximation method for the throughput The approximation is carried out for a few cases. Because of the promising results the method deserves a closer consideration.
The insight provided by the method leads to some interesting results; in particular, it appears that both the throughput and the mean end-to-end delay are almost independent of N, if N becomes large.
In the last section we indicate some points which are considered for further research.

Chapter 2

Flow control in computer communication networks

A computer communication network consists of a number of facilities (processors, links between processors, etc.) shared by competing users (messages, programs). The finite capacities of these facilities cause conflicts between the users. Queues of users are formed in front of certain facilities, finite buffers are filled to completion, etc. These conflicts may lead to a drastic reduction of the performance of the system, what is reflected in ever longer waiting times and a reduction of the effective throughput of the network (cf. Fig. 2.1).

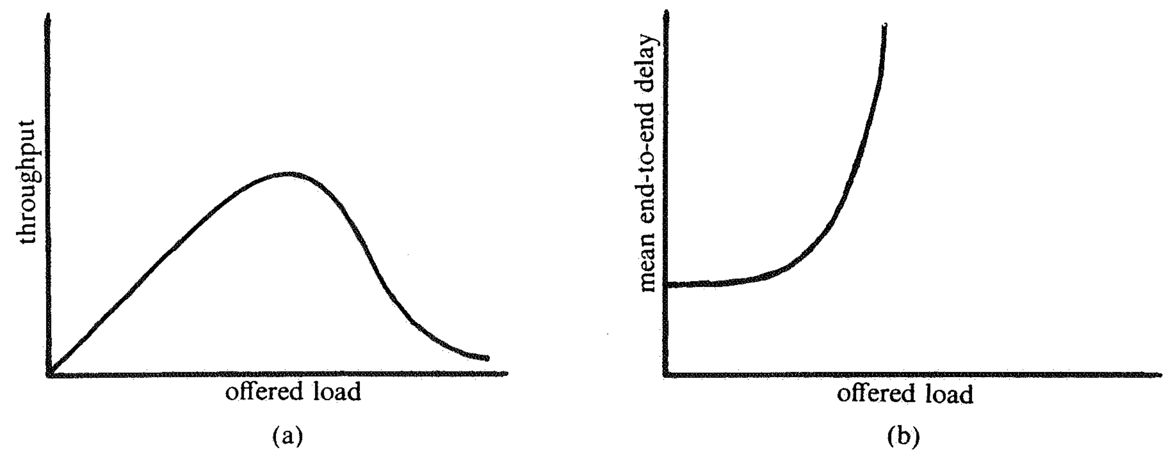


Fig. 2.1 Effective throughput (a) and mean end-to-end delay (b) versus offered load in an uncontrolled network.

If the number of users becomes too large, then the throughput may even decrease to zero because of the occurrence of "deadlock". Deadlock is usually caused by a lack of (free) buffer space in the nodes of the network: messages sent between two nodes cannot be accepted in each others' buffers.

In a computer communication network, the acceptance of a message by a receiver usually consists of (i) storing the message in a buffer, and (ii) sending an acknowledgment (ACK) to the sender. As soon as the sender gets back an ACK, he knows that there is no longer any need to store a copy of that message; a buffer space can be emptied. However, when the sender has not received an ACK within a certain period of time (the time-out) after transmission of the message, he will retransmit that message in order to prevent endlessly long waiting times.

The following simple example illustrates how a heavy traffic load can lead to deadlock. Suppose that two nodes, A and B, send each other messages which initially arrive externally at A and B (cf. Fig. 2.2). In a heavy-traffic situation, the buffers of A and B are quickly filled to completion. If A wants to transmit a message to B, then A must keep a copy of this message in his buffer space until he has received an ACK from B for it. The same holds for a message that B wants to transmit to A. Now a situation arises in which A and B endlessly transmit messages to each other which cannot be accepted due to full buffer spaces: the throughput between A and B is effectively reduced to zero. (Note: in Kleinrock [12] and Gerla and Kleinrock [9] the occurrence of deadlock in overloaded networks is extensively described).

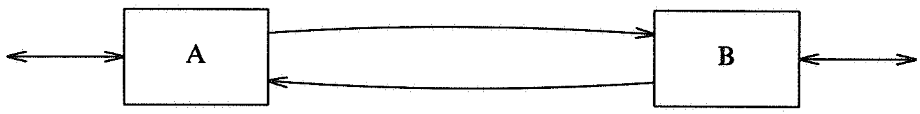


Fig. 2.2 Example of deadlock.

The phenomenon that a receiver cannot accept a message does not only occur because of a lack of free buffer space (as is the case in the above example of deadlock), but also because of transmission errors. In particular in long-distance transmission ("wide-area networks", for example in satellite communication) messages may get lost or may arrive at the receiver in distorted form. In local area networks transmission errors hardly occur, and the only reason for retransmission of a message usually is a lack of buffer space at the receiver (see Kleinrock [12] and Tanenbaum [21]). The foregoing makes it clear that a network should not without any restriction accept all traffic offered. Rules should be made to control externally offered traffic and to coordinate the traffic in the network. These rules to control and regulate information flows are called *flow control procedures.*

Figure 2.3 shows the throughput as a function of the offered traffic in a network with flow control. For comparison purposes, the same figure also displays this function for an ideal network (with supposedly perfect control of offered traffic) and for a real, uncontrolled network (cf. Fig. 2.1).

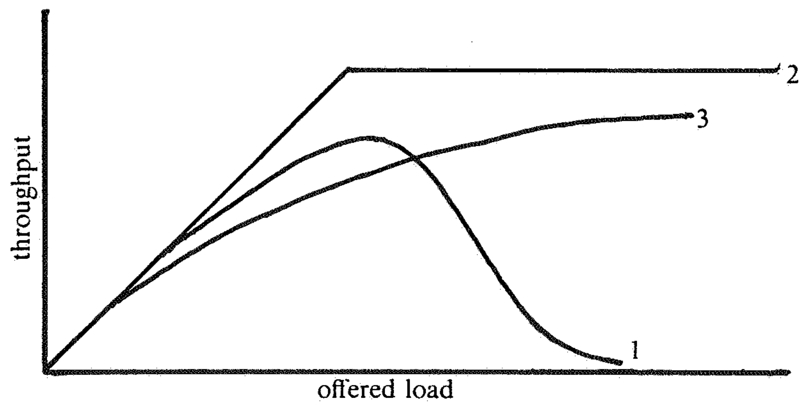


Fig. 2.3 The throughput in an uncontrolled network (1), an ideal network (2), and a network with flow control (3).

It appears that flow control prevents a sharp decrease of the throughput in the case of a large traffic offer, at the cost of a minor loss of efficiency in the case of a small traffic offer (see Gerla and Kleinrock [9] and Kleinrock [13]).

Generally a distinction is made between *local* and *global* flow control. Local flow control is the collection of procedures that regulate the traffic *within* the network, for example between two nodes; global flow control is the collection of procedures that regulate the externally offered traffic (see Gerla and Kleinrock [9]). Obviously, the application of global flow control leads to a reduction of sojourn times of messages within the network, as well as to an increase of waiting times until admission to the network. In order to be able to compare various flow control procedures, a general network performance measure has been introduced. We will leave this matter aside, referring to Kleinrock [13] for further information.

In the following our attention will be devoted to a form of global flow control that is being exercised on so-called virtual circuits of a network (a virtual circuit is a fixed route, along which messages are transmitted between a particular sender and a particular receiver, see Fig. 2.4).

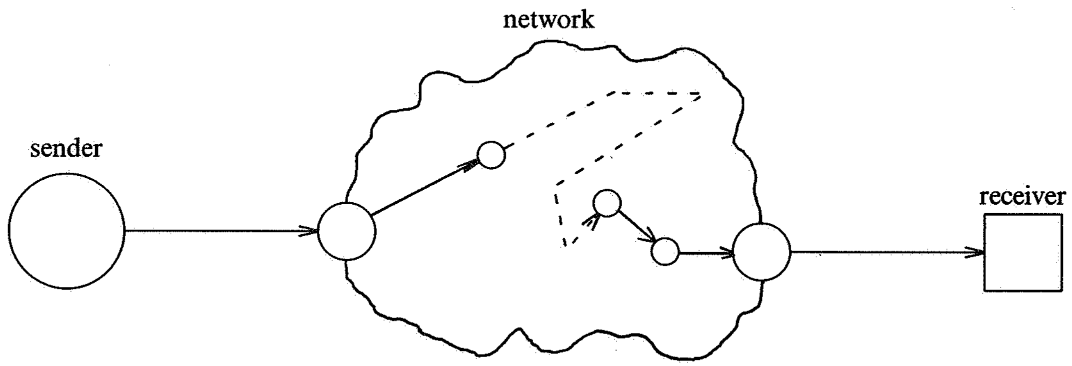


Fig. 2.4 A virtual circuit.

The most common protocol for this situation is the *window flow control protocol* (see Cerf and Kahn [5] and Reiser [15]). The principle behind this protocol is that an upper limit is imposed on the number of not yet acknowledged messages that can simultaneously be present in the virtual circuit. This maximum is called the *window size* of the virtual circuit.

Several variants of the window flow control protocol are in existence. One of them, which is called the *sliding window protocol,* plays an important role in this report, and will therefore now be studied in more detail (see Reiser [15,16]).

The sliding window mechanism operates as follows (see Fig. 2.5).

1. For each message (packet) that is being transmitted over the virtual circuit, a counter (which is initially set to N, the window size) is decreased by one.
2. As soon as the counter reaches the value zero, no more new packets are admitted to the virtual circuit (the sender stops transmission).
3. Each message is individually acknowledged by the receiver upon reception. As soon as the sender receives an ACK, the counter is increased by one. The sender is reactivated when the counter goes up from zero to one.

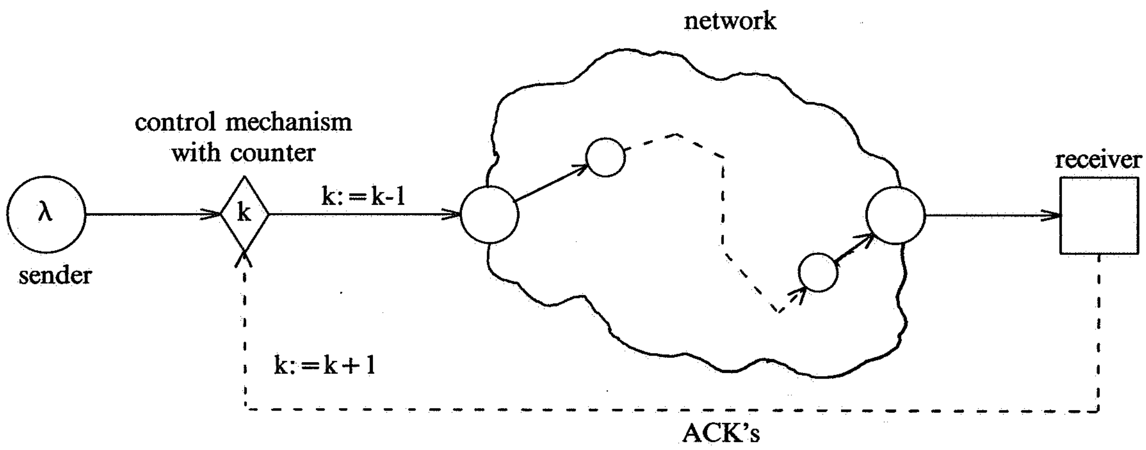


Fig. 2.5 Virtual circuit operating a sliding window.

It is often assumed that ACK's are received after a negligibly short transmission time (see Fig. 2.5).

A virtual circuit can, like other related networks (cf. Kobayashi [14] and Tobagi, Gerla et al. [22]) easily be represented by a queueing network. In that case the sender is a source that, with a certain intensity, generates customers (messages); these customers travel along the successive links of the virtual circuit. Each link can be modelled as a service system with one server, with service times that equal the transmission times of messages over this link. Thus the service time of a message at a service system is determined by the length of that message (see Reiser [15]).
For a further analysis it is important to remark that the sliding window mechanism keeps the sum of the total number of customers and of the ACK's present on the virtual circuit, and of the counter, constant and equal to N. Consequently, a virtual circuit with a sliding window mechanism can be modelled as a closed cyclic queueing network (see Fig. 2.6, and cf. Reiser [16]). It is thereby of course assumed that each node in the original system has a buffer capacity of at least N, so that retransmission is never necessary. The queue length at the service system with service intensity λ now represents the counter in the original system. Of course, the server with service intensity μ_i represents the i-th node of the virtual circuit (with average transmission time $1/\mu_i$), $i=1,...,M$.

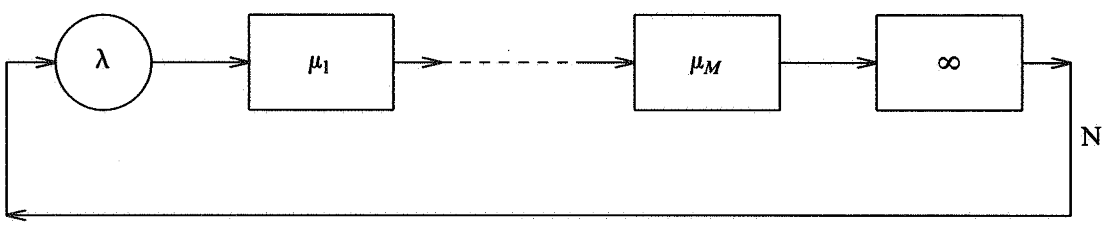


Fig. 2.6 Queueing model of the virtual circuit pictured in Fig. 2.5.
The queue with infinite service intensity represents the receiver.

A complete network, with sliding window mechanisms on the virtual circuits, can hence be modelled as a queueing network consisting of several closed chains. If we suppose that the transmission times of one and the same message at successive links are independent (Kleinrock's

"Independence Assumption", see Kleinrock [12]) and negative exponentially distributed, then an exact analysis of the joint queue-length distribution in the model is possible (Baskett et al. [1]). However, the computational complexity rapidly grows with the number of closed chains so that, actually, exact solutions are only possible in rather simple cases (see Reiser and Kobayashi [17]). In [16], Reiser presents an efficient heuristic method to handle this problem. This method yields good approximations for the throughput, the end-to-end delay of messages (i.e., the sojourn time of messages in the network), the average queue length for each node and the utilisation of the links.

We now return to the case of only one virtual circuit, with a sliding window mechanism. As remarked above, the buffer capacities of the nodes in a local area network cannot always be assumed to be infinite. Therefore it is possible that a message, upon arrival at a node, cannot be accepted because the buffer space is full. In that case this message has to be retransmitted by the sender. Clearly, a system in which blocking due to finite buffer space can occur is much harder to analyse than a system without this problem. One obvious complication is that the blocking probability at a node depends on the (unknown) throughput, which in turn depends on the blocking probability.

In the next chapter we shall describe a queueing model of a virtual circuit - with a sliding window mechanism - in which only the last node has a finite buffer capacity. As soon as overflow occurs in this node, *all* messages (customers) in the other nodes are instantaneously sent back to the sender, from which they have to be retransmitted. The rationale behind this procedure is the wish to maintain the order of the messages (or parts of them), so that it is not necessary to sort messages at the receiver. Consequently, the message that caused the overflow in the last node is the first message to be retransmitted by the sender.

It is clear from the foregoing that window flow control protocols give rise to consideration of a large number of different queueing network models (with or without blocking; various ways to send ACK's; etc.). In general these models are so complicated as to make an exact analysis of the important performance measures (throughput, end-to-end delay) prohibitive. One therefore often tries to acquire more information concerning these performance measures via approximation or simulation methods, hoping thus to be able to choose suitable (or even optimal) window sizes. Studies concerning the choice of window size for various networks have led to interesting results (see Kleinrock [12], Schwartz [20], Kermani [11] and Reiser [15]).

In this report we develop an approximation method for the analysis of a variant of the virtual circuit described above (with finite buffer capacity at the last node). A detailed discussion of the assumptions concerning service times, order of service, etc., will be presented in the next chapter.

Finally some general references to the relevant literature. Various types of computer communication networks are described in Tanenbaum [21]. Tanenbaum shows which problems can occur in the area of network communication, and how these problems are handled in practice. A good survey of the modelling and analysis of communication networks, along with many references, is presented by Reiser [15]. Here several (window) flow control procedures are shortly described and analysed. Gerla and Kleinrock [9] describe and compare flow control procedures which are applied at various levels of the network (e.g., local, global); the many references enable the reader to acquire more detailed information concerning these procedures.

Chapter 3

Model description

3.1 A virtual circuit with finite buffer space

In the preceding chapter we saw that a virtual circuit which operates a sliding window protocol can be modeled as a closed queueing network. In this chapter we will describe a virtual circuit which does operate a sliding window protocol, but has finite buffer space of size L(<N) at the last node M (see Fig. 3.1).

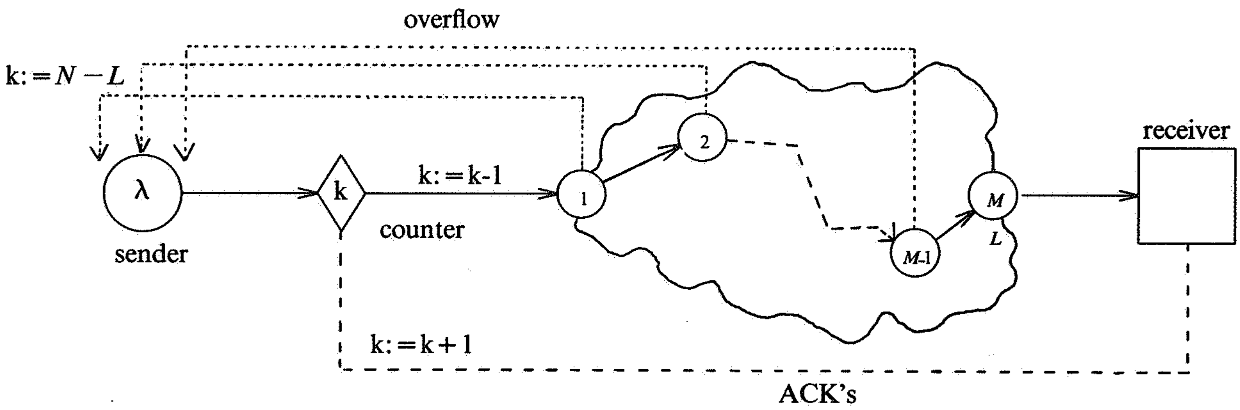


Fig. 3.1 Virtual circuit operating a sliding window protocol with window size N. Node M has finite buffer space $L < N$.

As soon as overflow occurs at this node all messages are sent back to the sender. Then these messages are sent again to the first node of the virtual circuit. In this way it is possible to save the order of the messages, so the receiver doesn't have to sort them at arrival. The message which caused the overflow is therefore the first one to be sent again.

To simplify the analysis we assume that the send process is not influenced by sending back messages in case of overflow. Thus these messages also arrive at node 1 with mean interarrival time $1/\lambda$ (λ is the send intensity).

The most important performance measures of this system are

the throughput : the mean number of messages arriving at the receiver per unit of time,

the (total) end-to-end delay :the time between the moment a message arrives at node 1 for the first time and the moment this message arrives at the receiver.

