Centrum voor Wiskunde en Informatica
Centre for Mathematics and Computer Science

J.P.C. Blanc

The transient behaviour of networks with infinite server nodes

Department of Operations Research and System Theory Report OS-R8408 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.).

THE TRANSIENT BEHAVIOUR OF NETWORKS WITH INFINITE SERVER NODES

J.P.C. BLANC
Centre for Mathematics and Computer Science, Amsterdam

What amount of time does the mean number of jobs in a Jackson network need to reach its steady-state value within, say, 1%, when the system starts working at time 0 with given numbers of jobs at the various nodes? How is this amount of time related to the relaxation time of the mean number of jobs in the network? These questions are discusses in detail for networks with infinitely many servers at each node. Conjectures are formulated for networks with finitely many servers at the nodes on the basis of simularities with the infinite case for networks with one or two nodes.

1980 MATHEMATICS SUBJECT CLASSIFICATION: 90B22, 60K25.
KEY WORDS & PHRASES: Jackson network, transient behaviour, relaxation time.
NOTE: This report will be submitted for publication elsewhere.

Report OS-R8408
Centre for Mathematics and Computer Science
P.O. Box 4079, 1009 AB Amsterdam, The Netherlands

1. Introduction

In the following we will study the time-dependent behaviour of the number of jobs in a network with a Poisson arrival stream, infinitely many servers at each node, general service time distributions and a homogeneous transition matrix, i.e., an $[M/G/\infty]^K$ network. Although such networks are of independent interest in the modeling of, e.g., migration and illness processes, cf. [3],[6], the main purpose of our investigations is to obtain some general insights into the transient behaviour of Jackson (queueing) networks, cf. [5], which are frequently used to model computer-communication systems, cf. [7]. The advantage of networks with infinite server nodes is that their transient analysis and the resulting formulas are much simpler than those of networks with finitely many servers at the nodes. This becomes already obvious when one compares the time-dependent state probabilities for the $M/M/\infty$ system with those for the M/M/1 system, cf.[9]. Therefore, we will proceed as follows. We shall derive formulas for the time-dependent state probabilities and mean number of jobs in a network with infinite server nodes. Then we will be concerned with questions like: if a system starts (resumes) working at time 0 with a given number of jobs at each of the service centers, after what amount of time τ_ε is (and stays) the average number of jobs in the network within ε% of its limiting value ($0 < \varepsilon < 100$); and how does this initial transient period τ_ε depend on the system parameters? The answer to these questions is important when it has to be decided whether a steady-state approximation is justified or not for describing a network in a given situation. It may also be helpful by determining the length of the initial part of a simulation run which should be ignored when estimating the steady-state distributions of a network. Finally, we shall formulate some conjectures on the time-dependent behaviour of the number of jobs in a queueing network with (exponential) single server nodes. For this purpose we shall express the initial transient periods τ_ε in terms of the relaxation time T of the network, cf. [2]. Let $N_S(t)$ be the total number of jobs in the network at time t, $t \geq 0$. Then the relaxation time of $N_S(t)$ is defined to be the smallest value T for which holds

$$E\{N_S(\infty)\} - E\{N_S(t) \mid N_S(0)\} = g(t)\ e^{-t/T}, \tag{1.1}$$

where

$$g(t) = o(e^{\delta t}), \quad t \to \infty, \quad \text{for all } \delta > 0. \tag{1.2}$$

The relaxation time T is a rather robust measure for the transient characteristics of a system, because the same value T determines the asymptotic behaviour of many quantities related to a system, and because it is independent of the initial state of a system, cf. [4, §III.7.3], [2], and section 2. Recently, the asymptotic behaviour of Jackson networks with two single server nodes was studied ([1],[2]). These results, together with those of the present paper, will be used to formulate conjectures on the asymptotic behaviour of Jackson networks with an arbitrary number of single server nodes. In particular, we will give an upper bound for the initial transient period τ_ε which depends on the system parameters only through the number of nodes K and the relaxation time T of the network. For the latter a conjecture was formulated in [2,§4.3].

