Centrum voor Wiskunde en Informatica
Centre for Mathematics and Computer Science

J.P.C. Blanc, E.A. van Doorn

Relaxation times for queueing systems

Department of Operations Research and System Theory   Report OS-R8407   May

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.).

RELAXATION TIMES FOR QUEUEING SYSTEMS

J.P.C. BLANC, E.A. VAN DOORN
Centre for Mathematics and Computer Science, Amsterdam

When a stochastic queueing model is used for performance analysis of, e.g., a computer or communication system, the steady-state situation is usually assumed to prevail. Since many systems exist where the validity of this assumption is questionable, while determination of the time-dependent behaviour of the system is difficult or even impossible, some simple means to characterize the speed with which system performance measures tend to their steady-state values is called for. In this paper the concept of relaxation time is put forward to provide such a characterization. We give a survey of results pertaining to relaxation times for a variety of queueing models. Also, some conjectures and open problems will be mentioned.

1980 MATHEMATICS SUBJECT CLASSIFICATION: 60K25, 90B22.
KEY WORDS & PHRASES: queueing system, relaxation time, time-dependent behaviour.
NOTE: This paper will appear in *Proceedings of the CWI Symposium Mathematics and Computer Science*, J.W. de Bakker, M. Hazewinkel, J.K. Lenstra (eds.). CWI Monograph 1, North-Holland, Amsterdam.

Report OS-R8407
Centre for Mathematics and Computer Science
P.O. Box 4079, 1009 AB Amsterdam, The Netherlands

1. Introduction

Queueing models are widely used for modelling and performance evaluation of computer and communication systems, as well as in many other fields. Suppose we want to predict the behaviour of a system for various values of the system parameters on the basis of such a model. The usual procedure is then to determine, either analytically or by simulation, steady-state characteristics of the model and to draw conclusions from the results thus found. Although acceptable in a majority of cases, many examples can be found where this procedure leads to conclusions which are unacceptably inaccurate, because no account is taken of the fact that the system starts working at a certain time t_0, say, under initial conditions which are known not to reflect steady-state behaviour. As a consequence, system behaviour at a time t, $t_0<t<\infty$, might deviate considerably from what steady-state results predict. A typical example occurs in a simulation context when one starts at t_0 with an empty system and draws conclusions from observations obtained during the period $[t_0,t_1]$ (see [26] for other examples).

The analytical approach to remedy this difficulty would be to determine the system's time-dependent behaviour under certain initial conditions, but this is notoriously difficult and, even if it is successful, does not often lead to results which are easy to interpret. This being so, it would be helpful to have at least some criterion for deciding whether the use of steady-state results is justified after some time $t-t_0$ has elapsed. Indeed, this is precisely what we need when studying system behaviour by simulation. What we seek, therefore, is some means to characterize the speed with which system characteristics tend to their steady-state values.

Since most time-dependent results available indicate that tendency to steady-state is exponential, we are led to the concept of *relaxation time,* which we define, for a function $f(t)$ tending to a finite limit $f(\infty)$, as

$$T(f) = \mathit{inf}\{T \mid f(t)-f(\infty)=O(e^{-t/T}) \ \ (t\to\infty)\} \tag{1.1}$$

(In some contexts $T^{-1}(f)$ is called the *decay parameter* of f.) Since we can think of numerous functions related to a particular queueing system (e.g., virtual waiting time, average queue length, probability of the system being empty), it is not a priori clear how to relate the concept of relaxation time to a queueing system, indeed, whether it is possible to do so in a sensible way. However, there are good reasons (about which later) to define the *relaxation time T of a queueing system* as

$$T = T(p_{00}), \tag{1.2}$$

where $p_{00}=p_{00}(t)$ is the probability that the system is empty at time $t>0$, given that there are no customers in the system at time 0, and that 0 is an arbitrary point of time with respect to the arrival process to the queueing system.

We remark that Morse [23] seems to have been the first to use the term "relaxation time" in a queueing context, but his definition differs from ours. Cohen's [6] definition of relaxation time for a queueing system, although not explicitly related to p_{00}, coincides with ours (cf. also Kingman [18]).

We were motivated to choose definition (1.2) by the following considerations. A description of the state of a system at t_0 - the initial conditions - requires (probabilistic) specification of at least
- the number of customers in the system,
- the distribution of the residual interarrival time,
- the distribution of the residual service times if there are customers being served.

Any deviation at t_0 from the steady-state situation in any of these factors leads to time-dependent phenomena. Since we cannot hope to be able to deal with all these effects simultaneously (cf. § 3.1), we will concentrate on deviations from the equilibrium distribution of the number of customers in the system. Thus we shall mostly assume that t_0 is an arbitrary point of time with respect to the arrival process and that residual service times are distributed as they are in equilibrium. Then, there are indications, such as Kingman's [19, 20] solidarity theorems for transition probabilities in a Markovian system, and certain results for specific models, that a large number of functions related to a queueing system have a common relaxation time $T(p_{00})$.

In the following sections we shall determine relaxation times for several queueing models, and we shall show that these are the relaxation times for functions other than $p_{00}(t)$ as well, whereby we restrict ourselves to functions related to the number of customers in the system. It should be noted that knowledge of a relaxation time is not sufficient to answer such concrete questions as: how much time does it take for $p_{00}(t)$ to be within 5% of its steady-state value? For, by definition,

$$p_{00}(t)-p_{00}(\infty) = e^{-t/T}g(t),$$

where $g(t)=O(e^{\epsilon t})$ $(t\to\infty)$ for all $\epsilon>0$, so that we need information on $g(t)$ as well. In many cases, however, such information can be given. Birth-death queueing models will be discussed in Section 2, and the $GI/G/1$ queue in Section 3. In Section 4 we present some results and a conjecture for Jackson queueing networks.

We conclude this introduction with some remarks pertaining to the computation of relaxation times. In many instances we can obtain an explicit representation for the Laplace transform

$$\int_0^\infty e^{-\phi t}(p_{00}(t)-p_0)dt, \quad \mathrm{Re}\phi>0 \tag{1.3}$$

(here, and in what follows, $p_0=p_{00}(\infty)$), which is analytic in the half plane $\mathrm{Re}\phi\leqslant 0$ as well, apart from a finite number of isolated algebraic singularities. In that case the asymptotic behaviour of $p_{00}(t)-p_0$ is determined by the singularity ϕ^* which is closest to the imaginary axis; in fact, it is readily seen from [28], that

$$T(p_{00}) = -\{\mathrm{Re}\phi^*\}^{-1}. \tag{1.4}$$

If $p_0>0$ and, instead of the continuation of (1.3), we consider the continuation of the Laplace transform

$$\Omega(\phi) = \int_0^\infty e^{-\phi t}p_{00}(t)dt, \quad \mathrm{Re}\phi>0, \tag{1.5}$$

then there will be an additional pole at $\phi=0$ which, of course, has no influence on $T(p_{00})$.

2. Birth-death queueing models

In this section we will discuss single server queueing systems for which the queue length process $\{N(t), 0\leqslant t<\infty\}$ is a birth-death process. The birth (arrival) rates and death (service) rates will be denoted by λ_n and μ_n, respectively, where n, $n=0,1,...,K$, is the queue length (including the customer in service) and $K-1$ $(1\leqslant K\leqslant\infty)$ the size of the waiting room. We clearly have $\mu_0=0$ and, if $K<\infty$, $\lambda_K=0$. The parameters λ_n and μ_{n+1} for $n=0,1,...,K-1$ are assumed to be positive. The model will be referred to as $M_{(n)}/M_{(n)}/1/K-1$. With appropriate interpretation of the service rates this model encompasses any Markovian multiserver delay and/or loss system with state-dependent arrival and service rates.

It will be convenient to treat the cases $K<\infty$ and $K=\infty$ separately. Before starting off with the finite case we remark that the exponential distributions involved make that of the three factors mentioned in the introduction only the initial queue length distribution influences the time-dependent behaviour of the pertinent model.