3.2 The queueing model

It is clear that we can model the system described above as a closed queueing network in the same way as the system without blocking (see Ch. 2). However, a complication arises if overflow occurs at the queue with finite waiting room, which represents the last node of the virtual circuit. In Fig. 3.2 the dotted line indicates how overflow in the system is modeled. Observe that the send process is exactly represented by the departure process from Q_0 ($\alpha_0 = 1/\lambda$), and that the number of customers present in Q_0 represents the value of the counter in the original system. In Fig. 3.2 the receiver has not been modeled because we assume that as soon as a message arrives, the receiver immediately sends an ACK which has infinitely short transmission time and delay (see Fig. 3.1).

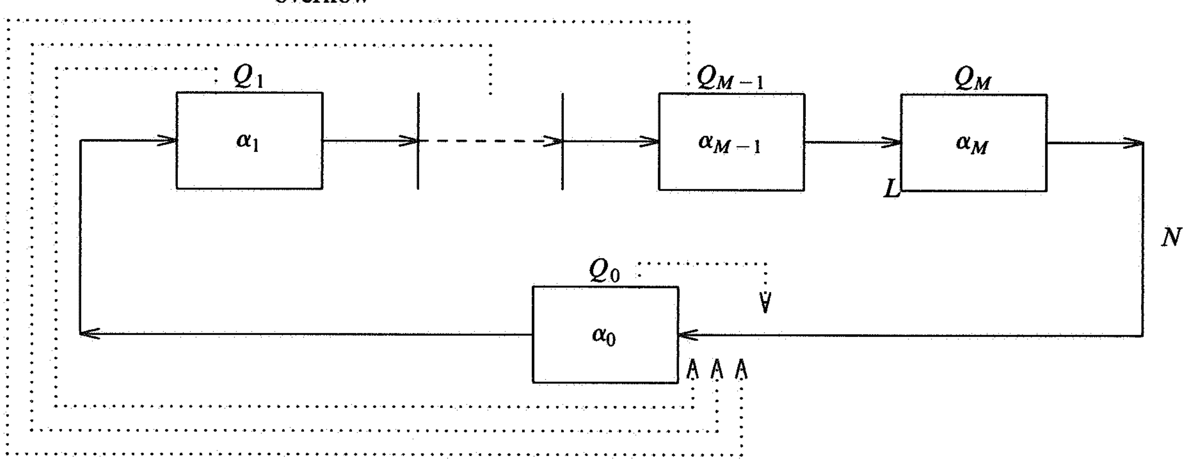


Fig.3.2 Queueing model of the virtual circuit pictured in Fig. 3.1.
α_i is the mean service time at Q_i, $i=0,...,M$, $\alpha_0 = 1/\lambda$.

Looking at Fig.2 we will now indicate in detail the assumptions we make about service times, service disciplines, etc.
$Q_1, \ldots, Q_M$ are service facilities with one server, and first come first served (FCFS) service discipline. The system contains N customers who successively visit $Q_0, \ldots, Q_M$. Q_0 can also be considered as a single server queue. Subsequent service times $\tau_i^{(1)}, \tau_i^{(2)}, \ldots$, in Q_i are independent random variables, which are negative exponentially distributed with mean α_i, $i=0,1,...,M$.
Thus

$$A_i(t) := Pr\{\tau_i^{(1)} < t\} = 1 - e^{-t/\alpha_i}, \qquad t > 0,$$
$$= 0, \qquad t \leqslant 0 \; ; \; i = 0,1.$$

$\tau_i^{(1)}, \tau_i^{(2)}, ...$, $i = 0,1,...,M$ are independent families of stochastic processes.
$Q_0, ..., Q_{M-1}$ have "infinite" waiting room (we assume that there are at least N places available at each queue). Q_M has finite waiting room of size $L-1$. , no more than L customers can be present in Q_M at the same time (one customer in service and L-1 waiting customers). We assume that N>L. As long as no overflow occurs at Q_M, service at Q_0 is also FCFS.

When a customer, say K, leaves Q_{M-1} and finds L customers in Q_M upon arrival (overflow occurs), then all customers present in $Q_1, ..., Q_{M-1}$ at that moment are sent back to Q_0, including customer K. Now there are N-L customers present in Q_0 and L customers in Q_M. The order of the customers in the network is not influenced by an overflow occurrence. If overflow takes place, the service of a customer in Q_0 (if one present) is blocked and the customers sent from $Q_1, ..., Q_{M-1}$ are placed in front of this customer, saving the original order. The server of Q_0 then immediately starts serving customer K. Service blocking and transmission of customers to Q_o in case of overflow take place instantaneously, and do not influence the service processes described above. The customers continue their way through the network until overflow takes place again.

Considering the system described above in equilibrium, we can define

T:= *throughput* : the mean number of customers leaving Q_M per unit of time (cf. the definition on page 12).

d := *(total) end-to-end delay* : the time between the first departure of a customer from Q_0 after he arrived from Q_M at Q_0, and the next departure of that customer from Q_M (cf. the definition on page 12).

$\mathbf{s}_i$:= *response time* at Q_i , $i = 0,1,....,M$.

c := *cycle time* : the time between two successive departures of the same customer from Q_M.

Our model is a closed cyclic network consisting of M queues with negative exponentially distributed service times, if $N \leq L$, because overflow will never occur. The Laplace-Stieltjes transform (L.S.T.) of both the cycle time distribution and the joint distribution of the successive response times at $Q_0, ..., Q_M$ for this system are known (Schassberger and Daduna [19], Boxma, Kelly and Konheim [4]). Remark that for $N \leq L$ the cycle time of a customer is the same as the sum of the successive response times at $Q_0, ..., Q_M$. If N>L this is not true of course, because in this case a customer can be sent back to Q_0 one or more times before arriving at Q_M.

Because of the complications arising when overflow takes place, an exact analysis seems to be impossible for the case $N > L$. We will therefore make use of an approximation method, which is developed in Chapter 5. To get more insight and to test the approximation method we consider in the next chapter a special case of our model, for which an exact analysis *is* possible: the case that the system consists of only two queues ($M = 1$). It will appear that even if we restrict ourself to the case $M = 1$, it is not possible in general to obtain distribution functions, but that we have to be content with means. For only two cases ($M = 1$, $L = 1$ and $M = 1$, $L = N-1$) we have been able to derive distribution functions (see Section 4.5).

Chapter 4

Analysis of a cyclic model with two queues

4.1 Introduction

In this chapter we analyse the queueing model of a virtual circuit with finite buffer space described in Chapter 3. However, here we restrict ourself to the case that the virtual circuit contains only one node ($M=1$). Thus we consider the model pictured in Fig. 4.1 (cf. Fig. 3.2).

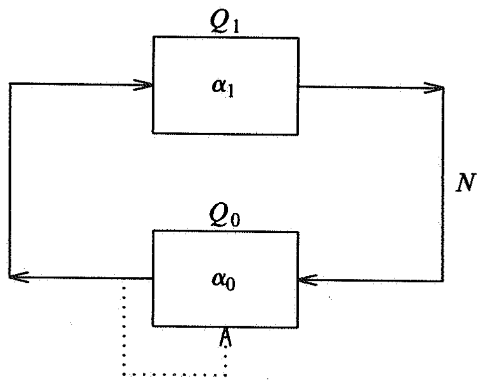


Fig. 4.1 Queueing model of a virtual circuit consisting of one node.

If in this model L customers are present at Q_1 and at the same time a customer, say K, finishes his service at Q_0, then K immediately starts a new service at Q_0. This procedure is repeated until at the end of a service of K at Q_0 there are less than L customers present in Q_1. K then moves from Q_0 to Q_1 and the service of the next customer in Q_0 is started. Starting another service time at Q_0 if overflow takes place will in the following be referred to as the "return mechanism" of the system.

In Chapter 3 we assumed that the successive service times $\tau_i^{(1)}, \tau_i^{(2)}, \cdots$ in Q_i are independent and identically distributed, and

$$A_i(t) := Pr\{\tau_i^{(1)} < t\} = 1 - e^{-t/\alpha_i}, \quad t > 0, \tag{4.1}$$
$$= 0, \qquad t \leqslant 0\ ;\ i = 0,1.$$

The service processes in Q_0 and Q_1 are independent.

The total number of customers in the system is equal to N. Thus if Q_1 contains j customers, then there are N-j customers present in Q_0, j=0,1,...,L . Because the service times are negative exponentially distributed, the number of customers present in Q_1 at a certain moment describes the state of the *whole* system.
Next to the definitions of T, **d**, **c**, $\mathbf{s}_i$, $i=1,2$, given in Chapter 3 we define

$\mathbf{x}_i$:= number of customers in Q_i, $i=0,1$,

$p_j^{(i)} := \Pr\{\mathbf{x}_i = j\}$, $j=0,1,...,N$; $i=0,1$,

y := number of customers in Q_1 seen by an arriving customer,

z := number of customers left behind in Q_1 by a departing customer.

In the following sections we shall compute the queue length distribution, the throughput, the mean cycle time and the mean end-to-end delay. For the cases $L=1$ and $L=N-1$, the cycle time distribution and the joint response time distribution will also be derived. In Section 4.6 we describe and analyse a related model where a customer who causes overflow does not immediately get a new service at Q_0, but has to join the rear of the queue. At the end of this chapter we shall compare this variant with the original model.

4.2 The queue length distribution

It is easily seen that Q_1 behaves like an M/M/1 queue with finite waiting room L (an M/M/1-L queue). Indeed, there are always at least N-L>0 customers present in Q_0, so Q_0 can be seen as a source from which customers arrive at Q_1 with negative exponentially distributed interarrival times. If there are L customers present in Q_1, the loss of an arriving customer in an M/M/1-L system is in our system exactly modeled by the return mechanism described above.

If N≤L, our model is a closed cyclic model in which no overflow occurs. It has often been remarked that Q_1 (and also Q_0) can be considered as an M/M/1-N queue in this case (see e.g. Kobayashi [14]).
Now, it is clear that the queue length process of Q_1 does *not* depend on N, if N≥L. For in that case the departure process from Q_0 is not influenced by N.

From the discussion above it follows that $p_j^{(1)}$ equals the probability that there are j customers present in an M/M/1-L queue with arrival intensity $1/\alpha_0$ and mean service time α_1. Therefore (see Cohen and Boxma [8])

$$\begin{aligned} p_j^{(1)} &= \frac{1-a}{1-a^{L+1}} a^j , \qquad a \neq 1 , \\ &= \frac{1}{L+1} , \qquad a=1 , \quad j=0,1,...,L, \end{aligned} \tag{4.2}$$

with

$$a := \frac{\alpha_1}{\alpha_0} . \tag{4.3}$$

In the preceding section we already noticed that

$$\mathbf{x}_0 = N - \mathbf{x}_1 . \tag{4.4}$$

Thus

$$p_j^{(0)} = p_{N-j}^{(1)}, \quad j = N-L,\ N-L+1, \ldots, N, \tag{4.5}$$

and hence, $p_j^{(0)}$ immediately follows from (4.2).

The mean number of customers in Q_1 is given by:

$$E\{\mathbf{x}_1\} := \sum_{j=1}^{L} j p_j^{(1)} = \frac{a}{1-a} - \frac{L+1}{1-a^{L+1}} a^{L+1}, \quad a \neq 1, \tag{4.6}$$

$$= \frac{1}{2} L, \quad a = 1.$$

Now, the mean number of customers in Q_0 can easily be calculated by using (cf. (4.4))

$$E\{\mathbf{x}_0\} = N - E\{\mathbf{x}_1\}. \tag{4.7}$$

From the analysis of the M/M/1-L model, it follows that (see Cohen [6]):

$$Pr\{\mathbf{y} = j\} = Pr\{\mathbf{z} = j\} = \frac{1-a}{1-a^L} a^j, \quad a \neq 1, \tag{4.8}$$

$$= \frac{1}{L}, \quad a = 1; \quad j = 0, 1, \ldots, L-1.$$

4.3 The throughput and the mean cycle time

In Chapter 3 we defined the throughput (T) to be the mean number of customers leaving Q_1 per unit of time.
Because the queue length distribution at Q_1 is known, the throughput can easily be found from

$$T = \frac{1}{\alpha_1}(1 - p_0^{(1)}). \tag{4.9}$$

Indeed, the departure intensity from Q_1 equals $1/\alpha_1$ as long as Q_1 contains at least one customer (an event which occurs with probability $1 - p_0^{(1)}$), and equals 0 as long as Q_1 is empty.
With the help of (4.2) we can now write

$$T = \frac{1}{\alpha_1} \frac{a - a^{L+1}}{1 - a^{L+1}}, \quad a \neq 1, \tag{4.10}$$

$$= \frac{1}{\alpha_1} \frac{L}{L+1}, \quad a = 1.$$

Remark 4.1.
It follows from (4.2) and (4.5) that

$$\frac{1}{\alpha_1}(1 - p_0^{(1)}) = \frac{1}{\alpha_0}(1 - p_L^{(1)}). \tag{4.11}$$

This is obvious because in Q_0 service completions take place with intensity $1/\alpha_0$, and a served customer moves to Q_1 if less than L customers are present in Q_1 (an event with probability $1 - p_L^{(1)}$). Therefore, the mean number of customers arriving at Q_1 equals $(1/\alpha_0)(1 - p_L^{(1)})$. Now, because we assume the system to be in equilibrium, relation (4.11) should hold.

The mean cycle time can easily be obtained by using Little's formula:

$$E\{\mathbf{c}\} = \frac{N}{T}.$$

Thus

$$E\{\mathbf{c}\} = N\frac{\alpha_1}{1-p_0^{(1)}} ,$$

or, using (4.2),

$$E\{\mathbf{c}\} = N\alpha_1\frac{1-a^{L+1}}{a-a^{L+1}} , \quad a\neq 1 , \tag{4.12}$$
$$= N\alpha_1\frac{L+1}{L} , \quad a=1 .$$

From (4.12) it follows that $E\{\mathbf{c}\}$ decreases if L increases, as long as $L \leqslant N$. Consequently, more waiting room at Q_1 leads to a smaller mean cycle time (and, of course, a higher throughput), for, then overflow occurs less often. If $L \geqslant N$, then of course $E\{\mathbf{c}\}$ and T do not depend on L.

We shall now study the behaviour of $E\{\mathbf{c}\}$ for large and for small values of a ($a\neq 1$). We rewrite (4.12) as follows

$$E\{\mathbf{c}\} = N\alpha_0(1+ 1/ \sum_{j=0}^{L-1}(\frac{1}{a})^{j+1}) , \quad a\neq 1 , \tag{4.13}$$

and also

$$E\{\mathbf{c}\} = N\alpha_1(1+ 1/ \sum_{j=0}^{L-1} a^{j+1}) , \quad a\neq 1 . \tag{4.14}$$

Now it is easily seen that for the case $L=N$ (no overflow) we can write

$$E\{\mathbf{c}\} = N\max\{\alpha_0,\ \alpha_1\}[1+O((\frac{\min\{\alpha_0,\alpha_1\}}{\max\{\alpha_o,\alpha_1\}})^N)] , \quad N\to\infty . \tag{4.15}$$

So for N sufficiently large,

$$E\{\mathbf{c}\} \approx N\max\{\alpha_0,\ \alpha_1\} . \tag{4.16}$$

Thus the cycle time is in this case almost completely determined by the queue with the slowest service time (see also Boxma [2]).
From (4.13) and (4.14) it follows that (4.16) also holds for $L<N$, if L (and N) is large enough. However, it is seen that for smaller values of L the influence of the queue with the smallest service time increases. For $L=1$, we even find that

$$E\{\mathbf{c}\} = N\alpha_0+N\alpha_1 . \tag{4.17}$$

At the end of this section we give expressions for the mean response times of Q_0 and Q_1. From (4.6) and (4.9) we find, by applying Little's formula, that

$$E\{\mathbf{s}_1\} = \frac{E\{\mathbf{x}_1\}}{T} = \alpha_1 L+\frac{\alpha_1}{1-a}-\frac{\alpha_1}{1-a}L , \quad a\neq 1 , \tag{4.18}$$
$$= \alpha_1\frac{1}{2}(L+1) , \quad a=1 .$$

The mean response time at Q_0 can now easily be calculated by means of

$$E\{\mathbf{s}_0\} = \frac{E\{\mathbf{x}_0\}}{T} = \frac{N-E\{\mathbf{x}_1\}}{T} = E\{\mathbf{c}\}-E\{\mathbf{s}_1\} .$$

We then find

$$E\{s_0\} = \frac{N\alpha_0-(N-L)\alpha_1 a^L}{1-a^L} - \frac{\alpha_1}{1-a}, \quad a\neq 1, \tag{4.19}$$

$$= N\alpha_1\frac{L+1}{L} - \frac{1}{2}\alpha_1(L+1), \quad a=1.$$

4.4 THE MEAN END-TO-END DELAY

In Section 3.2 we already defined the end-to-end delay for the general model ($M\geqslant 1$). For the case $M=1$ we can describe the end-to-end delay as follows:
Suppose a customer, say K, leaves Q_1 and arrives at Q_0. The service of K at Q_0 starts when all customers present upon his arrival have left. As soon as this service finishes, the end-to-end delay of K is started. The end-to-end delay ends at the next departure of K from Q_1.

A service (at Q_0), of which the completion initializes the end-to-end delay of a customer, will in the sequel be called the "first" service of that customer (at Q_0). The time spent by a customer at Q_0 after his first service (if overflow takes place), is called the residual response time s_0^* of that customer at Q_0 (see Fig. 4.2).

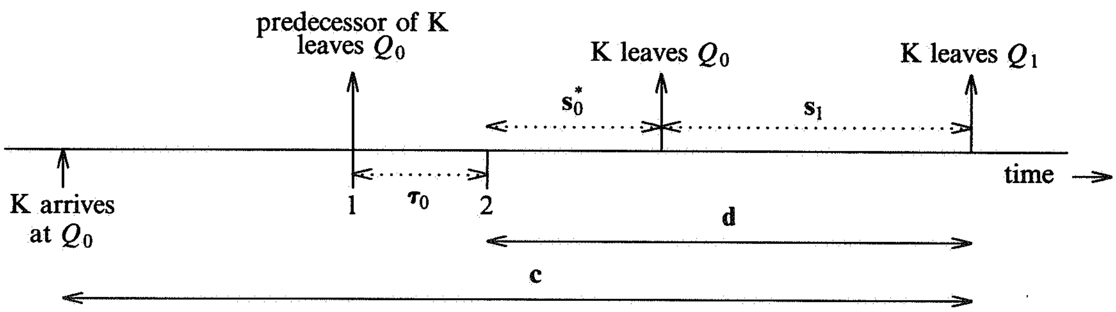


Fig. 4.2 The movement of a customer (K) through the system.
1= start of the first service of K at Q_0.
2= end of the first service of K at Q_0.