This introduction is concluded with some definitions. A Jackson network with infinitely many servers at each node is determined by the following data:

- K — the number of service centers;
- R — the number of job classes;
- α — the mean interarrival time for jobs entering the network;
- $c_{j,s}$ — the probability that a job enters the network at center j as job of class s ($j = 1,..,K$, $s = 1,..,R$);
- $q_{kj,rs}$ — the probability that a job of class r of which the service is completed at center k goes to center j as job of class s ($j,k = 1,..,K$, $s,r = 1,..,R$);
- $B_{j,s}(x)$ — the service time distribution at center j for jobs of class s ($j = 1,..,K$, $s = 1,..,R$).

Let further for $j,k = 1,..,K$, $s,r = 1,..,R$,

- $u_{k,r}$ — the probability that a job of class r leaves the network when a service of this job at center k has been completed, $u_{k,r} = 1 - \Sigma_{j,s}\, q_{kj,rs}$;

$\beta_{j,s}$ mean service time for jobs of class s at center j;

$\beta^*_{j,s}(\theta)$ Laplace-Stieltjes (L.S.) transform of $B_{j,s}(x)$, Re $\theta \geq 0$;

$S(x)$ sojourn time distribution for jobs in the network;

$\sigma^*(\theta)$ L.S. transform of $S(x)$, Re $\theta \geq 0$;

$\lambda_{j,s}$ arrival rate (external and internal) of jobs of class s at center j;

The traffic intensity at service center j is defined to be

$$\rho_j := \sum_{s=1}^{R} \lambda_{j,s}\, \beta_{j,s}, \qquad j=1,..,K, \tag{1.3}$$

and we let $\rho := \sum_{j=1}^{K} \rho_j$. If the network is stable, then $\lambda_{j,s}$ is determined by the following set of linear equations, cf. [5],

$$\lambda_{j,s} = c_{j,s}/\alpha + \sum_{k=1}^{K} \sum_{r=1}^{R} q_{kj,rs}\, \lambda_{k,r}, \qquad j=1,..,K, \quad s=1,..,R. \tag{1.4}$$

2. Transient behaviour of the number of jobs in an initially empty network

In this section the time-dependent behaviour of the total number of jobs in a network with infinitely many servers at each node is studied for the case that the system starts working at time 0 with no jobs present in the network. Because jobs do not have to wait in the networks which we are investigating, they move independently of each other through the networks. Therefore, as long as we only consider the total number $N_S(t)$ of jobs in such a network, this network is probabilistically equivalent to a single M/G/∞ system with mean interarrival time α and service time distribution $S(x)$, the sojourn time distribution in the original network. In [9, p.160] it was shown that for the M/G/∞ system

$$\Pr\{N_S(t)=n \mid N_S(0)=0\} = e^{-\phi(t)}\, [\phi(t)]^n/n!, \qquad n=0,1,.., \tag{2.1}$$

with

$$\phi(t) := \frac{1}{\alpha}\int_0^t [1-S(x)]\, dx = \rho - \frac{1}{\alpha}\int_t^\infty [1-S(x)]\, dx; \tag{2.2}$$

hence

$$E\{N_S(t) \mid N_S(0)=0\} = \phi(t) = \frac{1}{\alpha}\int_0^t [1-S(x)]\, dx. \tag{2.3}$$

From these relations it follows that the asymptotic behaviour of $N_S(t)$ as $t \to \infty$

is determined by the tail of the distribution $S(x)$. For instance, if $1-S(x) = Cx^{-\delta}+o(x^{-\delta})$, $x \to \infty$, $\delta > 1$, then $\phi(t) = \rho - C/\alpha\, t^{1-\delta}/(\delta-1) + o(t^{1-\delta})$, $t \to \infty$. In this case the relaxation time of $N_S(t)$ is infinite, cf. (1.1),(1.2). On the other hand, if $S(x)$ has a finite support, then $\phi(t) = \rho$ for t large enough and the relaxation time of $N_S(x)$ is equal to zero. We will further concentrate on distributions with exponential tails. If