2.1. The $M_{(n)}/M_{(n)}/1/K-1$ queueing system ($K<\infty$)

From Karlin and McGregor [17] (see also [9] and references there) we know that the transition probabilities

$$p_{ij}(t) = Pr\{N(t)=j \mid N(0)=i\}, \quad i,j=0,1, ..., K,$$

for the queue length process of the $M_{(n)}/M_{(n)}/1/K-1$ queue can be represented as

$$p_{ij}(t) = p_j + \pi_j \sum_{n=1}^{K} \exp(-x_n t) Q_i(x_n) Q_j(x_n) \sigma(x_n), \tag{2.1}$$

where the π_n are constants defined by

$$\pi_0 = 1, \quad \pi_n = \frac{\lambda_0 \lambda_1 ... \lambda_{n-1}}{\mu_1 \mu_2 ... \mu_n} \quad (n>0), \tag{2.2}$$

and the p_n are the steady-state probabilities of having n customers in the system, which satisfy

$$p_n = \pi_n / \sum_{m=0}^{K} \pi_m. \tag{2.3}$$

Further, the Q_n are polynomials defined by the recurrence relation

$$\lambda_n Q_{n+1}(x) = (\lambda_n + \mu_n - x) Q_n(x) - \mu_n Q_{n-1}(x) \tag{2.4}$$

$$Q_{-1}(x)=0, \quad Q_0(x)=1,$$

and the x_n are the (distinct and positive) zeros of

$$S(x) = \{(\mu_K - x) Q_K(x) - \mu_K Q_{K-1}(x)\}/x \tag{2.5}$$

(note that $Q_n(0)=1$ for all $n \geq 0$, so that $S(x)$ is a polynomial of degree K). Finally, $\sigma(x)$ is given by

$$\sigma^{-1}(x) = \sum_{m=0}^{K} \pi_m Q_m^2(x) \tag{2.6}$$

(note that $\sigma(x)<1$). Assuming that the zeros x_n are numbered in increasing order of magnitude, we have in particular

$$p_{00}(t) - p_0 = \exp(-x_1 t)\{\sigma(x_1) + o(1)\} \quad (t \to \infty), \tag{2.7}$$

from which it follows that the relaxation time of the $M_{(n)}/M_{(n)}/1/K-1$ queue is given by

$$T = x_1^{-1}. \tag{2.8}$$

Thus determination of the relaxation time of the $M_{(n)}/M_{(n)}/1/K-1$ queue amounts to determination of the smallest zero of a polynomial of degree K. In most cases this will have to be done numerically (cf. [22]), but it may happen that an explicit expression for T exists, as in the example of § 2.1.1.

It is evident from (2.1) that if $Q_n(x_1) \neq 0$ for $n=0,1 ..., K$, then T is also the relaxation time for $p_{ij}(t)$ ($i>0$ or $j>0$), and consequently for $p_{\omega j}(t)$ as well, ω indicating any initial queue length distribution which is not the equilibrium distribution. If $Q_n(x_1)=0$ for some n, which is exceptional but not impossible (see § 2.2.1), then T is larger than the relaxation time of $p_{ij}(t)$ ($i=n$ or $j=n$). It can be shown that there is at most one value of n for which $Q_n(x_1)=0$.

Obviously, the state space being finite, T is also the relaxation time for the mean queue length and any higher order moment, irrespective of the initial state distribution.

2.1.1. The $M/M/1/K-1$ queueing system

Our first example illustrating the results of the previous section is the $M/M/1$ queue with finite waiting room of size $K-1$. The arrival and service rates are now independent of the queue length: $\lambda_n=\lambda$, $\mu_n=\mu$, except $\mu_0=\lambda_K=0$. From [16] we obtain the following expression for the polynomials Q_n of (2.4):

$$Q_{n+1}(x) = \left[\frac{\mu}{\lambda}\right]^{n/2}\left\{\left[1-\frac{x}{\lambda}\right]U_n(\xi(x))-\left[\frac{\mu}{\lambda}\right]^{1/2}U_{n-1}(\xi(x))\right\},\ n\geq 0, \tag{2.9}$$

where

$$\xi(x) = \tfrac{1}{2}(\lambda\mu)^{-1/2}(\lambda+\mu-x), \tag{2.10}$$

and the U_n are Chebysev polynomials of the second kind, which satisfy

$$U_n(\cos\theta) = \sin(n+1)\theta/\sin\theta. \tag{2.11}$$

Some simple algebra involving the recurrence formula

$$\begin{aligned} 2\xi U_n(\xi) &= U_{n-1}(\xi)+U_{n+1}(\xi) \\ U_{-1}(\xi)&=0,\quad U_0(\xi)=1 \end{aligned} \tag{2.12}$$

for these polynomials, gives us

$$S(x) = -\left[\frac{\mu}{\lambda}\right]^{K/2} U_K(\xi(x)). \tag{2.13}$$

We see from (2.11) that the zeros of $U_K(\xi)$ are $\cos n\pi/(K+1)$, $n=1,2\ ...,\ K$, so that, by (2.10) and (2.13), the zeros of $S(x)$ are

$$x_n = \lambda+\mu-2(\lambda\mu)^{1/2}\cos n\pi/(K+1),\quad n=1,2,\ ...,\ K. \tag{2.14}$$

It follows that

$$T = x_1^{-1}=\{\lambda+\mu-2(\lambda\mu)^{1/2}\cos\pi/(K+1)\}^{-1}. \tag{2.15}$$

This result is also implicitly contained in [24, p. 65] and [30, p. 13].

We finally note the following. Since $Q_1(x)=1-x/\lambda$, we have $Q_1(\lambda)=0$. Now, if λ and μ are such that

$$\mu = 4\lambda(\cos\pi/(K+1))^2,$$

then

$$x_1 = \lambda+\mu-2(\lambda\mu)^{1/2}\cos\pi/(K+1) = \lambda,$$

so that $Q_1(x_1)=0$. This is an example of the exceptional situation referred to in the previous section, since the relaxation time of, e.g., $p_{10}(t)$ is smaller than T.

2.1.2. The $M/M/K/0$ loss system

The loss system $M/M/K/0$ ($K<\infty$) can be viewed as an $M/M_{(n)}/1/K-1$ queue where $\lambda_n=\lambda$ and $\mu_n=n\mu$, $n=0,1,\ ...,\ K$, except $\lambda_K=0$. Hence the technique of § 2.1 can be applied to find the relaxation time of this system. As observed in [16], the polynomials Q_n of (2.14) can now be identified in terms of Charlier polynomials $c_n(x,a)$, viz.

$$Q_n(x) = c_n\left(\frac{x}{\mu},\frac{\lambda}{\mu}\right), \tag{2.16}$$

where

$$c_n(x,a) = \sum_{m=0}^{n} m!\binom{n}{m}\binom{x}{m}(-a)^{-m}. \tag{2.17}$$

Using the recurrence relation for Charlier polynomials [16], it can subsequently be shown that

$$S(x) = \frac{\lambda}{x}\left\{c_{K+1}\left(\frac{x}{\mu},\frac{\lambda}{\mu}\right)-c_K\left(\frac{x}{\mu},\frac{\lambda}{\mu}\right)\right\}. \tag{2.18}$$

Unfortunately, no explicit expressions for the zeros of $S(x)$ seem to exist in general, so that x_1 has to be evaluated numerically.

2.2. *The $M_{(n)}/M_{(n)}/1/\infty$ queueing system*

Let us assume that the arrival and service rates λ_n and μ_n of the $M_{(n)}/M_{(n)}/1/\infty$ queue satisfy the condition

$$\sum_{n=0}^{\infty}(\lambda_n\pi_n)^{-1}\sum_{m=0}^{n}\pi_m = \infty, \tag{2.9}$$

where the π_n are given by (2.2). This condition, which is fulfilled in all practical applications, ensures that there is a unique set $\{p_{ij}(t)\}$ of transition probabilities satisfying the usual requirements (see, e.g., [27]).