Thus

$$\mathbf{d} = s_0^* + s_1 . \tag{4.20}$$

If immediately after the first service of K at Q_0 less than L customers are present at Q_1, then of course

$$s_0^* = 0 . \tag{4.21}$$

If, however, overflow takes place after the first service of K at Q_0, then

$$s_0^* = \tau_0^* + \tau_1^* , \tag{4.22}$$

with τ_0^* and τ_1^* residual service times at Q_0 and Q_1. Indeed, it then takes a time τ_1^* before a place at Q_1 comes free and a subsequent time τ_0^* before the service of K at Q_0 is completed, and K moves to Q_1. Because the service times at Q_0 and Q_1 are negative exponentially distributed, it holds that τ_0^* and τ_1^* are independent, and

$$Pr\{\tau_0^* < t\} = A_0(t) , \quad Pr\{\tau_1^* < t\} = A_1(t) .$$

Now

$$E\{\mathbf{s}_0^*\} = P_L^* E\{\tau_0^* + \tau_1^*\} , \tag{4.23}$$

and hence from (4.20)

$$D := E\{\mathbf{d}\} = P_L^*(\alpha_0 + \alpha_1) + E\{\mathbf{s}_1\} , \tag{4.24}$$

where P_L^* is the probability that at the completion of the first service of a customer at Q_0, there are L customers present in Q_1.
For calculating P_L^* we use the fact that it concerns the *first* service of a customer, (say K again), at Q_0. This means that the last service at Q_0 was the service of the predecessor, say K-1, of K and that customer K-1 after the completion of that service moved to Q_1. Thus, at that moment there were less than L customers present in Q_1. Because K only finds L customers in Q_1 (after his first service) if K-1 found L-1 customers in Q_1 upon arrival, it holds that

$P_L^* = Pr\{$*immediately after the last service at* Q_0 *there were* $L-1$ *customers present in* Q_1, and *during the next (first) service of K, there is no departure from* Q_1 | *immediately after the last service at* Q_0 *there were less than L customers present in* $Q_1\}$.

Now, because of the independence of the service times and the memoryless property of the negative exponential distribution, it follows that

$$P_L^* = Pr\{\mathbf{y} = L-1\} Pr\{\tau_0 < \tau_1\} .$$

So (cf. (4.8))

$$\begin{aligned} P_L^* &= \frac{1-a}{1-a^L} a^{L+1} \frac{a}{1+a} , & a \neq 1 , \\ &= \frac{1}{2L} , & a = 1 . \end{aligned} \tag{4.25}$$

Hence from (4.24), (4.25) and (4.18)

$$\begin{aligned} D &= \alpha_1 \left[\frac{a^{L-1} - a^L - L}{1-a^L} + L + \frac{1}{1-a} \right] , & a \neq 1 , \\ &= \alpha_1 \left(\frac{1}{2} L + \frac{1}{L} + \frac{1}{2} \right) , & a = 1 . \end{aligned} \tag{4.26}$$

REMARK 4.2.
P_L^* can also be calculated as follows: consider the number of times a customer is served at Q_0 before he moves to Q_1. Denoting this number by the random variable $\mathbf{n}$, it holds that

$$E\{\mathbf{n}\} = 1(1-P_L^*) + 2P_L^* Pr\{\tau_0 > \tau_1\} + 3P_L^* Pr\{\tau_0 < \tau_1\} Pr\{\tau_0 > \tau_1\} + ...,$$

or

$$E\{\mathbf{n}\} = (1-P_L^*) + P_L^* \frac{a}{1+a} \sum_{j=2}^{\infty} j \left(\frac{1}{1+a} \right)^j .$$

However, $E\{\mathbf{n}\}$ can also be calculated by comparing the number of services at Q_0 with the number of departures from Q_0.

Obviously

$$E\{\mathbf{n}\} = \frac{E\{\#\ services\ at\ Q_0\ per\ unit\ of\ time\}}{E\{\#\ departures\ from\ Q_0\ per\ unit\ of\ time\}} = \frac{1/\alpha_0}{T}.$$

From the preceding and from (4.9) and (4.11) it follows that

$$P_L^* = \frac{p_L^{(1)}}{1-p_L^{(1)}} \frac{1}{1+a},$$

and hence (4.25) is obtained by using (4.2).

From (4.26) it follows that the mean end-to-end delay does not depend on N, if $L<N$. This holds because a customer's end-to-end delay does not start before the end of his first service at Q_0. Therefore, both $\mathbf{s}_0^*$ and $\mathbf{s}_1$ (and hence $\mathbf{d}$) are completely determined by the queue length distribution of Q_1 (and α_1). We already found in Section 4.1 that the queue length distribution of Q_1 is independent of N, if $L<N$.

4.5 The response time distribution and the cycle time distribution

In Section 4.3 it was seen that the mean cycle time and the mean response times at Q_0 and Q_1 can rather easily be derived. In this section we want to determine the cycle time distribution and the joint distribution of the successive response times at both queues. It is clear that this causes much bigger problems, because the successive response times at Q_0 and Q_1 in general strongly depend on each other. This, of course, is the result of the dependency of the queue lengths at Q_0 and Q_1 (see Section 4.2). For the derivation we shall closely follow a method which was developed by Boxma and Donk [3] to determine the joint response time distribution in a closed cyclic network without overflow ($L \geqslant N$). However, because of the complications arising from the possibility of overflow in our model, it will appear that only for two simple cases ($L=1$ and $L=N-1$) we are able to derive the joint distribution of the response times of a customer at Q_0 and Q_1.

Now, we shall briefly describe the method used for the model with $L \geqslant N$, see Fig. 4.3. The quantities concerning this model will be indicated by a hat (ˆ).
Boxma and Donk consider the model, in equilibrium, at departure epochs from $\hat{Q}_1$.

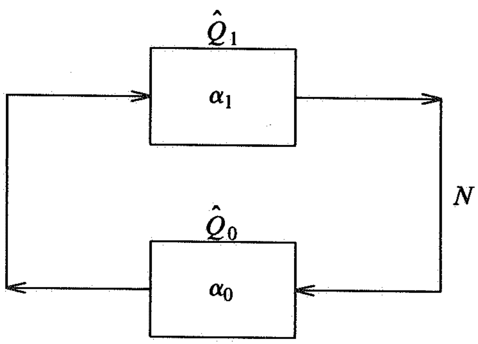


Fig. 4.3 Closed cyclic model with N customers and infinite waiting room at Q_0 and Q_1.

For a departing customer, say $\hat{\text{K}}$, the joint distribution of the past response time $\hat{\mathbf{s}}_1$ at $\hat{Q}_1$ and the successive response time $\hat{\mathbf{s}}_0$ at $\hat{Q}_0$ is derived by conditioning on the number of customers $\hat{\mathbf{z}}_1$ that is left behind in Q_1 by customer K.

Using Laplace-Stieltjes transforms it holds that

$$E\{e^{-\rho\hat{s}_0-\eta\hat{s}_1}\} = \sum_{i=0}^{N-1} Pr\{\hat{z}=i\}E\{e^{-\rho\hat{s}_0-\eta\hat{s}_1}|\hat{z}=i\} \tag{4.27}$$

$$= \sum_{i=0}^{N-1} Pr\{\hat{z}=i\}E\{e^{-\eta\hat{s}_1}|\hat{z}=i\}E\{e^{-\rho\hat{s}_0}|\hat{z}=i,\ \hat{s}_1\}\ ,\ \ \mathrm{Re}\ \rho,\eta\geqslant 0.$$

If customer K leaves behind i customers in $\hat{Q}_1$, then he sees $N-i-1$ customers in $\hat{Q}_0$ upon arrival. Hence, because of the memoryless property of the negative exponential distribution, it holds that

$$E\{e^{-\rho\hat{s}_0}|\hat{z}=i,\ \hat{s}_1\} = \left[\frac{1}{1+\alpha_0\rho}\right]^{N-i}\ ,\ \ \mathrm{Re}\ \rho\geqslant 0\,. \tag{4.28}$$

Next it is remarked that the queue length process of an M/M/1-N queue is a birth and death process and hence reversible. For a reversible stochastic process $\{x(t),\ t\geqslant 0\}$ it holds that

$$(x(t_1),\ x(t_2),\ \ldots,\ x(t_n)) \text{ and } (x(\tau-t_1),\ x(\tau-t_2),\ \ldots,\ x(\tau-t_n))$$

have the same distribution for all $t_1, t_2, \ldots, t_n, \tau$ (see Kelly [10]); (Kelly remarks: "Speaking intuitively, if we take a film of such a process and then run the film backwards, the resulting process will be statistically indistinguishable from the original process").
Denoting by $\hat{y}$ the number of customers in $\hat{Q}_0$ that is seen by an arriving customer it follows, because of the reversibility of the queue length process of $\hat{Q}_1$, that

$$E\{e^{-\eta\hat{s}_1}|\hat{z}=i\} = E\{e^{-\eta\hat{s}_1}|\hat{y}=i\}\ ,\ \ \mathrm{Re}\ \eta\geqslant 0,\ i=0,\ldots,\ N-1. \tag{4.29}$$

Because (cf. (4.28))

$$E\{e^{-\eta\hat{s}_1}|\hat{y}=i\} = \left[\frac{1}{1+\alpha_1\eta}\right]^{i+1}\ ,$$

it follows from (4.27), (4.28) and (4.29) that

$$E\{e^{-\rho\hat{s}_0-\eta\hat{s}_1}\} = \sum_{i=0}^{N-1} Pr\{\hat{z}=i\}\left[\frac{1}{1+\alpha_0\rho}\right]^{N-i}\left[\frac{1}{1+\alpha_1\eta}\right]^{i+1}\ ,\ \ \mathrm{Re}\ \rho,\eta\geqslant 0\ , \tag{4.30}$$

with (cf. (4.8))

$$Pr\{\hat{z}=i\} = Pr\{\hat{y}=i\} = \frac{1-a}{1-a^N}a^i\ ,\quad a\neq 1\ ,$$
$$= \frac{1}{N}\ ,\qquad a=1;\ i=0,1,\ldots,N-1.$$

The cycle time distribution can easily be obtained by substituting $\eta=\rho$ in (4.30).

We now hope to be able to apply the method described above to our model *with* overflow, for Q_1 behaves like an M/M/1-L queue of which the queue length process is reversible.
Again, we consider the system at departure epochs from Q_1 to derive the joint distribution of the past response time s_1 at Q_1 and the successive response time s_0 at Q_0 of a leaving customer K. If K leaves i customers behind in Q_1 then it is easily seen that the L.S.T. of the distribution of s_1 is given by (cf. (4.29))

$$E\{e^{-\eta s_1}|z=i\} = E\{e^{-\eta s_1}|y=i\} = \left[\frac{1}{1+\alpha_1\eta}\right]^{i+1}\ ,\ \ \mathrm{Re}\ \eta\geqslant 0,\ i=0,1,\ldots,L-1. \tag{4.31}$$

However, complications arise when trying to derive the L.S.T. of the distribution of s_0. Indeed, because overflow may take place, the distribution of the response time of K at Q_0 is not given by the distribution of the sum of $N-i$ independent service times at Q_0 (cf. (4.28)). The response time (distribution) of K at Q_0 does not only depend on the number of customers seen in Q_0 upon arrival, but also on the departure process from Q_1 during that (response) time. Therefore, in general, it is extremely difficult to derive the response time distribution of K at Q_0. However, for some special cases a solution can be presented. Indeed, if $L=1$ or $L=N-1$ then the problems arising in the general case can be solved. In spite of these limitations we shall analyse these two special cases to get more insight in the system. The case $L=1$ shall later be compared with the variant already mentioned, in which a customer who causes overflow does not immediately get a new service, but has to join the end of the queue (see Section 4.6).

For the case $L=1$ a customer, say K_N, who leaves Q_1, always sees $N-1$ customers in Q_0 upon arrival (see Fig. 4.4a). Numbering the customers from 1 to N (see Fig. 4.4a), it takes a residual service time $\tilde{\tau}_0^{(1)}$ before customer K_1 moves to Q_1. Then, K_1 starts a service at Q_1 and K_2 starts a service at Q_0. During the service time $\tau_0^{(1)}$ of K_1 at Q_1 no customers depart from Q_0, and by means of the return mechanism the service of K_2 at Q_0 is possibly repeated a number of times (see Fig. 4.4b).

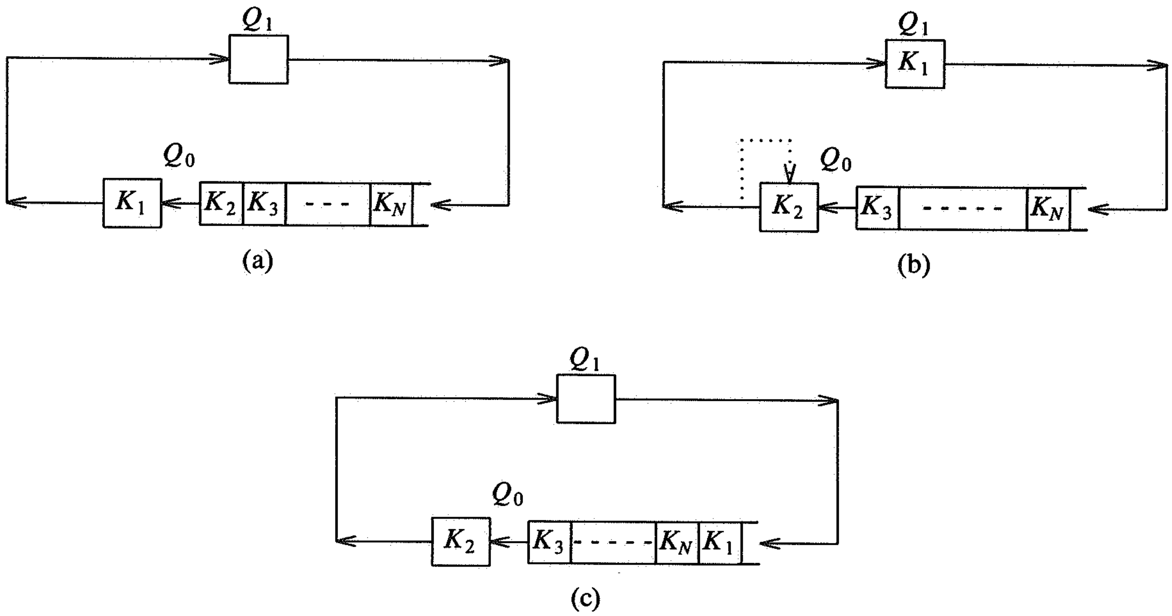


Fig. 4.4 The movement of the customers through the system for the case $L=1$.

After a time $\tau_1^{(1)}$ K_1 departs from Q_1 and, of course, leaves no customers behind. Now the situation is as in the beginning (see Fig. 4.4a), with the exception that all the customers have shifted one place (see Fig. 4.4c). During the response time s_0 of K_N at Q_0 this repeats until K_N himself leaves Q_0. Therefore it holds that

$$s_0 = \tilde{\tau}_o^{(1)} + \tau_1^{(1)} + \tilde{\tau}_0^{(2)} + \ldots + \tau_1^{(N-1)} + \tilde{\tau}_0^{(N)} , \qquad (4.32)$$

with $\tilde{\tau}_0^{(j)}$ a residual service time of K_j at Q_0, $j=0,1,\ldots,N$.
Because all service times are independent and negative exponentially distributed it follows from (4.32) that (cf. (4.22))

$$Pr\{s_0<t\} = A_0^{N*}(t)*A_1^{(N-1)*}(t) \ . \tag{4.33}$$

An arriving customer always sees Q_1 empty, so

$$Pr\{s_1<t\} = A_1(t) \ . \tag{4.34}$$

Remark that in this case (L=1), the successive response times at Q_0 and Q_1 are *independent.*
From the relations stated above it follows that (cf. (4.30))

$$E\{e^{-\rho s_0-\eta s_1}\} = \left[\frac{1}{1+\alpha_0\rho}\right]^N \left[\frac{1}{1+\alpha_1\rho}\right]^{N-1} \frac{1}{1+\alpha_1\eta} \ . \tag{4.35}$$

The cycle time distribution is given by

$$Pr\{c<t\} = A_0^{N*}(t)*A_1^{N*}(t) \ . \tag{4.36}$$

From (4.36) it easily follows that $E\{c\} = N\alpha_0+N\alpha_1$; this result was already found in Section 4.3.

For the case $L=N-1$, the derivation of the joint distribution of the response times is a bit more complicated because the number of customers left behind in Q_1 by a departing customer, say K, can vary from 0 to $L-1$. Suppose K leaves behind i customers in Q_1, i=0,1,..., $L-1$. Consequently, K sees $N-1-i$ customers in Q_0 upon arrival. Because $L=N-1$, we know that these $N-1-i$ customers will be served at Q_0 exactly once, and next move to Q_1. Hence, the waiting time w_0 of K at Q_0 is given by the sum of a residual service time and $N-i-2$ complete service times at Q_0. We now derive the L.S.T. of the conditional distribution of s_0 by looking at the number of customers who depart from Q_1 during the waiting time of K at Q_0. In particular, we consider the customer who starts his service at Q_1 immediately after the departure of K from that queue. The service time of this customer, say K', will be indicated by the random variable τ_1. (We assume that K leaves behind at least one customer, thus $1\leqslant i\leqslant L-1$). We now write

$$Pr\{s_0<t|z=i\} = Pr\{s_0<t,w_0<\tau_1|z=i\}+Pr\{s_0<t,w_0\geqslant\tau_1|z=i\} \ , \quad 1\leqslant i\leqslant L-1 \ . \tag{4.37}$$

For the case that $w_0<\tau_1$, the residual service time of K at Q_0 is indicated by $\tilde{\tau}_0$ (from the moment on that K' leaves Q_1).
Defining

$$\hat{\tau}_0 := \tau_0 \ , \quad \text{if } \tau_1\leqslant w_0 \ ,$$
$$:= \tilde{\tau}_0 \ , \quad \text{if } \tau_1>w_0 \ ,$$

it holds that

$$s_0 = \max\{w_0, \tau_1\}+\hat{\tau}_0 \ . \tag{4.38}$$

It is easily seen that $\max\{w_0, \tau_1\}$ and $\hat{\tau}_0$ are independent and that $\hat{\tau}_0$ is negative exponentially distributed with mean α_0.
If $w_0<\tau_1$, we can write

$$\tau_1 = w_0+\tilde{\tau}_1 \ , \tag{4.39}$$

with $\tilde{\tau}_1$ a residual service time at Q_1. w_0 and $\tilde{\tau}_1$ are independent, and $\tilde{\tau}_1$ is a negative exponentially distributed variable with mean α_1 (w_0 is the past service time of K' and $\tilde{\tau}_1$ is the residual service time of K' at Q_1).
Now, from (4.38) and (4.39) it follows that

$$Pr\{s_0<t, \ w_0<\tau_1|z=i\} = Pr\{w_0<t, \ w_0<\tau_1|z=i\}*A_1(t)*A_0(t) \ , \tag{4.40}$$

and

$$Pr\{\mathbf{s}_0<t,\ \mathbf{w}_0\geqslant\boldsymbol{\tau}_1|\mathbf{z}=i\} = Pr\{\mathbf{w}_0<t,\ \mathbf{w}_0\geqslant\boldsymbol{\tau}_1|\mathbf{z}=i\}*A_0(t)\,. \tag{4.41}$$