$$1 - S(x) = C\, e^{-\zeta x}\, [x^{\delta} + o(x^{\delta})], \qquad x \to \infty, \tag{2.4}$$

then

$$\phi(t) = \rho - C/(\alpha\zeta)\, e^{-\zeta t}\, [t^{\delta} + o(t^{\delta})], \qquad t \to \infty, \tag{2.5}$$

and the relaxation time of $N_S(t)$ is equal to $1/\zeta$. If δ is here a non-negative integer, then $\sigma^*(\theta)$ is regular for $\text{Re}\ \theta > -\zeta$ and it posseses a pole of order $\delta+1$ at $\theta = -\zeta$. Note that for $\phi(t)$ as in (2.5), cf. (2.1), for $n = 0,1,..,$ as $t \to \infty$,

$$\Pr\{N_S(t) = n \mid N_S(0) = 0\} = e^{-\rho}\, \rho^n/n!\ \{1 + \frac{C}{\alpha\zeta}\, (1 - n/\rho)\, e^{-\zeta t}\, [t^{\delta} + o(t^{\delta})]\}.$$

It is seen that in general these probabilities have the same relaxation time as $\phi(t)$ (apart from exceptional cases such as $n = \rho$ and $1-S(x) = Ce^{-\zeta x} + o(e^{-(\zeta+\xi)x})$, $x \to \infty$, for some $\xi > 0$). Therefore we shall restrict the discussion to the average total number of jobs in the network.

We continue this section by showing that the determination of the distribution $S(x)$ in (2.3) can be reduced to the solution of a set of linear equations. For this purpose we introduce for $j = 1,..,K$, $s = 1,..,R$, the L.S. transform $\eta^*_{j,s}(\theta)$ of the distribution of the residual sojourn time in the network of a job at the instant that it enters center j as job of class s. It is clear that

$$\sigma^*(\theta) = \sum_{j=1}^{K} \sum_{s=1}^{R} c_{j,s}\, \eta^*_{j,s}(\theta), \qquad \text{Re}\ \theta \geq 0. \tag{2.6}$$

Because the successive service times of a job and its routing through the network are all independent, the transforms $\eta^*_{k,r}(\theta)$ satisfy the set of linear equations: for $k = 1,..,K$, $r = 1,..,R$,

$$\eta^*_{k,r}(\theta) = \beta^*_{k,r}(\theta)\, [u_{k,r} + \sum_{j=1}^{K} \sum_{s=1}^{R} q_{kj,rs}\, \eta^*_{j,s}(\theta)], \qquad \text{Re}\ \theta \geq 0. \tag{2.7}$$

The L.S. transform $\sigma^*(\theta)$ is determined by (2.6) and (2.7). From (2.3) we have

$$\int_0^\infty e^{-\theta t}\, E\{N_s(t)\,|\,N_s(0)=0\}\,dt = [1-\sigma^*(\theta)]/(\alpha\theta^2), \qquad \text{Re } \theta > 0. \qquad (2.8)$$

This implies that the relaxation time of $N_s(t)$ is determined by the abscissa of convergence of $\sigma^*(\theta)$. If the transforms $\beta^*_{k,r}(\theta)$ are all rational or entire, then the only singularities of $\sigma^*(\theta)$ are poles due to zeros of the determinant of the set of equations (2.7); if one or more of the transforms $\beta^*_{k,r}(\theta)$ possess other singularities, e.g. branch points, then $\sigma^*(\theta)$ will possess similar singularities at the same positions as those of $\beta^*_{k,r}(\theta)$.