Karlin and McGregor [14] have shown that there is an analogue to (2.1) which reads

$$p_{ij}(t) = p_j+\pi_j\int_{0+}^{\infty}\exp(-xt)Q_i(x)Q_j(x)d\psi(x), \tag{2.20}$$

where the Q_n are the polynomials of (2.4), and the p_n, the steady-state probabilities, are now given by

$$p_n = \pi_n/\sum_{m=0}^{\infty}\pi_m, \tag{2.21}$$

which is to be interpreted as zero if $\Sigma\pi_n=\infty$. Finally, $d\psi$ is the unique mass distribution on $[0,\infty)$ of total mass 1 with respect to which the polynomials $\{Q_n\}_{n=0}^{\infty}$ are orthogonal:

$$\int_{0}^{\infty}Q_i(x)Q_j(x)d\psi(x) = \pi_j^{-1}\delta_{ij}. \tag{2.22}$$

In particular we have

$$p_{00}(t)-p_0 = \int_{0+}^{\infty}\exp(-xt)d\psi(x). \tag{2.23}$$

By $S(d\psi)$ we denote the support of $d\psi$, i.e.,

$$S(d\psi) = \{x\mid\int_{-\epsilon}^{+\epsilon}d\psi(x)>0 \quad \text{for all} \quad \epsilon>0\},$$

and we let

$$\gamma(d\psi) = \inf\{x>0 \mid x \in S(d\psi)\}. \tag{2.24}$$

Callaert [3,4] (see [11] for a simpler proof) has shown that

$$\gamma(d\psi) = \sup\{\eta \mid p_{00}(t)-p_0=O(e^{-\eta t})\ (t\to\infty)\}, \tag{2.25}$$

so that the relaxation time of the $M_{(n)}/M_{(n)}/1/\infty$ queue satisfies

$$T = \{\gamma(d\psi)\}^{-1}. \tag{2.26}$$

When it comes to calculation of T for specific parameters several possibilities present themselves.The first (and most convenient) possibility is that $\{Q_n\}$ constitutes a system of orthogonal polynomials for which the associated mass distribution is known. One such example will be presented in § 2.2.1.

The second possibility is to follow a suggestion by Karlin and McGregor [15] to try to find the Stieltjes transform

$$\Psi(\phi) = \int_{0+}^{\infty} \frac{d\psi(x)}{x+\phi}, \quad |\arg\phi|<\pi,\ |\phi|>0, \tag{2.27}$$

and to calculate the largest singularity of the analytic continuation of $\Psi(\phi)$. This singularity equals $-\gamma(d\psi)$, which is evident from our remark on Laplace transforms in the introduction and the fact that

$$\int_0^{\infty} e^{-\phi t}(p_{00}(t)-p_0)dt = \Psi(\phi), \quad \mathrm{Re}\,\phi>0, \tag{2.28}$$

as can be seen by substituting the representation (2.23) for $p_{00}(t)-p_0$ (cf. [31,§ VIII.4]). The example of § 2.2.3 was analyzed in this manner in [16].

A third possibiity to calculate T is of course to forget about the representation (2.23) and to try to find the Laplace transform of $p_{00}(t)-p_0$ by some other method. The model of § 2.2.2 was originally studied in this way; some of the results presented in § 2.2.3 were thus obtained as well.

If none of these approaches leads to a result which is computationally expedient, one has to content oneself with an approximation for T (cf. [11]).

We next address the question of whether T is the relaxation time of time-dependent functions other than $p_{00}(t)$. As regards the transition probabilities $p_{ij}(t)$ ($i>0$ or $j>0$), it can be shown (see [3,4]) that this is true indeed, except perhaps when $Q_i(\gamma)=0$ or $Q_j(\gamma)=0$ ($\gamma=\gamma(d\psi)$), in which case the relaxation time can be smaller than T. Setting aside this exceptional case (there is at most one value of n for which $Q_n(\gamma)=0$), it follows that the probabilities $p_{\omega j}(t)$, ω denoting any initial distribution with finite support, have relaxation time T. If ω has infinite support, it may happen that the relaxation time of $p_{\omega j}(t)$ is larger than T. An example of this phenomenon will be given in § 2.2.3.

It seems likely that the relaxation time for the mean queue length (and higher order moments) equals T, provided the initial distribution does not have the disturbing effects described above. A thorough investigation of this question on the basis of (2.20) requires more insight into the behaviour of the polynomials $\{Q_n\}$ than is present as yet.

2.2.1. The system $M/M/\infty$

The "queue length" process in the system $M/M/\infty$ is probabilistically equivalent to that of an $M/M_{(n)}/1/\infty$ queue where $\lambda_n=\lambda$ and $\mu_n=n\mu$, $n=0,1,\ldots$, so that the representation (2.20) is valid. The polynomials Q_n corresponding to this queue can be identified in terms of Charlier polynomials as in

(2.16). Since the Charlier polynomials $c_n(x,a)$ are orthogonal with respect to a distribution $d\psi$ which consists of masses $e^{-a}a^x/x!$ at the points $x=0,1,\ldots$, it follows that

$$T = \mu^{-1} \tag{2.29}$$

(cf. [16]). This result can also be obtained from Takács' findings for the system $M/G/\infty$ [30, p. 160].

2.2.2. *A queueing model where potential customers are discouraged by queue length*

We next consider a system of the type $M_{(n)}/M/1/\infty$, where $\mu_{n+1}=\mu$ and $\lambda_n=\lambda/(n+1)$, $n\geq 0$, which is intended to model decreasing willingness of a customer to join the queue as the queue length increases (see [25], [8] and references there; a more general model is studied in [5]).

Either directly from [5] or via the expression for the Laplace transform of $p_{00}(t)$ as given in [25], it follows that

$$T = 4\{(\lambda+4\mu)^{1/2}-\lambda^{1/2}\}^{-2}. \tag{2.30}$$

It is interesting to compare this result with the relaxation time for the $M/M/1/K-1$ queue of § 2.1.1, which can also be said to model customer reluctance to join a long queue. It is seen that for fixed μ the relaxation time of (2.30) tends to infinity, whereas the relaxation time of (2.15) tends to zero as λ goes to infinity.

2.2.3. *The $M/M/s$ queueing system* ($s<\infty$)

The distribution of the number of customers in the queue $M/M/s$ ($0<s<\infty$) can be studied in terms of an $M/M_{(n)}/1/\infty$ model where $\lambda_n=\lambda$ and

$$\mu_n = \begin{cases} n\mu & n=0,1,\ldots,s \\ s\mu & n=s+1,\, s+2,\ldots \end{cases}$$

For this model Karlin and McGregor [16] have explicitly determined the Stieltjes transform (2.27). On the basis of this result it was shown in [10, Ch. 6] that the relaxation time of the $M/M/s$ queue satisfies

$$T = (\lambda^{1/2}-(s\mu)^{1/2})^{-2}, \tag{2.31}$$

provided the traffic intensity $\rho=\lambda/s\mu$ is not smaller than some critical value ρ^*. If $\rho<\rho^*$, then T is larger than the right hand side of (2.31); actually, T^{-1} equals λ times the smallest positive root of the equation

$$R_s(x,\rho) = C(x), \tag{2.32}$$

where

$$C(x) = \tfrac{1}{2}\{1-x+\rho^{-1}-\{(1-x+\rho^{-1})^2-4\rho^{-1}\}^{1/2}\}, \tag{2.33}$$

and $R_s(x,y)$ is determined by the recurrence relations

$$\begin{aligned} R_{n+1}(x,y) &= 1-x+\frac{n}{sy}\{1-R_n^{-1}(x,y)\} \quad n=1,2,\ldots,s-1, \\ R_1(x,y) &= 1-x. \end{aligned} \tag{2.34}$$

The critical value ρ^* is the largest root <1 of the equation

$$R_s(1-x^{-1/2},x) = x^{-1/2}, \tag{2.35}$$

if $s>1$, and $\rho^*=0$ if $s=1$. Some values of ρ^* are given in Table 2.1.