Using

$$Pr\{\mathbf{w}_0<t,\ \mathbf{w}_0<\boldsymbol{\tau}_1|\mathbf{z}=i\} = \int_0^t (1-A_1(\tau))\, d_\tau A_0^{(N-i-1)*}(\tau)\ ,$$

and

$$Pr\{\mathbf{w}_0<t,\ \mathbf{w}_0\geqslant\boldsymbol{\tau}_1|\mathbf{z}=i\} = \int_0^t A_1(\tau)\, d_\tau A_0^{(N-i-1)*}(\tau)\ ,$$

it follows from (4.37), (4.40) and (4.41) that

$$E\{e^{-\rho\mathbf{s}_0}|\mathbf{z}=i\} = \left[\left[\frac{1}{1+\alpha_0\rho}\right]^{N-i-1} - \left[\frac{1}{1+\alpha_0(\rho+\frac{1}{\alpha_1})}\right]^{N-i-1}\right]\frac{1}{1+\alpha_0\rho} + \left[\frac{1}{1+\alpha_0(\rho+\frac{1}{\alpha_1})}\right]^{N-i-1}\frac{1}{1+\alpha_0\rho}\,\frac{1}{1+\alpha_1\rho}\ ,\ i=1,\ldots,L-1\ ,\ \mathrm{Re}\ \rho\geqslant 0\,. \tag{4.42}$$

It is easily seen that

$$E\{e^{-\rho\mathbf{s}_0}|\mathbf{z}=0\} = \frac{1}{1+\alpha_0\rho}E\{e^{-\rho\mathbf{s}_0}|\mathbf{z}=1\}\ . \tag{4.43}$$

Now, analogous to (4.30), the L.S.T. of the joint distribution of the response times of K at Q_0 and Q_1 can be derived (and also the L.S.T. of the cycle time distribution). We get

$$\begin{aligned} E\{e^{-\rho\mathbf{s}_0-\eta\mathbf{s}_1}\} = {} & \sum_{i=1}^{L-1} Pr\{\mathbf{z}=i\}\left[\frac{1}{1+\alpha_1\eta}\right]^{i+1}\left[\left[\left[\frac{1}{1+\alpha_0\rho}\right]^{N-i-1} - \left[\frac{1}{1+\alpha_0(\rho+\frac{1}{\alpha_1})}\right]^{N-i-1}\right]\frac{1}{1+\alpha_0\rho} + \left[\frac{1}{1+\alpha_0(\rho+\frac{1}{\alpha_1})}\right]^{N-i-1}\frac{1}{1+\alpha_0\rho}\,\frac{1}{1+\alpha_1\rho}\right] \\ & + Pr\{\mathbf{z}=0\}\frac{1}{1+\alpha_1\eta}\,\frac{1}{1+\alpha_0\rho}\left[\left[\left[\frac{1}{1+\alpha_0\rho}\right]^{N-2} - \left[\frac{1}{1+\alpha_0(\rho+\frac{1}{\alpha_1})}\right]^{N-2}\right]\frac{1}{1+\alpha_0\rho} \right. \\ & \left. + \left[\frac{1}{1+\alpha_0(\rho+\frac{1}{\alpha_1})}\right]^{N-2}\frac{1}{1+\alpha_0\rho}\,\frac{1}{1+\alpha_1\rho}\right],\quad \mathrm{Re}\ \rho,\ \eta\geqslant 0\,, \end{aligned} \tag{4.44}$$

with $Pr\{\mathbf{z}=i\}$ given by (4.8).

For the case $L=1$, $N=2$, (4.35) and (4.44) should be identical, which can easily be checked.

4.6 A VARIANT OF THE SERVICE DISCIPLINE AT Q_0

In the preceding sections we assumed that a customer, who caused overflow, immediately received a new service at Q_0. This was repeated until at the end of a service there were less than L customers present in Q_1. In this section we shall analyse a related model, which only differs in the way a customer causing overflow is handled. We shall assume that such a customer does not immediately get a new service, but has to join the end of the queue (of Q_0, see Fig. 4.5). Thus only for the case $L=N-1$, it is possible that a customer gets two or more subsequent services at Q_0 (remark that if $L=N-1$, this variant is equal to the original model).

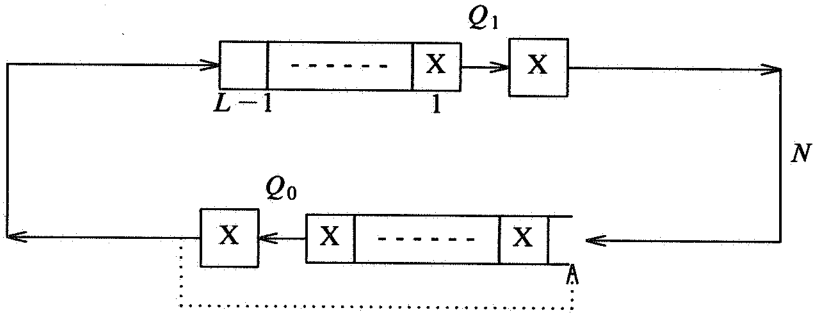


Fig. 4.5 Closed cyclic model with finite waiting room at Q_1.
A customer who causes overflow is placed at the end of the queue.

The most important difference with the original model is that in this model, (see Fig. 4.5), customers can *pass* each other at Q_0. It is easily seen that this does not influence the queue length distribution, the mean cycle time and the throughput. So, the results obtained in Section 4.2 and Section 4.3 still hold. We shall not define the end-to-end delay for this model. Of course, the joint distributions of the response times at Q_0 and Q_1 are different for the two models. In general it is extremely difficult to calculate cycle times and response times for systems in which customers can pass each other. Therefore, we shall restrict ourself to the case $L=1$.

To derive the L.S.T. of the joint distribution of the response times, we shall closely follow the method used in Section 4.5. We consider the system at a departure epoch of a customer, say K_N, from Q_1. K_N leaves no customers behind (in Q_1) and sees $N-1$ customers in Q_0 upon arrival. The situation at that moment is pictured in Fig. 4.6a on page 26 (the other customers are numbered from 1 to $N-1$).
After completion of the residual service time $\tilde{\boldsymbol{\tau}}_0^{(1)}$ of K_1 at Q_0, the situation is as shown in Fig. 4.6b. Denoting by $\mathbf{r}_0^{(1)}$ the (residual) response time of K_N at Q_0 measured from that moment, the total response time $\mathbf{s}_0$ of K_N at Q_0 is given by

$$\mathbf{s}_0 = \tilde{\boldsymbol{\tau}}_0^{(1)} + \mathbf{r}_0^{(1)} . \tag{4.45}$$

$\tilde{\boldsymbol{\tau}}_0^{(1)}$ and $\mathbf{r}_0^{(1)}$ are independent, and $\tilde{\boldsymbol{\tau}}_0^{(1)}$ is negative exponentially distributed with mean α_0.

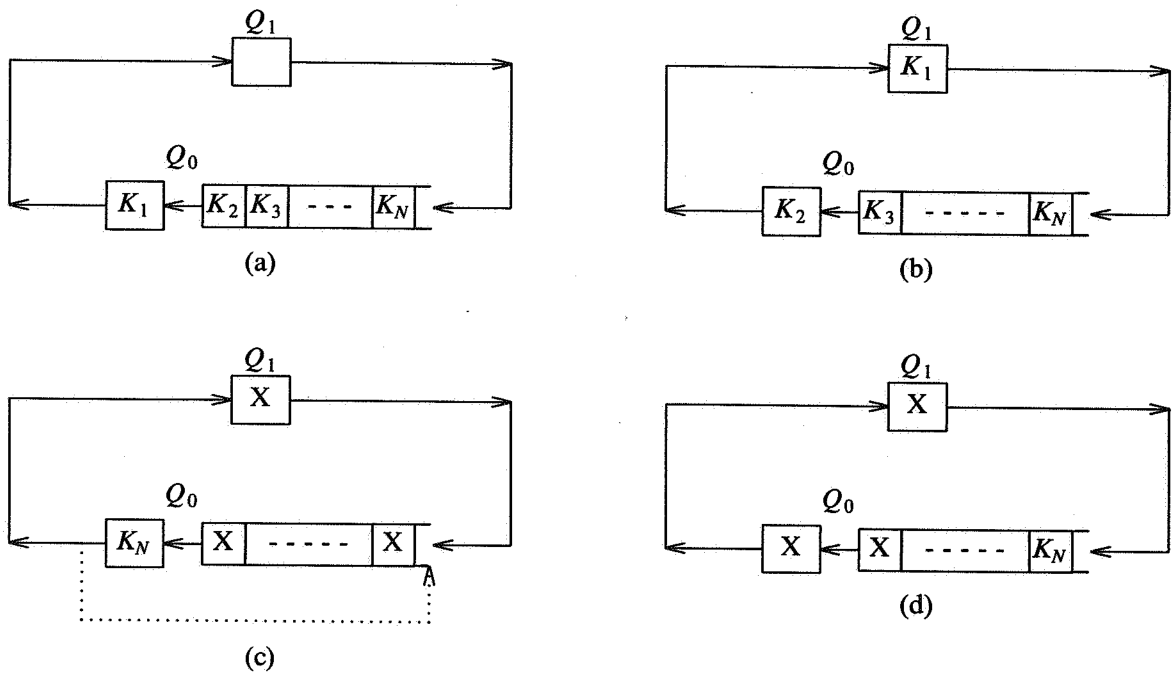


Fig. 4.6 The movement of the customers through the system.

Because of the service discipline mentioned above, it takes $N-2$ service times at Q_0 before K_N gets his first service after the start of his residual response time at Q_0. At the start of this first service the state of the system is as pictured in Fig. 4.6c. In this figure only the position of K_N is indicated because the order of the customers in the system may be changed. We remark that at this moment there is always *one* customer present in Q_1. If the (residual) service time $\tilde{\tau}_1$ of this customer in Q_1 is smaller than the service time $\tau_0^{(N)}$ of K_N in Q_0, then

$$\mathbf{r}_0^{(1)} = \tau_0^{(2)} + \cdots + \tau_0^{(N-1)} + \tau_0^{(N)}, \quad \tau_0^{(N)} > \tilde{\tau}_1 . \tag{4.46}$$

However, if the (residual) service time of the customer in Q_1 is larger than the service time of K_N at Q_0, then K_N, again, has to join the end of the queue after the completion of his service (see Fig. 4.6d). Concerning the response time of K_N at Q_0, however, now the situation is the same as at the start of his residual response time $\mathbf{r}_0^{(1)}$. The residual response time of K_N at Q_0, measured from this epoch, will be denoted by $\mathbf{r}_0^{(2)}$.
Thus

$$\mathbf{r}_0^{(1)} = \tau_0^{(2)} + ... + \tau_0^{(N-1)} + \tau_0^{(N)} + \mathbf{r}_0^{(2)}, \quad \tau_0^{(N)} < \tilde{\tau}_1 , \tag{4.47}$$

with

$$Pr\{\mathbf{r}_0^{(1)} < t\} = Pr\{\mathbf{r}_0^{(2)} < t\} . \tag{4.48}$$

Because all service times in the system are independent and negative exponentially distributed, it follows from (4.46), (4.47) and (4.48) that

$$Pr\{\mathbf{r}_0^{(1)} < t\} = A_0^{(N-2)*}(t) * \left[Pr\{\tau_0^{(N)} < t,\ \tau_0^{(N)} < \tilde{\tau}_1\} * Pr\{\mathbf{r}_0^{(1)} < t\} + Pr\{\tilde{\tau}_1 < \tau_0^{(N)} < t\} \right] . \tag{4.49}$$

Then, analogous to the derivation of (4.42), we get

$$E\{e^{-\rho \mathbf{r}_0^{(1)}}\} = \left[\frac{1}{1+\alpha_0 \rho} \right]^{N-2} \left[\frac{a}{1+a+\alpha_1 \rho} E\{e^{-\rho \mathbf{r}_0^{(1)}}\} + \frac{1}{1+\alpha_0 \rho} - \frac{a}{1+a+\alpha_1 \rho} \right] , \tag{4.50}$$

Now, $E\{e^{-\rho r_0^{(1)}}\}$ can easily be obtained from (4.50). Using (4.45) it follows that

$$E\{e^{-\rho s_0}\} = \left[\frac{1}{1+\alpha_0\rho}\right]^N \frac{1}{1+a+\alpha_1\rho-a\,/\,(1+\alpha_0\rho)^{N-2}}\,, \tag{4.51}$$

and hence (cf. (4.34) and (4.35))

$$E\{e^{-\rho s_0-\eta s_1}\} = \frac{1}{1+\alpha_1\eta}\left[\frac{1}{1+\alpha_0\rho}\right]^N \frac{1}{1+a+\alpha_1\rho-a\,/\,(1+\alpha_0\rho)^{N-2}}\,,\quad \mathrm{Re}\ \rho,\eta\geqslant 0\,. \tag{4.52}$$

The L.S.T. of the cycle time distribution is obtained by substituting $\eta=\rho$ in (4.52).

From the description of the service discipline given at the beginning of this section, it is (intuitively) clear that the variance of the response time at Q_0 is larger for this variant than for the original model. It is interesting to compare both models in this respect.
By differentiating (4.51) with respect to ρ twice, and next substituting $\rho=\eta$, we find

$$E\{s_0^2\}_{variant} = 2(N-1)^2\alpha_1^2+N(N+1)\alpha_0^2+2N(N-1)\alpha_0\alpha_1+a(N-2)(N-1)\alpha_0^2\,. \tag{4.53}$$

Analogously, it follows from (4.35) (with $\eta=0$) that

$$E\{s_0\}_{orig.} = N(N+1)\alpha_0^2+2N(N-1)\alpha_0\alpha_1+(N-1)N\alpha_1^2\,. \tag{4.54}$$

Since

$$E\{s_0\}_{variant} = E\{s_0\}_{orig.}\,,$$

it follows from (4.53) and (4.54), that

$$Var\{s_0\}_{variant}-Var\{s_0\}_{orig.} = (N-1)(N-2)(\alpha_1^2+\alpha_0\alpha_1)\,. \tag{4.55}$$

Thus the difference between the variances rapidly increases, if N grows. If $N=2$ the difference is equal to zero. As we remarked before, the two service disciplines are the same in that case ($N=2$, $L=1$; cf. (4.35) and (4.52)).

From (4.55) it is seen that for small values of α_1 ($\alpha_1<<1$) the difference between the variances is small. This holds because in that case the overflow probability is small (if, in addition, $\alpha_1<\alpha_0$, cf. (4.2)), and hence the service disciplines just slightly differ.

Chapter 5

Analysis of the general model - an approximation

5.1 Introduction

In this chapter we shall analyse the general model ($M \geqslant 1$) of the virtual circuit with finite buffer space (see Fig. 3.2). We have seen in the preceding chapter that, for the case $M = 1$, an exact analysis of this model is possible. This is mainly due to the fact that Q_M $(=Q_1)$ in this case behaves like an M/M/1-L queue, so that we can use the well-known results for that queueing model (e.g., the queue length distribution). The analysis of the model of only two queues is further simplified by the strong relation between the queue lengths at Q_0 and Q_1 (cf. (4.4)).

It is obvious that the general model is much harder to analyse. Firstly, Q_M for $M>2$ is not behaving like an M/M/1-L queue, because the arrival process at Q_M no longer is a Poisson process. Secondly, the fact that customers are sent back to Q_0 when overflow occurs leads here to extra complications, because the number of customers sent back depends on the number of customers present in $Q_1, \ldots, Q_{M-1}$ at the time of overflow.

The above-mentioned complications force us to refrain from an exact analysis of this (general) model. In the following sections we shall therefore develop an approximation method in order to obtain some insight in the behaviour of the most important performance measures.

The approximation method is mainly developed to estimate the throughput; indeed it yields (very) accurate results in most instances. The main tool to test the approximation has been simulation. A simulation program has been written that estimates the most important performance measures of the network; it can be found in Appendix B. For the case $M = 1$ the approximation results for the throughput become exact.

Due to the rather complicated structure of the end-to-end delay (see Chapter 3), it is much harder to derive a sharp approximation for this performance measure. However, the construction of the approximation method yields much insight, already allowing us to draw some interesting conclusions. For example, the throughput turns out to depend only slightly on the number of customers (N), while the mean end-to-end delay increases with N but stays below a certain limit as $N \uparrow \infty$. These results are consistent with what we have found in the preceding chapter for $M = 1$; indeed, in that case the throughput and mean end-to-end delay are completely independent of N (cf. (4.10) and (4.26)).

In the next section we shall describe the approximation method globally, stating some general observations. A detailed description of the method follows in Section 5.3. Subsequently we discuss some examples, comparing the approximation results with simulation results. The influence of N on throughput and mean end-to-end delay is also studied. Finally in Section 5.7 we mention the main conclusions, and we point at some problems for further investigation.

5.2 A GLOBAL DESCRIPTION OF THE METHOD

In this section we present a global description of the approximation method; we proceed in a stepwise fashion, guided by a few simple observations. Hereto we consider Figure 5.1, which once more displays the model (cf. Fig. 3.2).

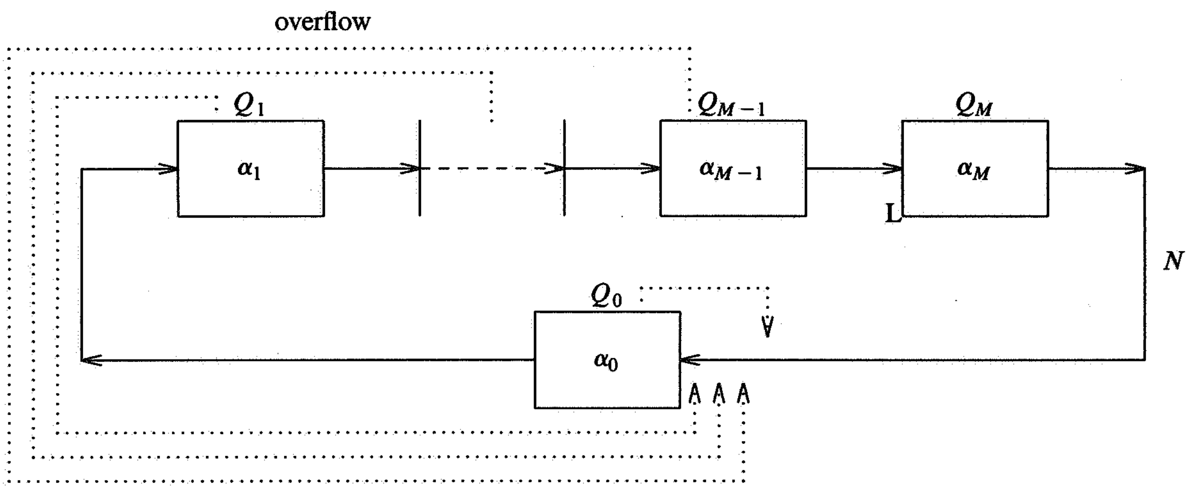


Fig. 5.1 Closed model of a virtual circuit with finite buffer space at the last node in a network with window flow control.

Our main goal is to approximate the throughput of this model, i.e., the average number of customers moving from Q_M to Q_0 per unit of time. As the system is assumed to be in equilibrium, the throughput also equals the average number of customers moving from Q_{M-1} to Q_M per unit of time. It would therefore be logical to estimate the number of customers who move from Q_{M-1} to Q_M during a - lengthy - period of time t, on the basis of the parameters $M, N, L, \alpha_0, \ldots, \alpha_M$. We do not know how large t should be before a good approximation of the throughput can be given; but we can restrict t to the length of one so-called regeneration period, which is representative for a very large period of time. To see this consider the following observation.

OBSERVATION 5.1.
Immediately after the occurrence of overflow at Q_M there are L customers in Q_M and $N-L$ customers in Q_0 (see Chapter 3).

From this observation we derive that, because of the fact that the service times are negative exponentially distributed, the system returns to the same state each time that overflow occurs. We can hence divide the time axis into successive cycles, the lengths of which are given by the periods of time between successive overflow epochs (see Fig. 5.2).

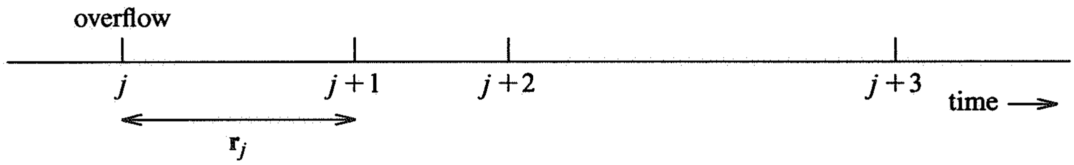


Fig. 5.2 The regenerative cycles.

The queue length and departure processes (in particular the departure process from Q_{M-1}) hence repeat themselves, probabilistically speaking, each time overflow occurs.

It can be easily seen that the corresponding successive cycle times $\mathbf{r}_j$, $j=0,1,...$, are independent and identically distributed random variables, so that the departure process from Q_{M-1} is a regenerative process (w.r.t. $\{\mathbf{r}_j, j=0,1,...\}$; see Cohen [7]). Because of the regenerative character of the departure process from Q_{M-1} it suffices to determine the average number of customers moving from Q_{M-1} to Q_M during just one cycle (of length $\mathbf{r}_j$). The following holds:

$$T = \frac{E\{\textit{number of counted customers during } \mathbf{r}_j\}}{E\{\mathbf{r}_j\}}. \tag{5.1}$$

$E\{\mathbf{r}_j\}$ (the average time between two successive overflows) depends strongly on the number of customers arriving at and departing from Q_M. Therefore we first concentrate on "counting" the customers who move from Q_{M-1} to Q_M during one cycle. First we define K_j to be the customer who causes the j-th overflow at Q_M, $j=1,2,...$ (see Fig. 5.2). Hence K_j is the customer who, at the end of the (j-1)-th cycle, finds L customers present at Q_M upon his departure from Q_{M-1}. All customers in $Q_1, \ldots, Q_{M-1}$, including K_j, are sent back to Q_0, where a new service of K_j is immediately started. It is now easy to make the following

OBSERVATION 5.2.
As long as K_j, after the j-th overflow, has not yet left Q_{M-1}, no customers move from Q_{M-1} to Q_M and hence no contribution to the throughput is made.

Indeed, the occurrence of overflow is followed by a transition period during which K_j passes through the queues $Q_0, \ldots, Q_{M-1}$ - and during which therefore no customers depart from Q_{M-1}.
We now define

I-period: the period between the j-th overflow and the next departure of K_j from Q_{M-1};

B-period: the period between the departure of K_j from Q_{M-1} and the first, (j+1)-th, time thereafter at which overflow occurs.

The length of an I-period (B-period) will be denoted by the stochastic variable $\mathbf{i}$ ($\mathbf{b}$).
A regenerative cycle thus consists of an I-period and a subsequent B-period (see Fig. 5.3).

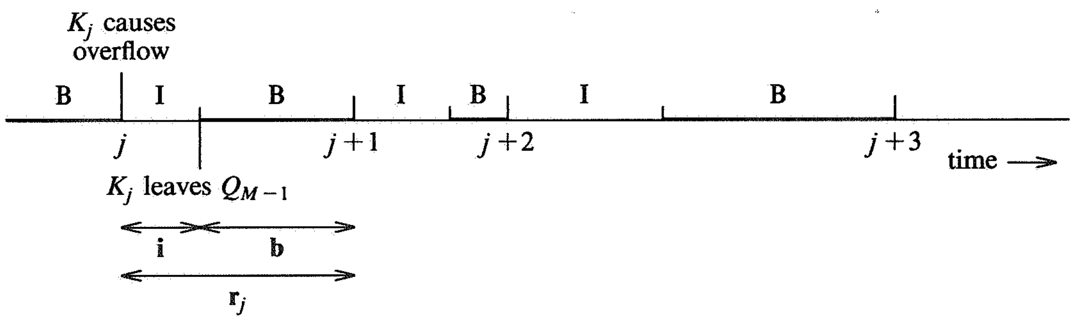


Fig. 5.3 Partition of the cycles in I-periods and B-periods.

It easily follows from the foregoing that

$$E\{\mathbf{r}_j\} = E\{\mathbf{i}\} + E\{\mathbf{b}\}\,, \tag{5.2}$$

$$Pr\{\mathbf{i} < t\} = A_0(t)^* \cdots {}^*A_{M-1}(t)\,. \tag{5.3}$$

Once K_j, at the beginning of a B-period, has departed from Q_{M-1}, other customers can also leave Q_{M-1} with varying mean times, until overflow takes place again.
Hence Observation 5.2 leads to

OBSERVATION 5.2A.
Only during B-periods contributions to the throughput are made.

If the departure process from Q_{M-1}, during a B-period, is a Poisson process, then Q_M can be considered as an M/M/1 queue during that time. Using the known results for that model, it is then possible to compute $E\{\mathbf{b}\}$. As, moreover, the service times at $Q_0, \ldots, Q_{M-1}$ are negative exponentially distributed, it is natural to approximate the departure process from Q_{M-1} during a B-period by a Poisson process with a suitable intensity $1/\alpha$. The thus obtained approximation $E\{\tilde{\mathbf{b}}\}$ for $E\{\mathbf{b}\}$ can also be used to estimate the number of customers who move from Q_{M-1} to Q_M during a B-period:

$$E\{\textit{number of counted customers}\} \approx E\{\tilde{\mathbf{b}}\}\,/\,\alpha\,. \tag{5.4}$$

Using (5.1) and (5.2) we now obtain the following approximation for the throughput:

$$T \approx \frac{1}{\alpha}\,\frac{E\{\tilde{\mathbf{b}}\}}{E\{\mathbf{i}\}+E\{\tilde{\mathbf{b}}\}}\,. \tag{5.5}$$

In the next section we shall deal with those details that have not yet been discussed. Using the theory of birth and death processes, we shall derive an approximation for $E\{\mathbf{b}\}$ and subsequently present ample motivation for the choice of the Poisson process and the parameter α.

5.3 DERIVATION OF THE APPROXIMATION FOR THE THROUGHPUT

In the foregoing section we have omitted the discussion of two matters which are of importance for a practical application of the approximation formula (5.5), viz., (i) the choice of a Poisson process (and the choice of its intensity) as an approximation of the arrival process at Q_M during a B-period and (ii) the derivation of the approximation $E\{\tilde{\mathbf{b}}\}$ for the mean duration of a B-period.
Next we shall derive an expression for $E\{\tilde{\mathbf{b}}\}$, based on the assumption that the arrival process at Q_M during a B-period is a Poisson process with (known) intensity $1/\alpha$. Furtheron in this section we shall present arguments to support this assumption, also discussing the choice of α.

Consider the system at the epoch at which customer K_j, at the start of the B-period, leaves Q_{M-1} and arrives at Q_M (see Fig. 5.3). The duration of this B-period clearly depends strongly on the number of customers $\mathbf{y}_M$ found by K_j in Q_M. For example, when $\mathbf{y}_M = L$, then obviously $\tilde{\mathbf{b}} = \mathbf{b} = 0$. Since the B-period ends at the next epoch at which a customer who leaves Q_{M-1} finds L customers present at Q_M, and since the arrival process during this period is assumed to be a Poisson process with intensity $1/\alpha$, the following holds:

$$\begin{aligned} E\{\tilde{\mathbf{b}}|\mathbf{y}_M{=}k\} &= E\{\alpha_{k+1,L+1}\}\,, \qquad k=0,\ldots,L-1\,,\\ &= 0\,, \qquad\qquad\qquad\quad\;\; k=L\,, \end{aligned} \tag{5.6}$$

where $\alpha_{i,j}$, $i,j \geqslant 0$, denotes the entrance time in state j starting from state i for the queue length process $\{\mathbf{x}_t,\ t \geqslant 0\}$ in an M/M/1 queue with arrival intensity $1/\alpha$ and mean service time α_M.
In (5.6) we have of course used the fact that customer K_j arrives at Q_M at the beginning of a B-period. Now, it easily follows that

$$E\{\tilde{\mathbf{b}}\} = \sum_{k=0}^{L-1} Pr\{\mathbf{y}_M{=}k\}E\{\boldsymbol{\alpha}_{k+1,L+1}\}\,. \tag{5.7}$$

Because there are exactly L customers present in Q_M at the beginning of an I-period (cf. Observation 5.1), it holds that $\mathbf{y}_M$ equals L minus the number of customers who departed from Q_M during the subsequent service times of K_j at $Q_0, \ldots, Q_{M-1}$ (cf. (5.3)).
Hence

$$Pr\{\mathbf{y}_M = k\} = \int_0^\infty \frac{(t/\alpha_M)^{L-k}}{(L-k)!} e^{-t/\alpha_M}\, dA_0(t) * \ldots * A_{M-1}(t)\,, \qquad k = 1,\ldots,L\,, \tag{5.8}$$

$$= 1 - \sum_{h=1}^{L} \int_0^\infty \frac{(t/\alpha_M)^{L-h}}{(L-h)!} e^{-t/\alpha_M}\, dA_0(t) * \ldots * A_{M-1}(t)\,, \qquad k = 0\,.$$

REMARK 5.1.
Using Laplace-Stieltjes transforms, the right-hand side of (5.8) can be computed as follows:

$$\int_0^\infty \frac{(t/\alpha_M)^{L-k}}{(L-k)!} e^{-t/\alpha_M}\, dA_0(t) * \ldots * A_{M-1}(t) = \tag{5.9}$$

$$\frac{(-1/\alpha_M)^{L-k}}{(L-k)!} \frac{d^{L-k}}{d\rho^{L-k}} \left[\frac{1}{1+\alpha_0 \rho} \cdots \frac{1}{1+\alpha_{M-1}\rho} \right]_{\rho = \frac{1}{\alpha_M}}, \quad \text{Re}\, \rho \geq 0\,, \quad k = 1,\ldots,L.$$

Because the queue length process of an M/M/1 system is a birth and death process with constant birth and death rates, $E\{\boldsymbol{\alpha}_{i,L+1}\}$ can be calculated by using known theory; e.g., it is possible to derive the L.S.T. $\phi_{i,L+1}(\rho)$ of the distribution function of $\boldsymbol{\alpha}_{i,L+1}$. It holds that

$$E\{\boldsymbol{\alpha}_{i,L+1}\} = \lim_{\rho \downarrow 0} \left[\frac{d}{d\rho} \phi_{i,L+1}(\rho) \right], \quad i = 0,1,\ldots,L\,. \tag{5.10}$$

A brief derivation and an expression for $\phi_{i,L+1}(\rho)$ is given in Appendix A (see also Cohen [6]).
For some practical calculations, carried out in the next section, we make use of the simple relations

$$E\{\boldsymbol{\alpha}_{i,j+1}\} = E\{\boldsymbol{\alpha}_{i,j}\} + E\{\boldsymbol{\alpha}_{j,j+1}\}\,, \quad i < j\,, \tag{5.11}$$

and

$$E\{\boldsymbol{\alpha}_{0,1}\} = \alpha\,. \tag{5.12}$$

Based on the assumption that there are j customers in the system, $E\{\boldsymbol{\alpha}_{j,j+1}\}$ can be calculated by just looking at the next event (the departure or the arrival of a customer). Denoting by $\boldsymbol{\tau}_M$ a service time at Q_M and by $\boldsymbol{\tau}$ a negative exponentially distributed random variable with $E\{\boldsymbol{\tau}\} = \alpha$, it holds that

$$E\{\boldsymbol{\alpha}_{j,j+1}\} = E\{\boldsymbol{\tau}(\boldsymbol{\tau}_M \geq \boldsymbol{\tau})\} + E\{\boldsymbol{\tau}_M(\boldsymbol{\tau}_M < \boldsymbol{\tau})\} + E\{\boldsymbol{\alpha}_{j-1,j+1}(\boldsymbol{\tau}_M < \boldsymbol{\tau})\}\,, \; j = 1,2,\ldots\,. \tag{5.13}$$

From (5.11), because of the mutual independency of $\boldsymbol{\alpha}_{j-1,j+1}$, $\boldsymbol{\tau}_M$ and $\boldsymbol{\tau}$ in the right hand side of (5.13), it follows that

$$E\{\boldsymbol{\alpha}_{j,j+1}\} = E\{\boldsymbol{\tau}(\boldsymbol{\tau}_M \geq \boldsymbol{\tau})\} + E\{\boldsymbol{\tau}_M(\boldsymbol{\tau}_M < \boldsymbol{\tau})\} + Pr\{\boldsymbol{\tau}_M < \boldsymbol{\tau}\}[E\{\boldsymbol{\alpha}_{j-1,j}\} + E\{\boldsymbol{\alpha}_{j,j+1}\}]\,.$$

From this $E\{\alpha_{j,j+1}\}$ can easily be solved. Using

$$E\{\boldsymbol{\tau}(\boldsymbol{\tau}_M \geq \boldsymbol{\tau})\} + E\{\boldsymbol{\tau}(\boldsymbol{\tau}_M < \boldsymbol{\tau})\} = \frac{\alpha_M}{1+a},$$

where $a = \alpha_M / \alpha$, we find

$$E\{\boldsymbol{\alpha}_{j,j+1}\} = \frac{\alpha_M + E\{\boldsymbol{\alpha}_{j-1,j}\}}{a}. \tag{5.14}$$

Hence we can now calculate $E\{\alpha_{i,j}\}$ using (5.11), (5.12) and (5.14).

In the preceding derivation we have used the assumption that the arrival process at Q_M during a B-period is a Poisson process. We shall now support this (approximation) assumption, and the choice of α, with some (heuristic) arguments.

Because no overflow occurs during a B-period, it is natural to view the system during such a period as an "ordinary" closed cyclic system (with infinite waiting rooms); cf. Fig. 5.4.

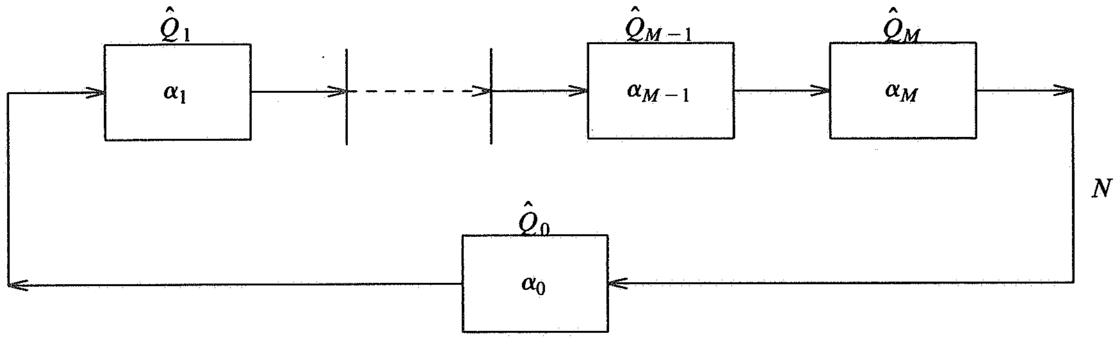


Fig. 5.4 Closed cyclic model with infinite waiting room

As far as the arrival process at Q_M is concerned, we can make use of some known results for that model. Those results have been derived under the condition of ergodicity. This is in general not a serious problem, as each B-period is preceded by a transient (I-) period. We distinguish two cases:

1. There is a unique queue with largest mean service time
2. There are several ($l > 1$) queues with the same, largest, mean service time.

For case 1, Boxma [2] shows that the arrival process at $\hat{Q}_i$, $i=0,1,...,M$ (see Fig. 5.4) can be approximated by a Poisson process with intensity $1/\alpha$ (if $N \to \infty$), where

$$\alpha = \max\{\alpha_i,\ i=0,1,...,M\}. \tag{5.15}$$

For case 2, too, it is shown that the arrival process at $\hat{Q}_i$, $i=0,1,...,M$ can be approximated by a Poisson process with intensity $1/\alpha$ (if $N \to \infty$), but now

$$\alpha = \frac{N+l-1}{N} \max\{\alpha_i,\ i=0,1,...,M\}. \tag{5.16}$$

These approximations are very accurate for large values of N, but they even yield acceptable results for rather small values of N.

At first sight the obvious thing to do now seems to approximate the arrival process at Q_M during a B-period by a Poisson process with intensity $1/\alpha$ as given by (5.15) and (5.16). However, in the foregoing we have not taken into account that there can be at most L customers present in Q_M. Suppose, e.g., that the mean service time in Q_M is by far the largest of all mean service times. In the model with all waiting rooms infinite, we have the well-known - and obvious - result that $\hat{Q}_M$ almost never is empty ($Pr\{\hat{Q}_M \text{ is empty}\} \approx 0$). This explains the above-mentioned fact that the arrival process at all queues of the cyclic model can be accurately approximated by a Poisson process with intensity (in this case) $1/\alpha_M$ (see Boxma [2]). But this argument breaks down when the waiting room at Q_M is finite. E.g., in the case $L=1$ it can be easily shown that the probability that Q_M is empty is at least $\alpha_{M-1}/(\alpha_M+\alpha_{M-1})$.

The above reasoning, and the observation that there are always *at least* $N-L$ customers present in $Q_0, \ldots, Q_{M-1}$, have led us to the decision to leave α_M out of consideration for the determination of the arrival intensity at Q_M. Hence we choose, in case 1 (cf. (5.15)):

$$\alpha := \max\{\alpha_i,\ i=0,1,\ldots,M-1\}. \tag{5.17}$$

Case 2 presents some additional difficulties, because the total number of customers present in $Q_0, \ldots, Q_{M-1}$ during a B-period is not constant (varying between $N-L$ and N). We have chosen

$$\alpha := \frac{\tilde{N}+l-1}{\tilde{N}} \max\{\alpha_i,\ i=0,1,\ldots,M-1\}, \tag{5.18}$$

with $\tilde{N}$ the mean total number of customers present in $Q_0, \ldots, Q_{M-1}$ at the start of a B-period (for the determination of l, α_M is again left out of consideration, cf. Section 5.4 , Example 2). We shall present numerical evidence to support the choice made in (5.18).
Formula (5.8) can be used for the calculation of $\tilde{N}$, for

$$\tilde{N} = N - E\{y_M\}. \tag{5.19}$$

We can now give an approximation of the throughput, using (5.5). In the next section we shall demonstrate the use of the approximation method in two examples.

5.4 SOME EXAMPLES

For two concrete cases we shall carry out the approximation method for the throughput, which was described in the preceding sections. In the next section more numerical results are presented and compared with results obtained by simulation.