Example. For $s=2$ and $\rho<\rho^*=\frac{1}{9}$ an explicit expression can be obtained from the above calculation scheme, viz.,

$$T = 2\{2\lambda+\mu+(\mu^2-4\lambda\mu)^{1/2}\}^{-1}, \tag{2.36}$$

in agreement with [4].

s	ρ^*
1	0
2	0.111
3	0.211
4	0.284
5	0.340
10	0.498

Table 2.1. Critical values for the traffic intensity in the system $M/M/s$

There is an interesting difference between the cases $\rho<\rho^*$ and $\rho\geqslant\rho^*$ as regards the asymptotic behaviour of $p_{00}(t)-p_0$ (cf. [10, Ch. 6]). Namely, if $\rho<\rho^*$ then the distribution $d\psi$ contains an isolated point mass m, say, in the point $\gamma(d\psi)=T^{-1}$, and we have

$$p_{00}(t)-p_0 = \exp(-t/T)\{m+o(1)\} \quad (t\to\infty). \tag{2.37}$$

If $\rho\geqslant\rho^*$, however, the distribution $d\psi$ has zero mass concentrated at $\gamma(d\psi)$, and we have

$$p_{00}(t)-p_0 = \exp(-t/T)\{o(1)\} \quad (t\to\infty). \tag{2.38}$$

The order term in (2.38), although not exponential, can enlarge considerably the speed with which $p_{00}(t)-p_0$ tends to zero as t goes to infinity. This phenomenon will also present itself when one considers functions like the transition probabilities $p_{ij}(t)$ ($i>0$ or $j>0$) and the mean queue length, and explains the fact that Odoni and Roth [26] found the relaxation time to be a conservative measure for the speed to steady state in the $M/M/1$ queue (where $\rho^*=0$). Indeed, for this system a more precise result than (2.38) can be obtained from [6, p. 84] to the effect that

$$p_{00}(t)-p_0 = \exp(-t/T)(t/T)^{-3/2}\left\{\frac{(1-\rho)\rho^{1/4}}{2\pi^{1/2}(1+\rho^{1/2})}+O(t^{-1})\right\} \quad (t\to\infty), \tag{2.39}$$

provided $\rho=\lambda/\mu\neq1$. It can be shown (cf. [29]) that also for $s>1$ a factor $t^{-3/2}$ dominates the order term in (2.38), provided $\rho>\rho^*$ (and $\rho\neq1$).

We remark that Cohen [6, p. 180] has obtained asymptotic expressions for the expected queue length in the system $M/M/1$, which show that the relaxation time for this function is given by (2.31) (with $s=1$) if the initial distribution has finite support.

The $M/M/1$ queue will now provide us with an example of the deteriorating effect an initial distribution with infinite support can have on the relaxation time of $Pr\{N(t)=0\}$. From [6, p. 80] it follows that for $i=0,1,...,$

$$\int_0^\infty e^{-\phi t}p_{i0}(t)dt = \frac{\{C(-\phi/\lambda)\}^{i+1}}{\mu\{1-C(-\phi/\lambda)\}}, \quad \mathrm{Re}\phi>0, \tag{2.40}$$

where $C(\cdot)$ is the function defined in (2.33). Since $C(-\phi/\lambda)$ is bounded in absolute value by 1 for

Reϕ>0, this implies

$$\sum_{i=0}^{\infty}(1-r)r^i \int_0^{\infty} e^{-\phi t} p_{i0}(t)dt = \frac{(1-r)C(-\phi/\lambda)}{\mu\{1-C(-\phi/\lambda)\}\{1-rC(-\phi/\lambda)\}}, \tag{2.41}$$

$$\text{Re}\phi>0, \quad 0<r<1.$$

Apart from the pole at $\phi=0$ ($C(0)=1$), which accounts for the steady-state probability p_0, the largest singularities of the right hand side of (2.41) are the branch point $\phi=-(\lambda^{1/2}-\mu^{1/2})^2$ of the function $C(-\phi/\lambda)$ and the pole at $\phi=-(1-r)(\mu-\lambda/r)$ where $C(-\phi/\lambda)=1/r$. It is readily verified that these singularities coincide for $r=\rho^{1/2}$, and that for $\rho^{1/2}<r<1$ the pole is larger than the branch point. Hence, the relaxation time T_ω of $Pr\{N(t)=0\}$ when the initial distribution ω satisfies $Pr\{N(0)=i\}=(1-r)r^i$, $i=0,1,...$, and $\rho^{1/2}<r<1$, is given by

$$T_\omega = \{(1-r)(\mu-\lambda/r)\}^{-1}. \tag{2.42}$$

Note that $T_\omega \to \infty$ as $r \uparrow 1$ so that any relaxation time between $(\lambda^{1/2}-\mu^{1/2})^{-2}$ and infinity can be realized by a proper choice of r.

For completeness' sake we remark that if the initial distribution is equal to the steady-state distribution, i.e., $r=\rho$, then $p_{00}(t)=1-\rho$ for all $t\geqslant 0$, so that the relaxation time of $Pr\{N(t)=0\}$ is equal to zero.

3. The GI/G/1 queueing system

In this section we will discuss relaxation times for $GI/G/1$ queueing systems. Customers arrive at the service facility according to a renewal process with interarrival time distribution $A(x)$, $x\geqslant 0$, and mean interarrival time α. The service times form a sequence of independent random variables with a common distribution $B(x)$, $x\geqslant 0$, and with mean service time β. Unless stated otherwise, the traffic intensity $\rho=\beta/\alpha$ is assumed to be smaller than 1. We let

$$\alpha^*(\theta) = \int_0^{\infty} e^{-\theta x} dA(x), \quad \text{Re}\theta\geqslant 0, \tag{3.1}$$

$$\beta^*(\theta) = \int_0^{\infty} e^{-\theta x} dB(x), \quad \text{Re}\theta\geqslant 0.$$

The relaxation time for the $GI/G/1$ queueing system was extensively studied in [6, § III.7.3]. Because explicit expressions for the probability $p_{00}(t)$, cf. Section 1, are only available in a few special cases, the asymptotic behaviour of $p_{00}(t)$ as $t\to\infty$ was studied on the basis of its Laplace transform $\Omega(\phi)$, cf. (1.5). The discussion is restricted to queueing systems with service time distributions having Laplace-Stieltjes transforms $\beta^*(\theta)$ with abscissas of convergence $\theta_b<0$, while

$$\beta^*(\theta)\uparrow\infty, \quad \text{as} \quad \theta\downarrow\theta_b. \tag{3.2}$$

It turned out that in this case the Laplace transform $\Omega(\phi)$ possesses an analytic continuation into a part of the left half plane, and the relaxation time of the system is determined by the singularity of $\Omega(\phi)$ with the largest real part (apart from a pole at $\phi=0$). Before stating the general result concerning the relaxation time of the $GI/G/1$ system the queueing systems $GI/M/1$ and $M/G/1$ will be discussed in more detail.

3.1. The GI/M/1 queueing system

For a general, but not lattice, interarrival time distribution $A(x)$ the Laplace transform of $p_{00}^{+}(t)$ is given by, cf. [6, § II.3.4],

$$\Omega^{+}(\phi) = \frac{1}{\phi} - \frac{\zeta(\phi)-\phi}{1-\alpha^{*}(\phi)} \frac{\beta}{\zeta(\phi)}, \quad \mathrm{Re}\phi>0; \tag{3.3}$$

here the superscript + indicates that the first customer arives at $t=0$ (i.e. $N(0+)=1$), and $\zeta=\zeta(\phi)$, $\mathrm{Re}\phi>0$, is the unique root of the equation

$$\alpha^{*}(\zeta) = 1+\beta\phi-\beta\zeta, \quad \mathrm{Re}\zeta>0. \tag{3.4}$$

If the assumption that the first customer arrives at $t=0$ is replaced by the assumption that the first customer arrives at a random instant $t=t_1\geqslant 0$, then it is easily verified, that the Laplace transform $\Omega^{\gamma}(\phi)$ of $p_{00}^{\gamma}(t)$ is given by

$$\Omega^{\gamma}(\phi) = \frac{1}{\phi} - \gamma(\phi)\frac{\zeta(\phi)-\phi}{1-\alpha^{*}(\phi)} \frac{\beta}{\zeta(\phi)}, \quad \mathrm{Re}\phi>0; \tag{3.5}$$

here $\gamma(\phi)=E\{\exp(-\phi t_1)\}$. In particular, if t_1 is distributed as the residual interarrival time at an arbitrary instant, so that $t=0$ is an arbitrary instant in the arrival process — in agreement with our definition of relaxation time, cf. Section 1 — , then we find

$$\Omega(\phi) = \frac{1}{\phi} - \frac{\zeta(\phi)-\phi}{\alpha\phi} \frac{\beta}{\zeta(\phi)}, \quad \mathrm{Re}\phi>0. \tag{3.6}$$