We first recall the approximation formula (5.5):

$$T = \frac{1}{\alpha} \frac{E\{\tilde{\mathbf{b}}\}}{E\{\mathbf{i}\}+E\{\tilde{\mathbf{b}}\}}.$$

EXAMPLE 1: $M=3$, $N=6$, $L=1$, $\alpha_0=5$, $\alpha_1=\alpha_2=\alpha_3=1$.

From (5.3) it easily follows that

$$E\{\mathbf{i}\} = \alpha_0+\alpha_1+\alpha_2 = 7 .$$

Because $\alpha_0>\alpha_1=\alpha_2$, we choose

$$\alpha := \alpha_0 = 5 .$$

Hence

$$a := \frac{\alpha_3}{\alpha} = \frac{1}{5} .$$

$E\{\tilde{\mathbf{b}}\}$ is computed by using (5.7). For that purpose we first have to calculate $Pr\{\mathbf{y}_3=0\}$, $Pr\{\mathbf{y}_3=1\}$ and $E\{\boldsymbol{\alpha}_{1,2}\}$.
From (5.8) and (5.9) it follows that

$$Pr\{\mathbf{y}_3=1\} = \frac{1}{1+\frac{\alpha_0}{\alpha_3}} \frac{1}{1+\frac{\alpha_1}{\alpha_3}} \frac{1}{1+\frac{\alpha_2}{\alpha_3}} = \frac{1}{24},$$

$$Pr\{y_3=0\} = 1-\frac{1}{1+\frac{\alpha_0}{\alpha_3}}\frac{1}{1+\frac{\alpha_1}{\alpha_3}}\frac{1}{1+\frac{\alpha_2}{\alpha_3}} = \frac{23}{24} .$$

Using (5.12) and (5.14) we find

$$E\{\boldsymbol{\alpha}_{1,2}\} = \frac{\alpha_3+\alpha}{a} = 30 . \tag{5.20}$$

Hence

$$E\{\tilde{\mathbf{b}}\} = Pr\{y_3=0\}E\{\boldsymbol{\alpha}_{1,2}\} = \frac{23}{24}\times 30 = \frac{115}{4} .$$

We now have the following approximation for the throughput:

$$T \approx \frac{1}{5}\frac{115/4}{7+115/4} = \frac{23}{143} \approx 0.161 .$$

EXAMPLE 2: $M=3$, $N=6$, $L=2$, $\alpha_0=\alpha_1=\alpha_2=\alpha_3=1$.

Obviously $E\{\mathbf{i}\} = 3$.
Because $\alpha_0=\alpha_1=\alpha_2$, it is now more difficult to determine α. For that purpose we first have to calculate $E\{y_3\}$ (cf. (5.18) and (5.19)).
From (5.8) and (5.9):

$$Pr\{y_3=2\} = \frac{1}{1+\frac{\alpha_0}{\alpha_3}}\frac{1}{1+\frac{\alpha_1}{\alpha_3}}\frac{1}{1+\frac{\alpha_2}{\alpha_3}} = \frac{1}{8} ,$$

$$Pr\{y_3=1\} = \frac{1}{\alpha_3}\left[\alpha_0\left(\frac{1}{1+\frac{\alpha_0}{\alpha_3}}\right)^2\frac{1}{1+\frac{\alpha_1}{\alpha_3}}\frac{1}{1+\frac{\alpha_2}{\alpha_3}}+\alpha_1\frac{1}{1+\frac{\alpha_0}{\alpha_3}}\left(\frac{1}{1+\frac{\alpha_1}{\alpha_3}}\right)^2\frac{1}{1+\frac{\alpha_2}{\alpha_3}}\right.$$
$$\left.+\alpha_2\frac{1}{1+\frac{\alpha_0}{\alpha_3}}\frac{1}{1+\frac{\alpha_1}{\alpha_3}}\left(\frac{1}{1+\frac{\alpha_2}{\alpha_3}}\right)^2\right] = \frac{3}{16} ,$$

$$Pr\{y_3=0\} = 1-(Pr\{y_3=1\}+Pr\{y_3=2\}) = \frac{11}{16} .$$

So

$$E\{y_3\} = \frac{7}{16} .$$

From (5.19) it now follows that

$$\tilde{N} = 6-\frac{7}{16} = 5\frac{9}{16} .$$

Consequently, we choose (cf. (5.18))

$$\alpha = \frac{5\frac{9}{16}+3-1}{5\frac{9}{16}}\times 1 = \frac{121}{89} .$$

Hence

$$a := \frac{\alpha_3}{\alpha} = \frac{89}{121}\,.$$

Using (5.14) and (5.20) we find

$$E\{\boldsymbol{\alpha}_{2,3}\} = \frac{\alpha_3 + E\{\boldsymbol{\alpha}_{1,2}\}}{a} = \frac{121}{89} + \left[\frac{121}{89}\right]^2 + \left[\frac{121}{89}\right]^2 \approx 5.721\,,$$

$$E\{\boldsymbol{\alpha}_{1,3}\} = E\{\boldsymbol{\alpha}_{1,2}\} + E\{\boldsymbol{\alpha}_{2,3}\} \approx 8.929\,.$$

Now, from (5.7) it follows that

$$E\{\tilde{\mathbf{b}}\} = \frac{11}{16}E\{\boldsymbol{\alpha}_{1,3}\} + \frac{3}{16}E\{\boldsymbol{\alpha}_{2,3}\} \approx 7.211\,.$$

Finally

$$T \approx \frac{89}{121}\frac{7.211}{3+7.211} \approx 0.519\,.$$

It is clear that the complexity of the calculations increases if L or N becomes larger. In principle the distribution of $\mathbf{y}_M$ can be determined explicitly (cf. (5.9)), but for large values of L and M we have to take refuge to a numerical determination by the computer.

5.5 Numerical results for the throughput

In this section we compare approximation results with results obtained by simulation, and we discuss the influence of the total number of customers in the system (N), i.e. the window size, on the throughput. Subsequently, we show that for the case $M=1$ the approximation is exact.

In the same way as shown in the foregoing section, we have, for a number of cases, computed the approximation for the throughput given by (5.5). The approximations are based on a model consisting of four queues ($M=3$), with in all cases $\alpha_3=1$. The cases are chosen in such a way, that we get a representative view of the reliability of the approximation results for different combinations of the parameters N, α_0, α_1, α_2. In all cases we have chosen $L=1$ or $L=2$. This should suffice because the overflow intensity can be varied by different combinations of α_0, α_1, α_2.

The simulation results have been obtained by the IBM RESQ simulation packet, and by a SIMULA'67 program (see [23]). An extensive description of this program is given in Appendix B. In this appendix we also discuss the simulation method which we have used and the reliability of the results.

From Table 5.1, on page 37, it appears that the approximation yields (very) good results. The relative approximation errors indicated in the table are defined as:

$$100\% \times \frac{\textit{approximation result} \;-\; \textit{simulation result}}{\textit{simulation result}}\,.$$

We shall now discuss the results.

It appears that for the cases with a unique queue with largest mean service time, the results are, generally speaking, better than the results for the cases with more than one queue with largest mean service time. This was to be expected, because in the last mentioned cases an additional approximation is needed for determining the parameter α (cf. (5.18)).

For small values of N (properly $\tilde{N}$), the approximation results are slightly less accurate. Apparently this is related to the fact that in these cases the Poisson process is not such a good approximation for the arrival process at Q_M during a B-period (see page 33).

It also appears from Table 5.1 that the influence of N on the throughput is very small. Moreover, the approximation results are independent of N for the cases with a unique queue with largest mean service time. When two or more queues have largest mean service time then the approximation does depend on N (cf. (5.18) and Example 2, Section 5.4), but it is to be expected that the influence of N on the throughput rapidly becomes negligible, if N grows (see (5.18) and (5.19)).

N	L	α_0	α_1	α_2	$T_{sim.}$	$T_{approx.}$	% error	95% conf. interval
3	2	1	1	1	0.475	0.454	-4.4	0.471-0.479
3	2	5	1	1	0.189	0.192	+1.6	0.183-0.192
4	2	1	1	1	0.502	0.488	-2.8	0.484-0.520
4	2	0.2	1	1	0.556	0.557	+0.2	0.538-0.575
4	2	5	1	1	0.192	0.192	-0.0	0.188-0.196
6	2	1	1	1	0.505	0.519	+2.7	0.470-0.540
6	2	5	1	1	0.193	0.192	-0.5	0.186-0.199
3	1	0.2	1	1	0.400	0.386	-3.3	0.388-0.412
3	1	5	1	1	0.159	0.161	+1.3	0.156-0.162
6	1	0.2	1	1	0.402	0.407	+1.2	0.392-0.412
6	1	1	1	1	0.347	0.356	+2.6	0.336-0.357
6	1	5	1	1	0.160	0.161	+0.6	0.156-0.164
6	1	0.2	0.2	0.2	0.716	0.719	+0.4	0.706-0.726
6	1	1	0.2	0.2	0.483	0.483	-0.0	0.476-0.490
6	1	0.2	0.2	5	0.166	0.166	+0.0	0.158-0.173
6	1	1	0.2	5	0.165	0.164	-0.6	0.158-0.171

Table 5.1. Simulation- and approximation results for the throughput (T) in the model with $M=3$. In all cases $\alpha_3 = 1$.

In Table 5.2a and Table 5.2b the approximation is carried out for different values of N and compared with the simulation. The other parameters are kept fixed. It is seen that even for small values of N, the throughput hardly varies. This is consistent with the results obtained in the preceding chapter, where it was found that, if $M=1$, the throughput is totally independent of N (cf. (4.10)).

N	$T_{sim.}$	$T_{approx.}$	% error
3	0.189	0.192	+1.6
4	0.192	0.192	-0.0
6	0.193	0.192	-0.5
10	0.195	0.192	-0.5
15	0.195	0.192	-1.5
20	0.195	0.192	-1.5

a

N	$T_{sim.}$	$T_{approx.}$	% error
2	0.394	0.356	-9.6
3	0.400	0.386	-3.3
4	0.402	0.401	-0.2
6	0.402	0.407	+1.2
10	0.403	0.412	+2.2
15	0.402	0.414	+2.8

b

Table 5.2 The influence of N on T.
a. $M=3,\ L=2,\ \alpha_0=5,\ \alpha_1=\alpha_2=\alpha_3=1$.
b. $M=3,\ L=1,\ \alpha_0=0.2,\ \alpha_1=\alpha_2=\alpha_3=1$.

It appears that for the case $M=1$, the approximation results are exact. For in that case the arrival process at Q_1 *is* a Poisson process with intensity $1/\alpha_0$ (see Section 4.2). Hence, if $M=1$, it holds that

$$E\{\mathbf{b}\} = E\{\tilde{\mathbf{b}}\},$$

and

$$\frac{E\{\mathbf{b}\}}{\alpha_0} = E\{\textit{number of customers moving from } Q_0 \textit{ to } Q_1 \textit{ during a } B-\textit{period}\}.$$

From (5.1) it follows that

$$T = \frac{E\{\tilde{\mathbf{b}}\}/\alpha_0}{E\{\tilde{\mathbf{b}}\}+E\{\mathbf{i}\}}, \quad \text{if } M=1, \tag{5.21}$$

which is the approximation formula (5.5).

REMARK 5.2 .
For the case $M=1$, $L=1$, (5.21) can be proved explicitly. In that case (see Example 1, Section 5.4)

$$E\{\mathbf{b}\} = E\{\tilde{\mathbf{b}}\} = \left[1-\frac{1}{1+\frac{\alpha_0}{\alpha_1}}\right]\frac{\alpha_0+\alpha_1}{\alpha_1/\alpha_0},$$

$$E\{\mathbf{i}\} = \alpha_0 .$$

Hence

$$\frac{E\{\tilde{\mathbf{b}}\}/\alpha_0}{E\{\tilde{\mathbf{b}}\}+E\{\mathbf{i}\}} = \frac{1}{\alpha_0+\alpha_1}.$$

Using the exact formula (4.7), it is easily verified that $T=\frac{1}{(\alpha_0+\alpha_1)}$.

5.6 The mean end-to-end delay

We already remarked in Section 5.1 that the approximation of the end-to-end delay gives rise to problems, because of its complicated structure. In this section it is our intention to make a start with an approximation approach for the end-to-end delay. Hereto we make use of some rough estimates and of the approximations for T and $E\{\mathbf{b}\}$, which were found in the foregoing sections. Therefore the approximation results may not be expected to be very accurate. However, the method leads to some interesting conclusions, and it illustrates the possibilities offered by the approximation method for the throughput to estimate other performance measures (see also Section 5.7). Comparison of a number of approximation results with simulation results leads us to believe that the set-up is worthy of further study.

From the definition of the end-to-end delay (see Section 3.2), it follows that we can distinguish two kinds of customers in the system; viz., customers who already started a (new) end-to-end delay and customers who did not yet start a new end-to-end delay. In the sequel we shall indicate these two kinds of customers by, respectively, type 1 and type 2 customers. It is clear that type 2 customers can only be in Q_0. However, Q_0 may also contain type 1 customers who have been sent back because overflow occurred. Of course, the type 1 customers are always ahead of the type 2 customers. We now define:

$\mathbf{x}_0^{(i)}$:= number of type i customers present in Q_0, $i=1, 2$.

$\mathbf{s}_0^{(i)}$:= response time at Q_0 of a type i customer, $i = 1, 2$.

From the definition of the end-to-end delay it follows that

$$E\{\mathbf{d}\} = E\{\mathbf{c}\} - E\{\mathbf{s}_0^{(2)}\}. \tag{5.22}$$

The mean cycle time $E\{\mathbf{c}\}$ can be estimated by applying Little's formula to the approximation result for the throughput:

$$E\{\mathbf{c}\} = \frac{N}{T}. \tag{5.23}$$

It is clear that the mean number of type 2 customers arriving at Q_0 per unit of time is equal to T. Hence, applying Little's formula once again, we obtain

$$E\{\mathbf{s}_0^{(2)}\} = \frac{E\{\mathbf{x}_0^{(2)}\}}{T}. \tag{5.24}$$

From (5.22), (5.23) and (5.24) it follows that, if we are able to determine $E\{\mathbf{x}_0^{(2)}\}$, we can approximate $E\{\mathbf{d}\}$ (using the approximation for T).

We shall now try to find an approximation for $E\{\mathbf{x}_0^{(2)}\}$. For that purpose we observe the system at an overflow epoch, and define

$\mathbf{n}$:= number of customers that is sent back from $Q_1, \ldots, Q_{M-1}$ to Q_0 (because of overflow).

Hence, there are at least $\mathbf{n}$ type 1 customers present in Q_0 at the start of an I-period. Now, during one regeneration cycle, we shall observe the different types of customers in the system. Because of the order of the two types of customers in the system, as mentioned before, it is easily seen that

OBSERVATION 5.3.
As soon as all the type 1 customers, who were present in Q_0 at the start of the I-period, have left Q_0, there are only type 2 customers (or no customers at all) in Q_0 until overflow occurs again.

Hence, it seems to be much easier to determine the mean number of type 1 customers in Q_0 than the mean number of type 2 customers in Q_0. For suppose there are n type 1 customers present in Q_0 at the start of an I-period, then $E\{\mathbf{x}_0^{(1)}\}$ can be approximated as follows

$$E\{\mathbf{x}_0^{(1)}\} \approx \frac{n\alpha_0 + (n-1)\alpha_0 + \ldots + \alpha_0}{E\{\tilde{\mathbf{b}}\} + E\{\mathbf{i}\}}. \tag{5.25}$$

If, in addition, we have an approximation for $E\{\mathbf{x}_0\}$, then, using

$$E\{\mathbf{x}_0\} = E\{\mathbf{x}_0^{(1)}\} + E\{\mathbf{x}_0^{(2)}\}, \tag{5.26}$$

we are able to approximate $E\{\mathbf{x}_0^{(2)}\}$.

First, we try to approximate $E\{\mathbf{x}_0^{(1)}\}$. Hereto we propose to approximate the mean number of type 1 customers present in Q_0 at the start of an I-period by $E\{\mathbf{n}\}$.
For the approximation of $E\{\mathbf{n}\}$ we consider the system, at the end of a B-period, as a closed cyclic model with infinite waiting rooms; see Fig 5.4. Using the notation of this figure, we propose the following approximation $E\{\tilde{\mathbf{n}}\}$ for $E\{\mathbf{n}\}$:

$$E\{\tilde{\mathbf{n}}\} = 1 + E\{\textit{total number of customers present in } \hat{Q}_1, \ldots, \hat{Q}_{M-1} \mid \textit{there are } L+1 \textit{ customers present in } \hat{Q}_M\}. \tag{5.27}$$

The right-hand side of (5.27) can easily be calculated exactly (see Baskett e.a. [1]). Now, analogous to (5.25), $E\{x_0^{(1)}\}$ is estimated by

$$E\{x_0^{(1)}\} \approx \frac{\lfloor E\{\mathbf{n}\}\rfloor \alpha_0 + (\lfloor E\{\mathbf{n}\}\rfloor - 1)\alpha_0 + ... + (E\{\mathbf{n}\} - \lfloor E\{\mathbf{n}\}\rfloor)\alpha_0}{E\{\tilde{\mathbf{b}}\} + E\{\mathbf{i}\}} , \tag{5.28}$$

where $\lfloor E\{\mathbf{n}\}\rfloor$ equals the largest integer which is smaller than $E\{\mathbf{n}\}$.

For the approximation of $E\{x_0\}$ we first define

$E\{x_0 \,|_I\}$:= mean number of customers present in Q_0 during an I-period,

$E\{x_0 \,|_B\}$:= mean number of customers present in Q_0 during a B-period.

With the help of good estimates $E\{\tilde{x}_0 \,|_I\}$ and $E\{\tilde{x}_0 \,|_B\}$ for $E\{x_0 \,|_I\}$ and $E\{x_0 \,|_B\}$, we can approximate $E\{x_0\}$ as follows:

$$E\{x_0\} \approx \frac{E\{\mathbf{i}\}E\{\tilde{x}_0 \,|_I\} + E\{\tilde{\mathbf{b}}\}E\{\tilde{x}_0 \,|_B\}}{E\{\mathbf{i}\} + E\{\tilde{\mathbf{b}}\}} . \tag{5.29}$$

To estimate $E\{x_0 \,|_I\}$ we start with $N-L$ customers present in Q_0 and L customers present in Q_M at the beginning of an I-period. During a time $E\{\mathbf{i}\}$ we consider the system as if the service times at the queues are constant. Now it is easy to obtain an approximation $E\{\tilde{x}_0 \,|_I\}$.

To estimate $E\{x_0 \,|_B\}$, again, cf.(5.27), we consider the system during a B-period as a closed cyclic model with infinite waiting rooms (see Fig. 5.4). Now it is natural to choose

$$E\{\tilde{x}_0 \,|_B\} = E\{\textit{number of customers present in } \hat{Q}_0 | \textit{there are less than } L+1 \textit{ customers present in } \hat{Q}_M\} . \tag{5.30}$$

The right-hand side of (5.30) can exactly be determined (cf. (5.27)).