Let ζ_0 be the root of $\alpha^{*\prime}(\zeta)=-\beta$ with the largest real part (ζ_o is the unique positive root of $\alpha^{*\prime}(\zeta)=-\beta$ in the case $\rho<1$). Then $\zeta(\phi)$ has a branch point at $\phi_0=\zeta_0-[1-\alpha^{*}(\zeta_0)]/\beta$. The Laplace transform $\Omega(\phi)$ is regular in the domain $\mathrm{Re}\phi>\phi_0$, apart from a pole at $\phi=0$. This implies that the relaxation time of the $GI/M/1$ system is equal to

$$T = -1/\phi_0. \tag{3.7}$$

It should be noted however, that if $t=0$ is not an arbitrary instant in the arrival process, it can happen that the transient effects due to the arrival process dominate those due to the queueing mechanism, cf. (3.3), (3.5). To be more precise, the Laplace transform $\Omega^{+}(\phi)$ possesses poles in the domain $\mathrm{Re}\phi>\phi_0$ at points ϕ for which $\alpha^{*}(\phi)=1$ (the same statement holds for $\Omega^{\gamma}(\phi)$ as far as zeros of $1-\alpha^{*}(\phi)$ are not compensated by zeros of $\gamma(\phi)$). Let ϕ_a be the real part of the root(s) of $\alpha^{*}(\phi)=1$, $\phi\neq 0$, with the largest real part(s). Then the relaxation time of $p_{00}^{+}(t)$ is equal to $\max\{-1/\phi_0, -1/\phi_a\}$. As an example consider the $E_m/M/1$ queueing system. For this system ϕ_0 and ϕ_a can be obtained explicitly:

$$\phi_0 = -[1+m\rho-(m+1)\rho^{m/(m+1)}]/\beta, \tag{3.8}$$

$$\phi_a = -m\rho[1-\cos(2\pi/m)]/\beta.$$

Let β be fixed. Because $\phi_a\uparrow 0$ as $\rho\downarrow 0$ and $\phi_0\uparrow 0$ as $\rho\uparrow 1$, it is clear that ϕ_a dominates for small values of ρ and that ϕ_0 determines the relaxation time of $p_{00}^{+}(t)$ for ρ close to unity (see Table 3.2 for the values of ρ for which $\phi_0=\phi_a$ for some $E_m/M/1$ systems). For ρ fixed $\phi_a\uparrow 0$ as $m\to\infty$. This is in agreement with the fact that the limit of $p_{00}^{+}(t)$ as $t\to\infty$ does not exist in the $D/M/1$ system.

Remark. The zeros of $1-\alpha^{*}(\phi)$ are not cancelled by zeros of $\zeta(\phi)-\phi$ in (3.3) and (3.5), cf. (3.4), because $\mathrm{Re}\zeta(\phi)\geqslant\zeta_0>0$ for $\mathrm{Re}\phi\geqslant\phi_0$, while zeros of $1-\alpha^{*}(\phi)$ have a non-positive real part.

3.2. The $M/G/1$ queueing system

For the $M/G/1$ queueing system the Laplace transform $\Omega(\phi)$ of $p_{00}(t)$ is given by, cf. [6, § II.4.3],

$$\Omega(\phi) = 1/\xi(\phi), \quad \mathrm{Re}\phi>0; \tag{3.9}$$

here $\xi=\xi(\phi)$, $\mathrm{Re}\phi>0$, is the unique root of the equation

$$\beta^*(\xi) = 1+\alpha\phi-\alpha\xi, \quad \mathrm{Re}\xi>0. \tag{3.10}$$

Because of the assumption (3.2) the equation $\beta^{*\prime}(\xi)=-\alpha$ has a unique root ξ_0 on the interval $(\theta_b,0)$. Therefore, the function $\xi(\phi)$ possesses a branch point at $\phi_0=\xi_0+[\beta^*(\xi_0)-1]/\alpha$. In [6, p. 603] it is shown that $\xi(\phi)$ possesses an analytic continuation into the domain $\mathrm{Re}\phi>\phi_0$, and that it has exactly one zero in $\mathrm{Re}\phi>\phi_0$, viz. $\phi=0$. Hence ϕ_0 is the singularity with the largest real part of the Laplace transform $\Omega(\phi)$ -apart from a pole at $\phi=0$-, which implies that the relaxation time of the $M/G/1$ queueing system is equal to

$$T = -1/\phi_0. \tag{3.11}$$

In Table 3.1 relaxation times have been tabulated as a function of ρ for some $M/G/1$ systems in which the service time distributions have rational Laplace-Stieltjes transforms with denominators of degree 1 or 2. Such a Laplace-Stieltjes transform is completely determined by its mean β, its coefficient of variation C_s and its largest pole θ_b (both poles are necessarily real). Note that $T=T(\rho)\downarrow-1/\theta_b$ as $\rho\downarrow 0$ and that $(1-\rho)^2T(\rho)\to 2(C_s^2+1)$ as $\rho\uparrow 1$, cf. (3.19), (3.18). Further note that in the examples the relaxation time is an increasing function of ρ, C_s and θ_b.

Table 3.1. The relaxation time for some $M/G/1$ systems ($\beta=1$); C_s denotes the coefficient of variation of the service time distribution and θ_b the largest pole of $\beta^*(\theta)$.

	$M/E_2/1$	$M/M/1$	$M/H_2/1$	$M/H_2/1$	$M/H_2/1$	$M/H_2/1$
C_s^2	0.5	1	3	3	7	7
θ_b	-2	-1	-0.4	-0.2	-0.2	-0.1
ρ						
$\downarrow$ 0.0	0.500	1.000	2.500	5.000	5.000	10.00
0.1	1.413	2.139	4.962	8.064	9.973	16.59
0.2	2.244	3.273	7.371	10.98	14.83	22.76
0.3	3.428	4.889	10.77	14.99	21.66	31.18
0.4	5.275	7.403	16.01	20.99	32.19	43.74
0.5	8.411	11.66	24.80	30.83	49.86	64.19
0.6	14.35	19.68	41.28	48.84	82.92	101.4
0.7	27.56	37.48	77.58	87.64	155.7	181.0
0.8	66.46	89.72	183.5	198.4	367.8	406.7
0.9	283.1	379.7	767.7	797.3	1537	1617
$\uparrow$ 1.0 $(1-\rho)^2T(\rho)$:	3	4	8	8	16	16

From [6, p. 24] it follows that for $\mathrm{Re}\phi>0$,

$$\int_0^\infty e^{-\phi t}E\{N(t)\mid N(0)=0\}dt = \frac{1}{\alpha\phi^2} - \frac{\xi(\phi)-\phi}{\phi\xi(\phi)}\,\frac{\beta^*(\phi)}{1-\beta^*(\phi)}. \tag{3.12}$$

It seems possible that zeros of $1-\beta^*(\phi)$ rather than the branch point ϕ_0 of $\xi(\phi)$ play a dominating role in the asymptotic behaviour of $E\{N(t)\mid N(0)=0\}$ as $t\to\infty$. In the following discussion let ϕ_s be the real part of the zeros of $1-\beta^*(\phi)$, $\phi\neq 0$, with the largest real part. For the $M/E_k/1$ system ($k=2,3,..$) it is readily obtained that

$$\phi_0 = -[k+\rho-(k+1)\rho^{1/(k+1)}]/\beta, \quad (3.13)$$
$$\phi_s = -k[1-\cos(2\pi/k)]/\beta.$$

Let β fixed. The value of ϕ_s does not depend on ρ, while the branch point ϕ_0 is an increasing function of ρ for $0<\rho<1$. For $k=2,3,4$ we have $\phi_0 \geq \phi_s$ as $\rho\downarrow 0$, so that the relaxation time of $E\{N(t)|N(0)=0\}$ is given by (3.11) for every ρ, $\rho<1$. But for $k>4$ we have $\phi_0<\phi_s$ for ρ small enough. However, it is possible that the zeros of $1-\beta^*(\phi)$ are compensated by zeros of $\xi(\phi)-\phi$ in (3.12), cf. (3.10). This requires further investigation. Incidentally, it can be shown that for ρ large enough ($\rho>1$) the difference between $E\{N(t)|N(0)=0\}$ and its asymptote

$$(a-1)t/\beta + \tfrac{1}{2}[1-C_s^2] + [\beta\xi(0)]^{-1},$$

is $O(\exp(\phi_s t))$ as $t\to\infty$, in contrast with $p_{00}(t)$ which is still $O(\exp(\phi_0 t))$ as $t\to\infty$.