Now we are able to approximate $E\{x_0\}$ and $E\{x_0^{(1)}\}$ (see (5.28)). Then, using (5.26), $E\{x_0^{(2)}\}$ can be estimated. Subsequently, using (5.24), $E\{s_0^{(2)}\}$ can be estimated and, finally, using (5.22), we find an approximation for the mean end-to-end delay.

Because of the rough estimates which are used for the approximation, the results may not be expected to be very accurate. For a small number of cases we carried out the approximation. In Table 5.3 on page 41, the results are presented and compared with simulation results. The simulation was carried out by the program described in Appendix B.

We shall not discuss the results. We only remark that, because of the rather small errors, it might be worthwhile to continue the investigations at a later stage (see Section 5.7).

N	α_0	α_1	α_2	$E\{\mathbf{d}\}_{sim.}$	$E\{\mathbf{d}\}_{approx.}$	%error	95% conf. interval
4	0.2	1	1	9.65	9.71	+0.6	9.49-9.81
6	1	1	1	11.50	10.90	-5.2	11.08-11.92
6	0.2	1	1	14.64	12.87	-12.1	14.43-14.75
6	5	1	1	5.76	5.78	+0.3	5.32-6.20
6	0.2	0.2	0.2	6.88	6.56	-4.6	6.73-7.03
6	1	0.2	0.2	3.50	3.73	+6.6	3.38-3.62
6	1	0.2	5	33.55	34.22	+2.0	31.59-35.51
6	0.2	5	0.2	36.08	35.78	-0.8	35.48-36.68
6	1	5	5	45.05	41.23	-8.5	43.64-46.46

Table 5.3. Simulation - and approximation results for the mean end-to-end delay. In all cases $M=3$, $L=1$ and $\alpha_3=0$.

In Table 5.3, the approximation was carried out only for rather small values of N. We expect that the results become worse if N grows. This can be seen as follows.
Suppose that $(N-L)/\alpha_0 >> E\{\mathbf{i}\}+E\{\mathbf{b}\}$. This means that, usually, at the end of a B-period there are still many customers present in Q_0, who already were in Q_0 at the start of the foregoing I-period. Thus, in that case, a cycle time $(E\{\mathbf{i}\}+E\{\mathbf{b}\})$ is too "short" to spread the customers over the queues in such a manner that comparison of the system with a closed cyclic model with infinite waiting rooms, which is in equilibrium, is possible. However, the approximation is (partly) based on the assumption that these two systems are comparable (cf. (5.27) and (5.30)).

The discussion above leads to an interesting assumption. Because both the mean cycle time and the throughput are (almost) independent of N (cf. (5.23) and Section 5.5), we expect that $E\{\mathbf{n}\}$ is bounded (independent of N), and that

$$E\{\mathbf{n}\} < \frac{E\{\mathbf{i}\}+E\{\mathbf{b}\}}{\alpha_0} . \tag{5.31}$$

For all the customers who are sent back, have departed from Q_0 during the cycle time. Hence, because the returning customers are placed at the head of the queue, the mean response time at Q_0 of these type 1 customers will be bounded too if $N \uparrow \infty$. Now, we expect that the same holds for the mean end-to-end delay.
Simulation results suggest that this assumption is right. By varying N and keeping the other parameters fixed, we studied (using simulation) the influence of N on the mean end-to-end delay, see Fig 5.5, pp. 43-44. It is clear that in all three cases ((a), (b) and (c)) the mean end-to-end delay converges to an upper bound, if $N \uparrow \infty$. However, the speed of convergence varies strongly. In case (a) the upper bound is already almost reached if $N=6$, whereas in case (c) the mean end-to-end delay does not stabilize until $N=300$. In view of the foregoing this might be expected (cf. (5.31)); for in case (a) $(E\{\mathbf{i}\}+E\{\mathbf{b}\})/\alpha_0 = 2.7$, and in case (c) $(E\{\mathbf{i}\}+E\{\mathbf{b}\})/\alpha_0 = 39.7$. In case (b) $(E\{\mathbf{i}\}+E\{\mathbf{b}\})/\alpha_0 = 5.9$.

5.7 Conclusions

In this section we first indicate some points which will not be discussed in this report, but which are worthy of further study. After that we shall sum up the results obtained in this chapter.

Firstly we suggest a further investigation of the approximation method which was discussed in the previous section. In particular, it is interesting to investigate the maximum mean end-to-end delay

($\lim_{N\uparrow\infty} E\{\mathbf{d}\}$) and the speed of the convergence (from Fig. 5.5 it appears that, initially, $E\{\mathbf{d}\}$ grows linearly).

Two performance measures which have not been mentioned until now are the overflow intensity (O), and the probability that a customer, departing from Q_{M-1}, causes overflow ($Pr\{overflow\}$). Especially the overflow intensity may be an important measure, for if many customers have to be sent back very often, then this may give a substantial contribution to the total load of the system.
It is natural to approximate these performance measures as follows:

$$O \approx \frac{1}{E\{\mathbf{i}\}+E\{\tilde{\mathbf{b}}\}},$$

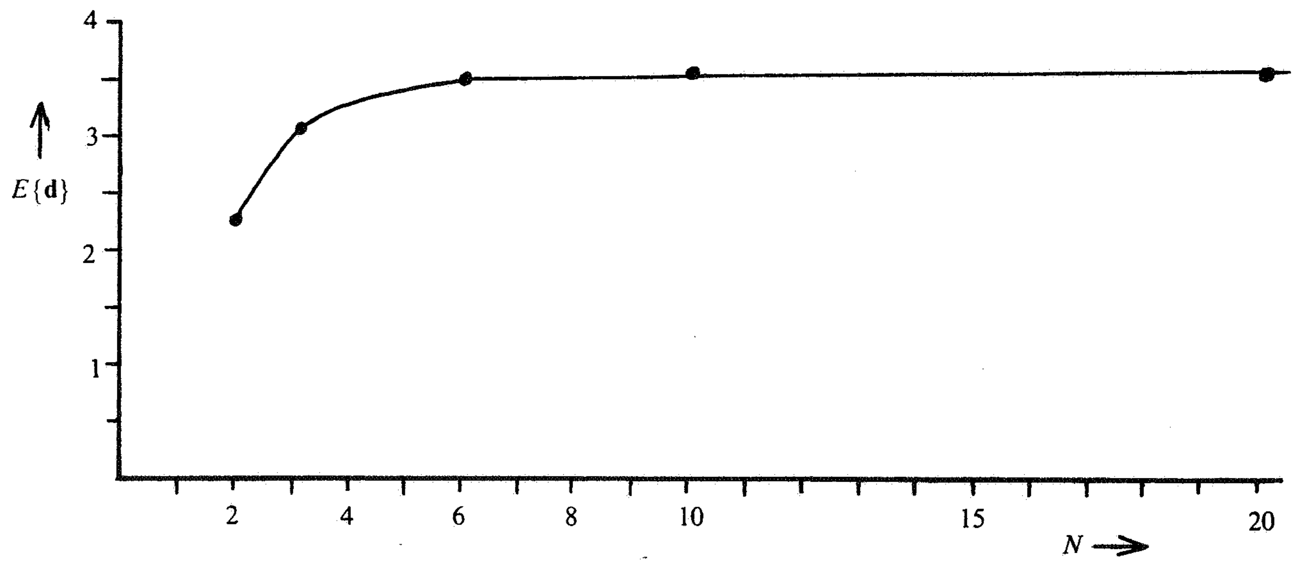
and

$$Pr\{overflow\} \approx \frac{1}{1+E\{\tilde{\mathbf{b}}\}/\alpha}.$$

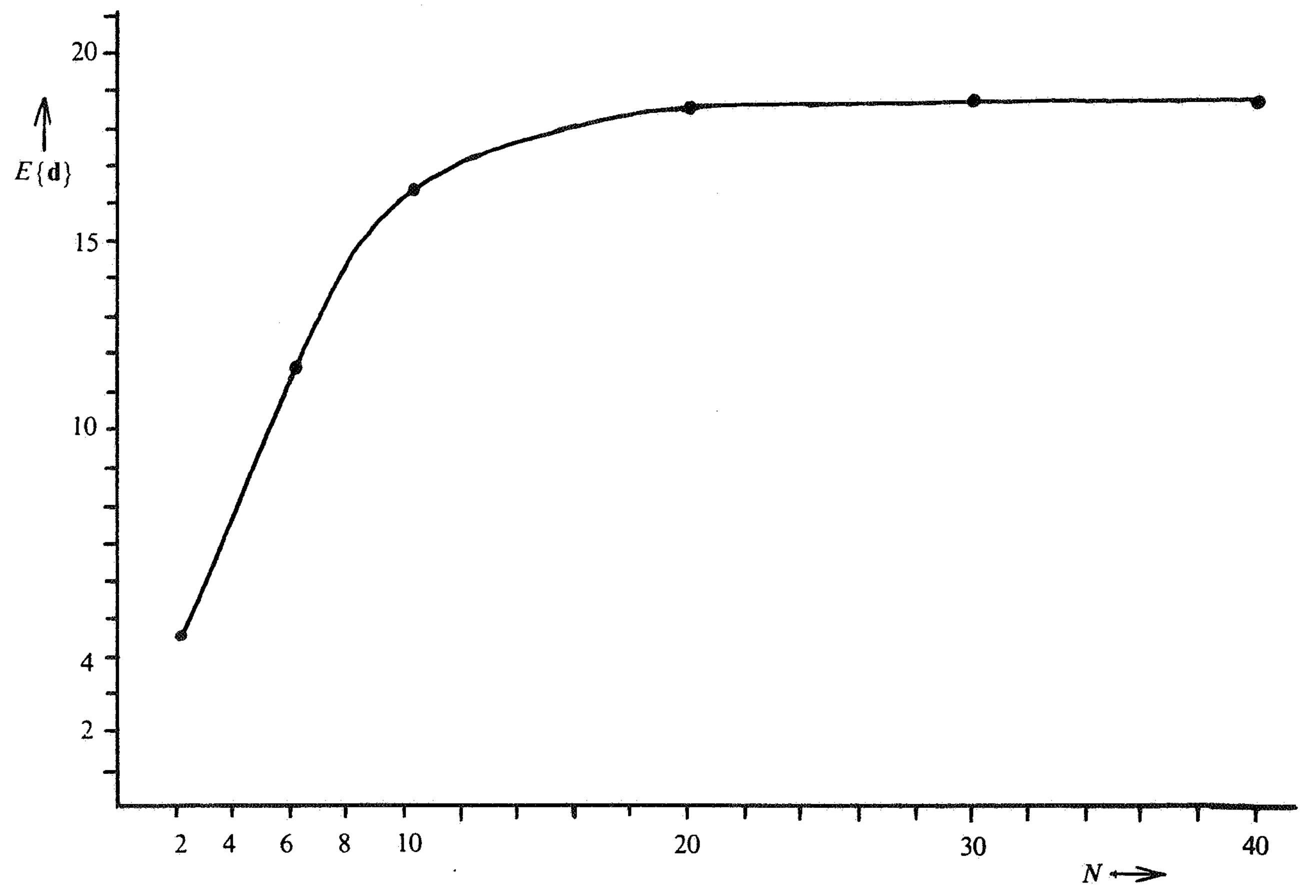
Thus, we can make use of the known values of $E\{\mathbf{b}\}$ and α, which were already derived for the approximation of the throughput.
With the help of the simulation program, the approximations can be tested.
A first start of the investigations shows good results for the approximation of $Pr\{overflow\}$. The accuracy of the approximation is comparable with the accuracy of the approximation of T, as, in view of the approximation formula, might be expected.

Finally, we mention a (realistic) variant of the model, in which the customers, in case of overflow, are *not* sent back to Q_0 (the sender, see Chapter 3), but immediately return to Q_1 (the first node of the virtual circuit). The approximation method (for T) for the original model should be easily adaptable to this variant. It is easily seen that for this variant the mean end-to-end delay is *not* bounded, if $N\uparrow\infty$; for if the customers, in case of overflow, return to Q_1, then the flow of type 2 customers from Q_0 to Q_1 is *not* blocked, contrary to the situation in the original model.
Adaptation of the approximation method to variants consisting of more than one queue with finite waiting room seems to be not quite as easy.

Now we shall briefly review the results obtained in this chapter.
- We developed a rather simple approximation method for the throughput, which is exact for the special case, treated in Chapter 4, in which the model consists of only two queues.
- For the (tested) case that the model consists of four queues ($M=3$) and Q_3 has a waiting room of size one or two ($L=1,2$), the relative approximation error is about a few percent. Only for small values of $\tilde{N}$ (see (5.19) and Section 5.5) the approximation results are less accurate. Because of the general formulation of the method, we may expect that the approximation is also useful for larger values of L and M.
- It appears that the throughput only slightly depends on N, and it becomes constant if $N\uparrow\infty$.
- The construction of the approximation method provides much insight in other performance measures, for instance the mean end-to-end delay.
- We have set up an approximation method for the mean end-to-end delay. The obtained results, for small values of N, are encouraging.
- Simulation results support the conjecture, arisen during the development of the approximation, that not only the throughput but also the mean end-to-end delay approaches a finite limit if $N\uparrow\infty$.

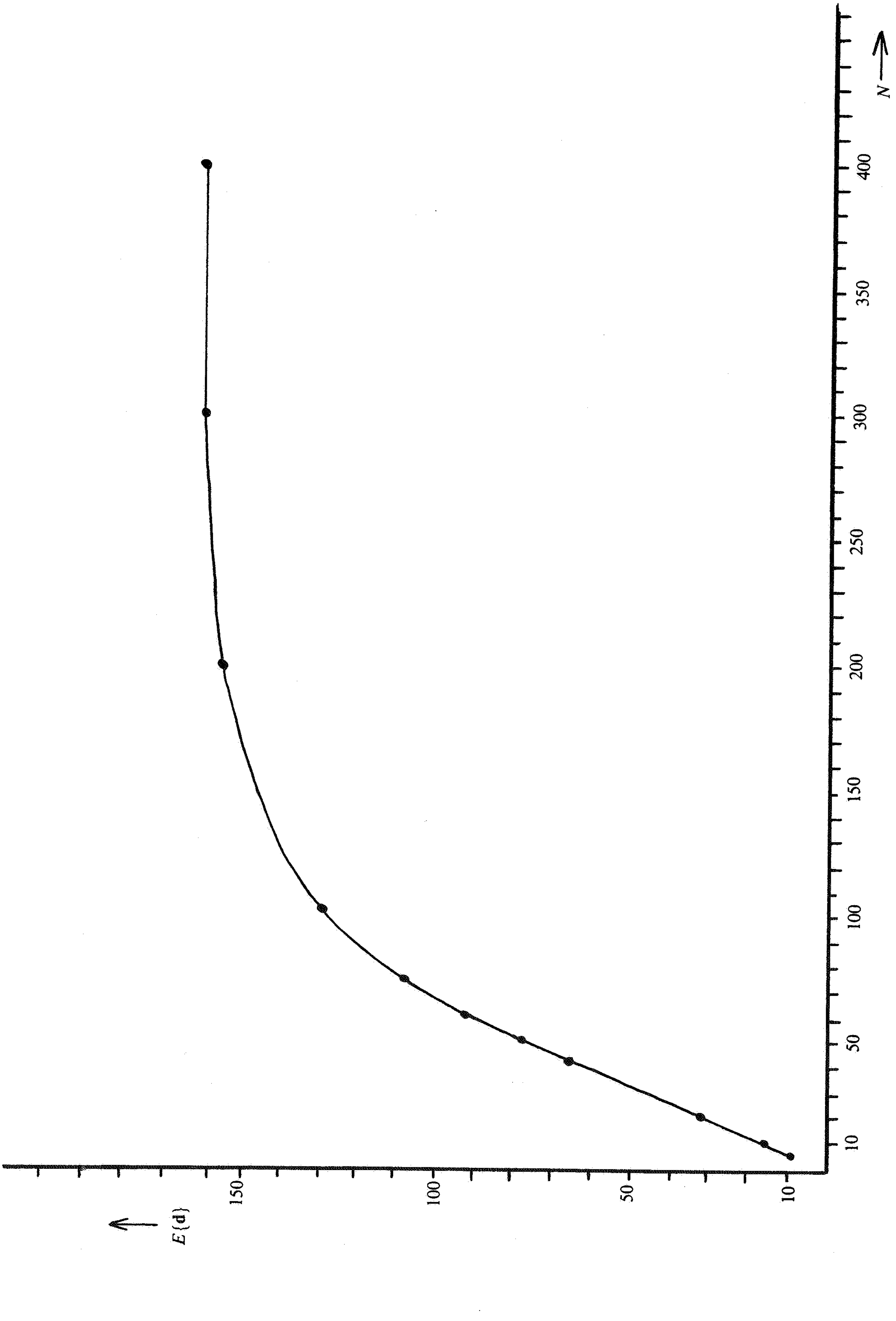


5.5a



5.5b

Fig. 5.5 The influence of N on the mean end-to-end delay ($E\{\mathbf{d}\}$).
a. $M=3,\ L=1,\ \alpha_0=1,\ \alpha_1=\alpha_2=0.2,\ \alpha_3=1.$
b. $M=3,\ L=1,\ \alpha_0=\alpha_1=\alpha_2=\alpha_3=1.$
c. $M=3,\ L=2,\ \alpha_0=0.2,\ \alpha_1=\alpha_2=\alpha_3=1.$



E{d}
150
100
50
10
N
10
50
100
150
200
250
300
350
400

REFERENCES

1. 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.
2. O.J. BOXMA (1985). *Response times in cyclic queues - the influence of the slowest server,* preprint 380, Mathematical Institute, University of Utrecht.
3. O.J. BOXMA AND P. DONK (1982). *On response time and cycle time distributions in a two-stage cyclic queue,* Perf. Evaluation 2, 181-194.
4. O.J. BOXMA, F.P. KELLY AND A.G. KONHEIM (1984). *The product form for sojourn time distributions in cyclic exponential queues,* J. ACM 31, 128-133.
5. V.G. CERF AND R. KAHN (1974). *A protocol for packet network intercommunication,* IEEE Trans. Commun. 22.
6. J.W. COHEN (1982). *The Single Server Queue,* North-Holland, Amsterdam, revised ed.
7. J.W. COHEN (1977). *Applied Stochastics,* lecture notes (in Dutch), Math. Inst. R.U. Utrecht.
8. J.W. COHEN AND O.J. BOXMA (1984). *Computer Systems Performance Analysis,* lecture notes (in Dutch), Math. Inst. R.U. Utrecht.
9. M. GERLA AND L. KLEINROCK (1980). *Flow control: a comparative survey,* IEEE Trans. Commun. 28, 553-574.
10. F.P. KELLY (1979). *Reversibility and Stochastic Networks,* Wiley, New York.
11. P. KERMANI (1977). *Switching and flow control techniques in computer communication networks,* Ph.D. Dissertation, School of Engineering and Applied Science, University of California.
12. L. KLEINROCK (1976). *Queueing Systems (Computer Applications),* Vol.2, Wiley, New York.
13. L. KLEINROCK (1978). *On flow control in computer networks,* In: Intern. Conf. Commun., Toronto, vol.2, 27.2.1-27.2.5 .
14. H. KOBAYASHI (1978). *Modeling and Analysis, an Introduction to System Performance Evaluation Methodology,* Addison-Wesley Publ. Co., Reading, MA.
15. M. REISER (1982). *Performance evaluation of data communication systems,* Proc. IEEE 70, 171-196.
16. M. REISER (1979). *A queueing network analysis of computer communication networks with window flow control,* IEEE Trans. Commun. 27, 1199-1209.
17. M. REISER AND H. KOBAYASHI (1975). *Queueing networks with multiple closed chains: theory and computatinal algorithms,* IBM J. Res. Develop. 19, 282-294.
18. C.H. SAUER AND K.M. CHANDY (1981). *Computer Systems Performance Modeling,* Prentice-Hall, Englewood Cliffs, New Jersey.
19. R. SCHASSBERGER AND H. DADUNA (1983). *The time for a roundtrip in a cycle of exponential queues,* J. ACM 30, 146-150.
20. M. SCHWARTZ (1982). *Performance analysis of the SNA virtual route pacing control,* IEEE Trans. Commun. 30, 172-184.
21. A.S. TANENBAUM (1981). *Computer Networks,* Prentice-Hall,Englewood Cliffs, New Jersey.
22. F.A. TOBAGI, M. GERLA, R.W. PEEBLES AND E.G. MANNING (1978). *Modeling and measurement techniques in packet communication networks,* Proc. IEEE 66, 1423-1447.
23. H. ZWEERUS-VINK AND H. OUDSHOORN (1980). *SIMULA voor ALGOL-kenners,* ACCU-series nr. 16, Academisch Computer Centrum, Utrecht.