3.3. *General results for the GI/G/1 queueing system*

In [6, § III.7.3] a theorem on the asymptotic behaviour of $p_{00}^+(t)$ (i.e. given $N(0+)=1$, cf. § 3.1) as $t\to\infty$ was proved for the $GI/G/1$ system. This result can be adapted to the case that $t=0$ is an arbitrary instant in the arrival process in a way similar to that of § 3.1:

$$\Omega(\phi) = \frac{1}{\phi} + \frac{1-\alpha^*(\phi)}{\alpha\phi}\left[\Omega^+(\phi)-\frac{1}{\phi}\right], \quad \mathrm{Re}\phi>0. \quad (3.14)$$

From [6, p. 601] and (3.14) it follows that for $\theta_b<\mathrm{Re}\theta<0<\mathrm{Re}\phi$,

$$\Omega(\phi) = \frac{1}{\phi} - \frac{1-\beta^*(\phi)}{\alpha\phi^2} \times \quad (3.15)$$
$$\exp\left\{\frac{-1}{2\pi i}\int_{L_\theta}\left[\frac{1}{\phi-\theta}+\frac{1}{\theta}\right]\log[1-\beta^*(\theta)\alpha^*(\phi-\theta)]d\theta\right\},$$

here L_θ is a line parallel to the imaginary axis. By using the same arguments as in [6, § III.7.3] relation (3.15) leads to the following result.

For $\rho<1$ there exists a unique real value $\phi=\phi_0$, $\theta_b<\phi_0<0$, for which the equation in θ,

$$\beta^*(\theta)\alpha^*(\phi-\theta)=1, \quad \theta \text{ real}, \quad \theta_b<\theta<\phi, \quad (3.16)$$

has a double root. The relaxation time T of the $GI/G/1$ system is equal to $-1/\phi_0$, and

$$p_{00}(t)-p_0 = O((t/T)^{-3/2}\exp(-t/T)), \quad t\to\infty, \quad (3.17)$$

if $A(x)$ and $B(x)$ are not lattice distributions.

Remark. If $A(x)$ or $B(x)$ is a lattice distribution, then the Laplace transform $\Omega(\phi)$ has beside the branch point $\phi=\phi_0$ other branch points with $\mathrm{Re}\phi=\phi_0$. Therefore, the relaxation time for such systems is still $T=-1/\phi_0$, but the behaviour of $p_{00}(t)-p_0$ as $t\to\infty$ is different from (3.17) because of the contributions of the other poles on the line $\mathrm{Re}\phi=\phi_0$.

It should be noted, that if $p_{00}^+(t)$ or $p_{00}^{\gamma}(t)$ is considered (as for the $GI/M/1$ system in § 3.1) then the Laplace transform possesses poles at zeros ϕ, $\phi\neq 0$, of the function $1-\alpha^*(\phi)$. In Table 3.2 the traffic intensity ρ_a for which the largest real part ϕ_a of these poles coincides with ϕ_0 is shown for several models. For $\rho<\rho_a$ the relaxation time of $p_{00}^+(t)$ is equal to $-1/\phi_a$, for $\rho>\rho_a$ to $-1/\phi_0$.

Table 3.2. The traffic intensity ρ_0 for which $\phi_0 = \phi_a$

System	ρ_a
$E_2/M/1$	0.125
$E_2/E_2/1$	0.172
$E_2/E_4/1$	0.212
$E_2/D/1$	0.278
$E_4/M/1$	0.134
$E_4/E_2/1$	0.192
$E_4/E_4/1$	0.250
$E_4/D/1$	0.368

In [6, p. 612] the following heavy traffic limit was derived for the relaxation time $T = T(\rho) = -1/\phi_0$. Let the Laplace-Stieltjes transforms $\beta^*(\theta)$ and $\alpha^*(\theta/\alpha)$ be fixed (thus only the mean interarrival time α varies). Then

$$\lim_{\rho \uparrow 1}(1-\rho)^2 T(\rho) = \lim_{\rho \uparrow 1} 4(1-\sqrt{\rho})^2 T(\rho) = 2\beta\{C_s^2 + C_a^2\}, \tag{3.18}$$

here $C_s(C_a)$ is the coefficient of variation of the service (interarrival) time distribution. On the other hand it is not difficult to see that for light traffic the following limit holds

$$\lim_{\rho \downarrow 0} T(\rho) = -1/\theta_b. \tag{3.19}$$

In the following cases explicit expressions were found for the branch point ϕ_0 of $\Omega(\phi)$, cf. [6, p. 613, 614]. For the $E_m/E_k/1$ system $(m,k = 1,2, \ldots)$

$$\phi_0 = -[k + m\rho - (m+k)\rho^{m/(m+k)}]/\beta; \tag{3.20}$$

for the $E_m/D/1$ system $(m = 1,2, \ldots)$,

$$\phi_0 = -m[\rho - 1 - \log\rho]/\beta. \tag{3.21}$$

These expressions have been used to obtain the relaxation times listed in Table 3.3.

Table 3.3. The relaxation times of some $GI/G/1$ systems as a function of the traffic intensity ρ ($\beta=1$)

ρ	$M/M/1$	$M/E_4/1$	$E_4/M/1$	$E_4/E_4/1$	$M/D/1$	$E_4/D/1$
↓ 0.0	1.000	0.250	1.000	0.2500	0.0000	0.0000
0.1	2.139	1.058	1.646	0.5347	0.7130	0.1782
0.2	3.273	1.736	2.379	0.8181	1.235	0.3089
0.3	4.889	2.703	3.429	1.222	1.984	0.4961
0.4	7.403	4.215	5.057	1.851	3.162	0.7904
0.5	11.66	6.791	7.797	2.914	5.177	1.294
0.6	19.68	11.68	12.94	4.921	9.023	2.256
0.7	37.48	22.60	24.27	9.370	17.64	4.411
0.8	89.72	54.83	57.33	22.43	43.21	10.80
0.9	379.7	234.8	239.8	94.93	186.5	46.64
↑ 1.0 $(1-\rho)^2T(\rho)$:	4.0	2.5	2.5	1.0	2.0	0.5

4. Queueing networks

In the last decade there has been a growing interest in the analysis of queueing networks, particularly because of their frequent occurrence in the modeling of computer systems, cf. [21]. Until recently little was known about the time-dependent behaviour of queueing networks. The concept of relaxation time has been discussed on the basis of diffusion approximations, but this will lead at most to heavy traffic limits for the relaxation times. Lately, however, some exact results were obtained for the relaxation times of networks of the type described by Jackson [12]. In § 4.1 the relaxation time of an open network with infinitely many servers at each node will be discussed. For such a network the relaxation time is determined by the eigenvalue with the smallest real part of a matrix which depends on the transition matrix and the mean service times at the nodes. In § 4.2 the relaxation time of an open network with two single server nodes will be presented. This relaxation time is equal to the maximum of two values which can be assigned each to a different node, and which are equal to the relaxation times of the two nodes when considered separately as $M/M/1$ queueing systems. This result will be used in § 4.3 to formulate a conjecture on the relaxation time of an open Jackson network with an arbitrary number of single server nodes.

In this section the following notation will be used. The network consists of N nodes ($N=2,3,..$). The external arrivals form a Poisson process with constant mean interarrival time α. Each node has an infinite capacity (for service or waiting). Each time a customer visits a node he awaits his turn (if necessary) and receives service during an exponentially distributed amount of time. At each node the mean serivce time is fixed. We define for $j,k=1,...,N$,

c_j : the probability that a customer enters the network at node j;
β_j : the mean service time at node j;
q_{kj} : the probability that a customer goes to node j when his service at node k has been completed;
λ_j : the arrival rate (external and internal) at node j,
ρ_j : the traffic intensity at node j.

The network is assumed to be stable. Then the arrival rates λ_j satisfy the set of equations, cf. [12],

$$\lambda_j = c_j/\alpha + \sum_{k=1}^{n} q_{kj}\lambda_k, \quad j=1,..,N. \tag{4.1}$$

The network with infinite server nodes is stable for all values of the parameters. The network with single server nodes is stable iff

$$\rho_j = \lambda_j\beta_j<1, \quad j=1,..,N. \tag{4.2}$$

The transition matrix $Q=(q_{kj})$ is assumed to be such that every customer leaves the system after having received a finite number of services, i.e., the network is open. Finally, the trivial case $\lambda_j=0$ for some j, $j=1,...,N$, is excluded.

4.1. Jackson networks with infinite server nodes

In [13] a method was developed for the solution of the functional equation for the generating function of the joint queue length distribution for a queueing system with a Poisson arrival stream and N infinite exponential server stations in series. With the aid of this result it is not difficult to derive that the relaxation time of this system is equal to

$$T = \max_{1\leq j\leq N}\{\beta_j\}, \tag{4.3}$$

i.e., to the maximum of the relaxation times of these service systems at each of the nodes when considered separately as $M/M/\infty$ systems, cf. § 2.2.1. Further, it can be seen that

$$p_{00}(t)-p_0 = O(P_{n-1}(t/T)e^{-t/T}), \quad t\to\infty; \tag{4.4}$$

here $P_n(\cdot)$ is a polynomial of degree n, and n, $1\leq n\leq N$, is equal to the number of nodes j for which $\beta_j=T$.This system provides us with an example in which the coefficient of $\exp(-t/T)$ in the expansion of $p_{00}(t)$ as $t\to\infty$ is not bounded. However, calculations have shown that for $m=1,2$ the function $P_m(t/T)\exp(-t/T)$ is decreasing for $t>0$.

The relaxation time of a network with a finite number of infinite exponential server stations — with general transition matrix $Q=(q_{kj})$ and arbitrary probabilities c_j, $j=1,...,N$ — can be obtained by noting that such a network is equivalent to a (one node) $M/G/\infty$ system when only the total number of customers in the system is considered. In this $M/G/\infty$ system the service time distribution $B(x)$ is equal to the distribution of the total sojourn time of a customer in the original network. Takács [30, p. 160] showed that for a $M/G/\infty$ system

$$p_{00}(t) = \exp\{-\frac{1}{\alpha}\int_0^t[1-B(x)]dx\}. \tag{4.5}$$

By determining the distribution $B(x)$ it follows that the relaxation time T of a network with N infinite exponential server stations (and that of the related $M/G/\infty$ system) is equal to $-1/\mathrm{Re}\gamma_{\min}$, where $\gamma_{\min}$ is the eigenvalue with the smallest real part of the matrix

$$\begin{bmatrix} (1-q_{11})/\beta_1 & -q_{12}/\beta_1 & \cdots & -q_{1N}/\beta_1 \\ -q_{21}/\beta_2 & (1-q_{22})/\beta_2 & \cdots & -q_{2N}/\beta_2 \\ \vdots & \vdots & & \vdots \\ -q_{N1}/\beta_N & -q_{N2}/\beta_N & \cdots & (1-q_{NN})/\beta_N \end{bmatrix}. \tag{4.6}$$

Note that T is independent of the mean interarrival time α and of the probabilities c_j, $j=1,..,N$. In the special case $N=2$, $q_{11}=q_{22}=0$, it is found that

$$T = 2\beta_1\beta_2[\beta_1+\beta_2-\sqrt{(\beta_1+\beta_2)^2-4\beta_1\beta_2(1-q_{12}q_{21})}]^{-1}. \tag{4.7}$$

4.2. Jackson networks with two single server nodes

With the aid of a recently developed method for the solution of functional equations for the generating functions of the joint time-dependent queue length distributions for queueing systems with two waiting lines, cf. [7], the following results have been obtained. In [1] it was shown that for the special network with $N=2$, $c_1=1$, $c_2=0$, $q_{12}=0$, $q_{11}=q_{22}=0$, i.e. two single server stations in series, the relaxation time is equal to

$$T = \max_{j=1,2}\{\beta_j/(1-\sqrt{\rho_j})^2\}, \tag{4.8}$$

i.e., T is equal to the maximum of the relaxation times of the service systems at the two nodes when considered separately as $M/M/1$ queueing systems with arrival rate $1/\alpha$, cf. § 2.2.3. Moreover, it was found that in the asymmetric case $\rho_1 \neq \rho_2$,

$$p_{00}(t)-p_0 = O((t/T)^{-3/2}e^{-t/T}), \quad \text{as} \quad t\to\infty, \tag{4.9}$$

cf. (2.39), while in the symmetric case $\rho_1=\rho_2$,

$$p_{00}(t)-p_0 = O((t/T)^{-1/2}e^{-t/T}), \quad \text{as} \quad t\to\infty. \tag{4.10}$$

Comparing this result with that of § 4.1 leads one to conjecture that the relaxation time of a system with an arbitrary number of single server nodes in series is equal to the maximum of the relaxation times of the nodes when considered separately, and that

$$p_{00}(t)-p_0 = O(P_{n-1}(t/T)(t/T)^{-3/2}e^{-t/T}), \quad t\to\infty,$$

where n is the number of nodes with relaxation time equal to T. Recently, the relaxation time of the general Jackson network with two single server nodes was obtained by the first author. It was found to be equal to

$$T = \max_{j=1,2}\{\beta_j m_j/(1-\sqrt{\rho_j})^2\}, \tag{4.11}$$

here m_j, $j=1,2$, is the mean number of visits of a fixed customer to node j given that this customer visits node j at least once:

$$m_j = (1-q_{jj})/[(1-q_{11})(1-q_{22})-q_{12}q_{21}]. \tag{4.12}$$

The following gives an interpretation of (4.11). The relaxation time of a single $M/M/1$ system with feedback is equal to

$$T = \beta(1-p)^{-1}/(1-\sqrt{\rho})^2, \tag{4.13}$$

where β is the mean service time, ρ the traffic intensity and p the probability that a customer returns to the queue after completion of one of his services. The result (4.13) is clear if one notes that the Laplace-Stieltjes transform of the distribution of the total service time which a customer receives is given by

$$\sum_{m=1}^{\infty}(1-p)p^{m-1}(1+\beta\theta)^{-m} = [1+\beta\theta/(1-p)]^{-1}. \tag{4.14}$$

In (4.13) the factor $(1-p)^{-1}$ can be interpreted as the mean number of services received by a fixed customer. Hence, from (4.11) it is seen that the relaxation time of a Jackson network with two single server nodes is equal to the maximum of the relaxation times of the service systems of the two nodes when considered separately as $M/M/1$ queues with feedback. Note the different effects which the transition matrix Q has on the relaxation time of networks with infinite server nodes, cf. (4.7), and on that of networks with single server nodes, cf. (4.11), (4.12).

We finally note that a result similar to (4.3) and (4.8) was found for quite a different model. In [2] the time-dependent behaviour of an $M/G/1$ system with two types of customers and a paired service discipline (i.e. a pair of customers of different type is served if possible, if the customer population consists of one type only, a single customer is served) was studied. Let α be the mean interarrival time, c_j, $j=1,2$, the probability that a customer is of type j, and $B(x)$ the service time distribution. Then the relaxation time of this system is equal to the maximum of the relaxation times of two ordinary $M/G/1$ systems with mean interarrival time α/c_j, $j=1,2$, and service time distribution $B(x)$, cf. § 3.2.

4.3. A conjecture on the relaxation time of an open Jackson network

Unfortunately, the method with which the results in § 4.2 were obtained is at present not applicable in the analysis of queueing systems with more than two waiting lines. However, the form of the relaxation time (4.11) for a Jackson network with two nodes lends itself to a conjecture on the relaxation time of a Jackson network with an arbitrary number of single server nodes. In fact we conjecture that the relaxation time of such a network is equal to

$$T = \max_{1 \leq j \leq N} \{\beta_j m_j / (1-\sqrt{\rho_j})^2\}; \tag{4.15}$$

here m_j, $j=1,..,N$, has the same meaning as in § 4.2. By interpreting the transition process as a Markov chain, m_j is readily found to be

$$m_j = |I-Q|_{jj} / |I-Q|, \quad j=1,\ldots,N; \tag{4.16}$$

here I is the $N \times N$ identity matrix, $Q=(q_{kj})_{1 \leq k,j \leq N}$, $|I-Q|$ is the determinant of $I-Q$, and $|I-Q|_{jj}$ is the determinant of the submatrix of $I-Q$ obtained by deleting the j^{th} row and column. Note that $|I-Q| \neq 0$, because the network is open.

Remark. Probably the relaxation time of a Jackson network with an arbitrary number of servers at each node will also have a form like (4.15) — with a proper modification, cf. (2.31) — provided that the traffic intensities ρ_j are not too small, cf. § 2.2.3. On the other hand it is not easy to generalize (4.15) to a network with more than one class of customers, because the relaxation time of a single $M/M/1$ system with K customer classes is equal to that of a $M/H_K/1$ system, and the latter has a form different from (2.31), cf. § 3.2.