Appendix A

De L.S.T. $\phi_{i,L+1}(\rho)$ of the distribution function of the entrance time $\boldsymbol{\alpha}_{i,L+1}$.

In (5.10) the L.S.T. $\phi_{i,L+1}(\rho)$ is used to calculate the mean entance time $E\{\boldsymbol{\alpha}_{i,L+1}\}$. We shall derive an expression for $\phi_{i,L+1}(\rho)$, $i=0,...,L$.

$\{\mathbf{x}_t\ ,\ t\geqslant 0\}$ (see (5.6)) is a birth-and-death process with constant birth-and-death rates (respectively $1/\alpha$ and $1/\alpha_M$).
We define for $i,j\geqslant 0$

$$P_{i,j}(t) := Pr\{\mathbf{x}_t=j|\mathbf{x}_0=i\}\ ,$$

$$F_{i,j}(t) := Pr\{\boldsymbol{\alpha}_{i,j}<t\}\ ,$$

$$\pi_{i,j}(\rho) := \int_0^\infty e^{-\rho t} P_{i,j}(t)\, dt\ , \qquad \mathrm{Re}\ \rho>0.$$

Hence

$$\phi_{i,j}(\rho) = \int_0^\infty e^{-\rho t}\, dF_{i,j}(t)\ , \qquad \mathrm{Re}\ \rho\geqslant 0,\ \ i,j\geqslant 0.$$

Using the relation

$$P_{i,j}(t) = \int_0^t P_{j,j}(t-\tau)\, dF_{i,j}(\tau) + \delta_{i,j} e^{-(1/\alpha+1/\alpha_M)}\ ,$$

it can be proved that (see Cohen [7])

$$\phi_{i,j}(\rho) = \frac{\pi_{i,j}(\rho)}{\pi_{j,j}(\rho)}\ , \qquad \mathrm{Re}\ \rho>0. \tag{A.1}$$

In Cohen [6], pp. 78-80, an expression is derived for $\pi_{i,j}(\rho)$, which, for $i\leqslant j$, can be written as follows:

$$\pi_{i,j}(\rho) = \alpha_M \frac{a^{j-i}x_2^{j-i}}{a(x_1-x_2)} + \frac{a^j}{\rho}\,\frac{x_2^{j+i}-2ax_2^{j+i+1}+a^2x_2^{j+i+2}}{a(x_1-x_2)}\ , \qquad i\leqslant j\ , \tag{A.2}$$

with

$$x_1 = \frac{1+a+\alpha_M\rho}{2a} + \frac{1}{2a}\surd((1+a+\alpha_M\rho)^2 - 4a)\,,$$
$$x_2 = \frac{1+a+\alpha_M\rho}{2a} - \frac{1}{2a}\surd((1+a+\alpha_M\rho)^2 - 4a)\,, \quad \text{Re } \rho>0,$$

and

$$a := \frac{\alpha_M}{\alpha}\,.$$

From (A.1) and (A.2) it easily follows that

$$\phi_{i,L+1}(\rho) = \frac{\alpha_M(ax_2)^{L+1-i} + (a^{L+1}/\rho)(x_2^{L+1+i} - 2ax_2^{L+2+i} + a^2x_2^{L+3+i})}{\alpha_M + (a^{L+1}/\rho)(x_2^{2L+2} - 2ax_2^{2L+3} + a^2x_2^{2L+4})}\,.$$

Appendix B

The simulation program

In this appendix we describe the program which has been developed to simulate the queueing model described in Chapter 3. First we present an introduction to the simulation followed by an explanation of the program. A listing of the program is given on pp. 50-53.

B.1 Introduction to the program

The program has been designed in such a way that all the parameters ($M>1$, N, L, α_0, ..., α_M) can be chosen by the user. Although the program also provides a simulation of other performance measures (e.g. $E\{s_0\}$, $Pr\{overflow\}$), we shall only discuss the throughput and the mean end-to-end delay. Measuring the throughput is done in almost the same way as suggested in Section 5.2: during a certain period (the simulation time) the customers who move from Q_{M-1} to Q_M are counted. Next, the throughput (during the simulation time) is found by dividing the number of counted customers by the simulation time.
Determining the end-to-end delays of the customers is more difficult, because one has to keep stock of the type of each customer and, if necessary, the type has to be adapted (see Section 5.6). In the program the determination of an end-to-end delay of a customer, say K, is done as follows. As soon as K, departing from Q_0, starts an end-to-end delay, he is "marked" with the time, say $t_1(>0)$, at which this event takes place. The next time, say at time t_2, that K moves from Q_M to Q_0, the end-to-end delay t_2-t_1 of K is noted and his mark t_1 is replaced by 0 (the "type 2" mark).
After a transient period, of which the length can be chosen by the user, the registration of the throughput and the end-to-end delay is started. After a subsequent period (the run time), the length of which can also be chosen by the user, the throughput and the mean of the registered end-to-end delays is computed. Subsequently a new (equally long) run time is started. The whole simulation consists of a transient period followed by ten run times. After the last run, the results for the throughput and the end-to-end delay of each run are averaged, and the sample variances are calculated. Using these quantities it is possible to determine an (approximately) 95% confidence interval for the final results.

The simulation method described above is called the "single run method". Actually, the simulation consists of *one* long run, which is divided in (in this case) 10 equally long parts. The 10 sample means are viewed as 10 observations. More about the single run method and other simulation methods can be found in Sauer and Chandy [18] and Cohen and Boxma [8].

B.2 Explanation of the program.

The program has been written in the simulation language SIMULA'67. An introduction to SIMULA'67 is given by Zweerus-Vink and Oudshoorn [23].

We shall now explain the most important points of the program.
- Lines 4-47 describe the way in which the customers are treated by the servers.

Line 4: nr is the server's number (nr=0,1,...,M).
Lines 10-44: If there are customers present in the queue, then the first customer is served during a service time which is generated by the server (lines 13 and 14). After that the customer is placed in the queue of the next server (if possible). We distinguish four cases.

i. 1⩽nr⩽M-2 (lines 15-20). The customer is just moved to Q_{nr+1}.

ii. nr$=M-1$ (lines 21-29). If less than L customers are present in Q_M then the customer moves to Q_M and the number of counted customers is increased by one.

iii. nr$=M$ (lines 30-37). The customer is moved to Q_0. The end-to-end delay of the customer is noted (line 33) and the customer is marked as a type 2 customer (line 34: k.aanktyd:=0).

iv. nr$=0$ (lines 38-43). the customer is moved to Q_1. If necessary a new end-to-end delay of the customer is started (line 40).

- Lines 49-56 describe the behaviour of a customer who has been placed in a queue. He has to activate the server if there are no waiting customers ahead of him (the server had become passive, see line 45).

- Lines 58-85 describe the way in which the customers are moved from $Q_1, \ldots, Q_{M-1}$ to Q_0 if overflow takes place.

-Lines 87-183 describe a number of procedures which are called in the main program (lines 185-227) to take care of the lay-out, to report the results of each run, to do the statistical calculations, etc.

-Lines 185-227. This is the main program. After the declarations, $M+1$ queues and servers are generated (lines 199-200). Subsequently N customers are generated and placed in Q_0 (lines 202-208). The mean service times, the initial values for the random generator and the duration of the transient period are read (lines 209-213). These values are printed by the procedure writeparams. At line 215 the system is started by activating the first customer in Q_0. After the transient period the run time is read and, by calling the procedure initialize, a table is made in which the results of each run can be stored. Subsequently the variables which keep statistical information are initialized (by using the procedure preparenewrun) and the first run is started (line 223). When the run has finished, the results are printed (line 223). This is 10 times repeated (lines 219-224). Finally, using the procedure report, some statistical calculations are done and the final results are printed.

```
BEGIN
INTEGER M,N; M:=ININT; N:=ININT;
SIMULATION BEGIN

  PROCESS CLASS LOKET(NR); INTEGER NR;
   BEGIN
     REF(KLANT) K;
     REAL BTYD;
     WHILE TRUE DO
     BEGIN
       WHILE NOT QUEUE(NR).EMPTY DO
       BEGIN
         K:-QUEUE(NR).FIRST;
         BTYD:=NEGEXP(1/B(NR),U(NR));
         HOLD(BTYD);
         IF (NR>0 AND NR<M-1) THEN
         BEGIN
           K.INTO(QUEUE(NR+1));
           K.QNO:=NR+1;
           REACTIVATE(K);
         END;
         IF NR=M-1 THEN
         BEGIN
           IF QUEUE(M).CARDINAL<L THEN
           BEGIN
             K.INTO(QUEUE(M)); K.QNO:=M;
             AANTAL:=AANTAL+1; REACTIVATE(K);
           END
           ELSE OVERFLOW;
         END;
         IF NR=M THEN
         BEGIN
           CYCLINO:=CYCLINO+1;
           VERZTYD:=VERZTYD+TIME-K.AANKTYD;
           K.AANKTYD:=0; K.QNO:=0; K.ARRQ0:=TIME;
           K.INTO(QUEUE(0));
             REACTIVATE(K);
         END;
         IF NR=0 THEN
         BEGIN
           IF K.AANKTYD=0 THEN K.AANKTYD:=TIME;
           VERBLTQ0:=VERBLTQ0+TIME-K.ARRQ0; T0:=T0+1;
           K.INTO(QUEUE(1)); K.QNO:=1; REACTIVATE(K);
         END;
       END;
       PASSIVATE;
     END;
   END;

 PROCESS CLASS KLANT(I); INTEGER I;
   BEGIN
     REAL AANKTYD,ARRQ0;
     INTEGER KLANTNO,QNO;
     KLANTNO:=I;
     WHILE TRUE DO
     IF BEDIENDE(QNO).IDLE THEN ACTIVATE BEDIENDE(QNO) ELSE PASSIVATE;
   END;

```

```
PROCEDURE OVERFLOW;
  BEGIN
    INTEGER I; REF(KLANT) WYZER; REF(HEAD) QHULP;
    FLOWNO:=FLOWNO+1;
    QHULP:-NEW HEAD;
    WHILE NOT QUEUE(0).EMPTY DO
    BEGIN
      WYZER:-QUEUE(0).FIRST;
      WYZER.INTO(QHULP);
    END;
    FOR I:=1 STEP 1 UNTIL M-1 DO
    WHILE NOT QUEUE(M-I).EMPTY DO
    BEGIN
      WYZER:-QUEUE(M-I).FIRST;
      WYZER.INTO(QUEUE(0)); WYZER.QNO:=0;
      WYZER.ARRQ0:=TIME;
    END;
    WHILE NOT QHULP.EMPTY DO
    BEGIN
      WYZER:-QHULP.FIRST;
      WYZER.INTO(QUEUE(0));
    END;
    FOR I:=0 STEP 1 UNTIL M-2 DO
    BEGIN
      CANCEL(BEDIENDE(I)); BEDIENDE(I):-NEW LOKET(I);
    END;
    ACTIVATE BEDIENDE(0);
  END;

PROCEDURE WRITEPARAMS;
  BEGIN
    INTEGER J;
    OUTTEXT(" M="); OUTINT(M,3); OUTIMAGE;
    OUTTEXT(" N="); OUTINT(N,3); OUTIMAGE;
    OUTTEXT(" L="); OUTINT(L,3); OUTIMAGE;
    FOR J:=0 STEP 1 UNTIL M DO
    BEGIN
      OUTTEXT(" ALPHA"); OUTINT(J,1);
      OUTTEXT("="); OUTFIX(B(J),1,4);
      OUTIMAGE; OUTIMAGE;
    END;
  END;

PROCEDURE INITIALIZE;
  BEGIN
   OUTTEXT(" INLOOPTYD: "); OUTFIX(INLOOPTYD,1,6);
   OUTIMAGE;
   OUTTEXT(" AANTAL RUNS: "); OUTINT(10,2);
   OUTIMAGE;
   OUTTEXT(" GESIMULEERDE TYD PER RUN: ");
   OUTFIX(RUNTYD,1,7);
   OUTIMAGE; OUTIMAGE;
   OUTTEXT(" =======SIMULATIERESULTATEN=======");
   OUTIMAGE; OUTIMAGE;
   OUTTEXT(" AANTAL AANKOMSTEN BY Q0 GEDURENDE DE INLOOPTYD: ");
   OUTINT(CYCLINO,6); OUTIMAGE;
   OUTTEXT(" THROUGHPUT GEDURENDE DE INLOOPTYD");
   OUTFIX(AANTAL/INLOOPTYD,3,6);
```

```
     OUTIMAGE; OUTIMAGE; OUTIMAGE; OUTIMAGE;
     OUTTEXT(" RUN"); SETPOS(POS+15);
     OUTTEXT(" THROUGHPUT"); SETPOS(POS+15);
     OUTTEXT(" VERZENDTYD"); SETPOS(POS+15);
     OUTTEXT(" PR[OVERFLOW]"); SETPOS(POS+15);
     OUTTEXT(" E[SO]");
     OUTIMAGE; OUTIMAGE; OUTIMAGE;
    END;

 PROCEDURE PREPARENEWRUN;
   BEGIN
     AANTAL:=0; CYCLINO:=0;
     VERZTYD:=0; FLOWNO:=0;
     VERBLTQ0:=0; T0:=0;
   END;

 PROCEDURE RUNREPORT(I); INTEGER I;
   BEGIN
     OUTINT(I,2); SETPOS(21);
     TPUT(I):=AANTAL/RUNTYD;
     VTYD(I):=VERZTYD/CYCLINO;
     OFLOW(I):=FLOWNO/(AANTAL+FLOWNO);
     VBTYDQ0(I):=VERBLTQ0/T0;
     OUTFIX(TPUT(I),3,6); SETPOS(47);
     OUTFIX(VTYD(I),3,7); SETPOS(72);
     OUTFIX(OFLOW(I),3,7); SETPOS(101);
     OUTFIX(VBTYDQ0(I),3,6);
     OUTIMAGE; OUTIMAGE;
   END;

 PROCEDURE REPORT;
   BEGIN
     INTEGER I;
     REAL MEANTPUT,MEANVTYD,MEANOFLOW,VARTPUT,VARVTYD;
     REAL MEANVTQ0;
     FOR I:=1 STEP 1 UNTIL 10 DO
     BEGIN
       MEANTPUT:=MEANTPUT+TPUT(I)/10;
       MEANVTYD:=MEANVTYD+VTYD(I)/10;
       MEANOFLOW:=MEANOFLOW+OFLOW(I)/10;
       MEANVTQ0:=MEANVTQ0+VBTYDQ0(I)/10;
     END;
     FOR I:=1 STEP 1 UNTIL 10 DO
     BEGIN
       VARTPUT:=VARTPUT+(TPUT(I)-MEANTPUT)**2/9;
       VARVTYD:=VARVTYD+(VTYD(I)-MEANVTYD)**2/9;
     END;
     OUTIMAGE;
     OUTTEXT(" -----EINDRESULTATEN-----");
     OUTIMAGE; OUTIMAGE;
     OUTTEXT(" THROUGHPUT: ");
     OUTFIX(MEANTPUT,3,6);
     OUTIMAGE; OUTIMAGE;
     OUTTEXT(" GEMIDDELDE VERZENDTYD: ");
     OUTFIX(MEANVTYD,3,7);
     OUTIMAGE;OUTIMAGE;
     OUTTEXT(" PR[OVERFLOW]:");
     OUTFIX(MEANOFLOW,3,7);
```

```
      OUTIMAGE; OUTIMAGE;
      OUTTEXT(" E[SO]:");
      OUTFIX(MEANVTQ0,3,6);
      OUTIMAGE; OUTIMAGE;
      OUTTEXT(" VARIANTIE THROUGHPUT:");
      OUTFIX(VARTPUT,6,8);
      OUTIMAGE;OUTIMAGE;
      OUTTEXT(" VARIANTIE VERZENDTYD:");
      OUTFIX(VARVTYD,3,8);
    END;

  INTEGER CYCLINO,FLOWNO;
  INTEGER ARRAY U(0:M);
  INTEGER L,AANTAL,T0,J;
  REAL ARRAY B(0:M);
  REAL VERZTYD,VERBLTQ0;
  REF(HEAD) ARRAY QUEUE(0:M);
  REF(LOKET) ARRAY BEDIENDE(0:M);
  REF(KLANT) EERSTE,VOLGENDE;
  REAL STARTTYD,INLOOPTYD,RUNTYD;
  REAL ARRAY TPUT,OFLOW,VTYD,VBTYDQ0(1:10);

  L:=ININT;
  FOR J:=0 STEP 1 UNTIL M DO
  BEGIN
    BEDIENDE(J):-NEW LOKET(J);
    QUEUE(J):-NEW HEAD;
  END;
  EERSTE:-NEW KLANT(1); EERSTE.QNO:=0;
  EERSTE.INTO(QUEUE(0)); EERSTE.AANKTYD:=0;
  FOR J:=2 STEP 1 UNTIL N DO
  BEGIN
    VOLGENDE:-NEW KLANT(J); VOLGENDE.QNO:=0;
    VOLGENDE.INTO(QUEUE(0)); VOLGENDE.AANKTYD:=0;
  END;
  FOR J:=0 STEP 1 UNTIL M DO
   B(J):=INREAL;
  FOR J:=0 STEP 1 UNTIL M DO
   U(J):=ININT;
  INLOOPTYD:=INREAL;
  WRITEPARAMS;
  ACTIVATE EERSTE;
  HOLD(INLOOPTYD);
  RUNTYD:=INREAL;
  INITIALIZE;
  FOR J:=1 STEP 1 UNTIL 10 DO
  BEGIN
    PREPARENEWRUN;
    HOLD(RUNTYD);
    RUNREPORT(J);
  END;
  REPORT;
  END
  END
```