Finally, the main implication of the hypothesis (4.15) will be discussed. Consider for example the following network with three nodes ($N=3$). Let $\beta_1=\beta_2=\beta_3=\beta$, $c_1=1$, $c_2=c_3=0$, and

$$Q = \begin{bmatrix} 0 & \epsilon & \epsilon \\ 0 & 0 & 1-\epsilon \\ 0 & 1-\epsilon & 0 \end{bmatrix}, \quad 0<\epsilon \leq \tfrac{1}{2},$$

see Figure 4.1. It is readily verified, that for this network the traffic intensities at the three nodes are the same, and equal to $\rho=\beta/\alpha$. However, $m_1=1$ and $m_2=m_3=[\epsilon(2-\epsilon)]^{-1} \geq 4/3$, cf. (4.16), so that the hypothesis (4.15) implies

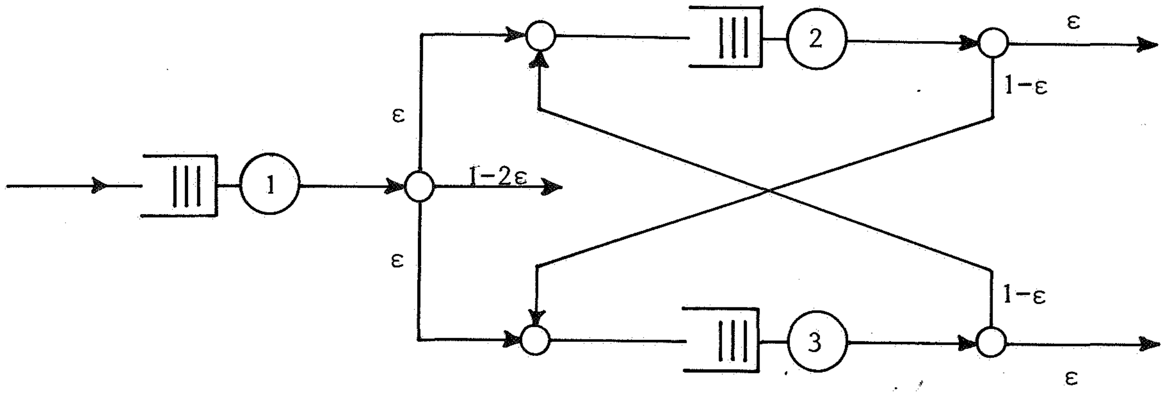

Figure 4.1. A three-node network

$$T = \frac{1}{\epsilon(2-\epsilon)} \frac{\beta}{(1-\sqrt{\rho})^2} . \qquad (4.17)$$

Note that the right hand side of (4.17) tends to infinity as $\epsilon \downarrow 0$ for fixed traffic intensity ρ. In general, the hypothesis (4.15) implies that for nodes with the same traffic intensity the relaxation time of a node (when considered separately) is larger when a few customers are served many times than when many customers are served a few times. In the above example the mean number of nodes visited by a fixed customer is equal to 3, but given that the customer reaches node 2 or 3, it is equal to $2+(1-\epsilon)/\epsilon$. Hence, for ρ fixed and ϵ small enough there are customers with an arbitrarily long expected sojourn time in the system. It is likely that this fact influences the relaxation time of the network. Therefore it supports the conjecture (4.15).

Remark. If the same network as above is considered, but with infinite server nodes, then the eigenvalues of the matrix (4.6) are $1/\beta$, ϵ/β and $(2-\epsilon)/\beta$. Hence, the relaxation time of this network is

$$T = \beta/\epsilon.$$

This relaxation time T also tends to infinity as $\epsilon \downarrow 0$.

References

[1] J.P.C. BLANC, *The relaxation time of two queueing systems in series,* Report BW 192/83. Centre for Mathematics and Computer Science, Amsterdam, 1983.

[2] J.P.C. BLANC, *Asymptotic analysis of a queueing system with a two-dimensional state space,* to appear in J. Appl. Probab.

[3] H. CALLAERT, *Exponentiële ergodiciteit voor geboorte- en sterfteprocessen,* Ph. D. thesis, University of Louvain, 1971.

[4] H. CALLAERT, *On the rate of convergence in birth-and-death processes,* Bull. Soc. Math. Belg. 26 (1974), 173-184.

[5] T.S. CHIHARA AND M.E.H. ISMAIL, *Orthogonal polynomials suggested by a queueing model,* Adv. in Appl. Math. 3 (1982), 441-462.

[6] J.W. COHEN, *The Single Server Queue,* Rev. ed., North-Holland, Amsterdam, 1982.

[7] J.W. COHEN AND O.J. BOXMA, *Boundary Value Problems in Queueing System Analysis,* North-Holland, Amsterdam, 1983.

[8] E.A. VAN DOORN, *The transient state probabilities for a queueing model where potential customers are discouraged by queue length,* J. Appl. Probab. 18 (1981), 499-506.

[9] E.A. VAN DOORN, *On the time dependent behaviour of the truncated birth-death process,* Stochastic Process. Appl. 11 (1981), 261-271.

[10] E.A. VAN DOORN, *Stochastic Monotonicity and Queueing Applications of Birth-Death Processes,* Lecture Notes in Statistics 4, Springer-Verlag, New York, 1981.

[11] E.A. VAN DOORN, *Conditions for exponential ergodicity and bounds for the decay parameter of a birth-death process,* Report BW 174/82, Centre for Mathematics and Computer Science, Amsterdam, 1982.

[12] J.R. JACKSON, *Networks of waiting lines,* Oper. Res. 5 (1957), 518-521.

[13] R.R.P. JACKSON AND P. ASPDEN, *A transient solution to the multistage Poisson queueing system with infinite servers,* Oper. Res. 28 (1980), 618-622.

[14] S. KARLIN AND J.L. MCGREGOR, *The differential equations of birth-and-death processes and the Stieltjes moment problem,* Trans. Amer. Math. Soc. 85 (1957), 489-546.

[15] S. KARLIN AND J.L. MCGREGOR, *The classification of birth and death processes,* Trans. Amer. Math. Soc. 86 (1957), 366-400.

[16] S. KARLIN AND J.L. MCGREGOR, *Many server queueing processes with Poisson input and exponential service times,* Pacific J. Math. 8 (1958), 87-118.

[17] S. KARLIN AND J.L. MCGREGOR, *Ehrenfest urn models,* J. Appl. Probab. 2 (1965), 352-376.

[18] J.F.C. KINGMAN, *On queues in which customers are served in random order,* Proc. Cambridge Philos. Soc. 58 (1962), 79-91.

[19] J.F.C. KINGMAN, *The exponential decay of Markov transition probabilities,* Proc. London Math. Soc. 13 (1963), 337-358.

[20] J.F.C. KINGMAN, *Ergodic properties of continuous-time Markov processes and their discrete skeletons,* Proc. London Math. Soc. 13 (1963), 593-604.

[21] L. KLEINROCK, *Queueing Systems* II, John Wiley, New York, 1976.

[22] F. MACHIHARA, *On the property of eigenvalues of some infinitesimal generator,* Oper. Res. Lett. 2 (1983), 123-126.

[23] P.M. MORSE, *Stochastic properties of waiting lines,* Oper. Res. 3 (1955), 255-261.

[24] P.M. MORSE, *Queues, Inventories and Maintenance,* John Wiley, New York, 1958.

[25] B. NATVIG, *On the transient state probabilities for a queueing model where potential customers are discouraged by queue length,* J. Appl. Probab. 11 (1974), 345-354.

[26] A.R. ODONI AND E. ROTH, *An empirical investigation of the transient behaviour of stationary queueing systems,* Oper. Res. 31 (1983), 432-455.

[27] G.E.H. REUTER, *Denumerable Markov processes and the associated contraction semigroups on l,* Acta Math. 97 (1957), 1-46.

[28] J.P. SCHOUTEN, *Operatorenrechnung,* Springer-Verlag, Berlin, 1961.

[29] J.H.A. DE SMIT, *The time dependent behaviour of the queue length process in the system $M/M/s$,* CORE discussion paper no. 7217, University of Louvain, 1972.

[30] L. TAKÁCS, *Introduction to the Theory of Queues,* Oxford University Press, New York, 1962.

[31] D.V. WIDDER, *The Laplace Transform,* Princeton University Press, Princeton, 1946.