<u>Example.</u> Consider a network with two service centers, one job class, $q_{12,11} = \pi_1$, $q_{21,11} = \pi_2$, $q_{11,11} = q_{22,11} = 0$. Then from (2.6), (2.7), it follows for Re $\theta \geq 0$,

$$\sigma^*(\theta) = [c_{1,1}\,\beta^*_{1,1}(\theta)\,\{1-\pi_1+\pi_1(1-\pi_2)\beta^*_{2,1}(\theta)\} +$$

$$+ c_{2,1}\,\beta^*_{2,1}(\theta)\,\{1-\pi_2+\pi_2(1-\pi_1)\beta^*_{1,1}(\theta)\}][1-\pi_1\pi_2\beta^*_{1,1}(\theta)\beta^*_{2,1}(\theta)]^{-1}.$$

In the case that $\beta^*_{j,1}(\theta) = \exp\{-\beta_{j,1}\theta\}$, $j=1,2$, i.e., constant service times at both centers, the relaxation time of $N_s(t)$ is equal to

$$T = (\beta_{1,1} + \beta_{2,1})\,[-\ln(\pi_1\pi_2)]^{-1}.$$

3. Transient behaviour of the number of jobs in an initially non-empty network

If a network contains jobs at time 0, then the following quantities must be specified in order to be able to describe the number of jobs in the network for $t > 0$: the number of jobs of each class at each center and for each job the distribution of its residual service time. To avoid too intricate notations we will assume that jobs of the same class at the same center have the same residual service time distributions. Hence, we let for $k=1,..,K$, $r=1,..,R$, Re $\theta \geq 0$,

$m_{k,r}$ the number of jobs of class r at center k at time 0;

$\gamma^*_{k,r}(\theta)$ the L.S. transform of the residual service time at time 0 for jobs of class r at center k.

The set of these data will be indicated by the symbol M. Because the jobs move independently through the network, the distribution of the number of jobs in the network at time $t > 0$ given M is the convolution of the distribution of the number of jobs which arrived after time 0 and are present at time t (this distribution was determined in section 2) and the distribution of the number of jobs which were present at time 0 and are still present at time t (a death process). Let $S_{k,r}(x)$ be the distribution of the sojourn time for jobs of class r which are present at time 0 at center k ($k=1,..,K$, $r=1,..,R$), then the above implies that for $t > 0$,

$$E\{N_S(t)\ |M\} = \frac{1}{\alpha}\int_0^t [1-S(x)]\, dx + \sum_{k=1}^{K}\sum_{r=1}^{R} m_{k,r}\ [1 - S_{k,r}(t)]. \tag{3.1}$$

The L.S. transforms $\sigma^*_{k,r}(\theta)$ of $S_{k,r}(x)$ are given by, cf. (2.7), for $k=1,..,K$, $r=1,..,R$,

$$\sigma^*_{k,r}(\theta) = \gamma^*_{k,r}(\theta)\ [u_{k,r} + \sum_{j=1}^{K}\sum_{s=1}^{R} q_{kj,rs}\ \eta^*_{j,s}(\theta)], \qquad \text{Re } \theta \geq 0. \tag{3.2}$$

With (2.7) this can be reduced to: for $k=1,..,K$, $r=1,..,R$,

$$\sigma^*_{k,r}(\theta) = \gamma^*_{k,r}(\theta)\ \eta^*_{k,r}(\theta)/\beta^*_{k,r}(\theta), \qquad \text{Re } \theta \geq 0. \tag{3.3}$$

From (3.1) it follows that for Re $\theta > 0$,

$$\int_0^\infty e^{-\theta t}\ E\{N_S(t)\ |M\}\ dt = [1-\sigma^*(\theta)]/(\alpha\theta^2) + \sum_{k=1}^{K}\sum_{r=1}^{R} m_{k,r}\ [1-\sigma^*_{k,r}(\theta)]/\theta. \tag{3.4}$$

The relaxation time of $N_S(t)$ is determined by the largest of the abscissas of convergence of $\sigma^*(\theta)$, cf. section 2, and of $\gamma^*_{k,r}(\theta)$, $k=1,..,K$, $r=1,..,R$, cf. (3.2).

4. On the joint distribution of the number of jobs at the various centers

The results of the preceding sections, cf. (2.8),(3.4), are suitable for the determination of the relaxation time T of $N_S(t)$, but not so much for the calculation of the instants τ_ε which are defined to be the smallest values with the property that for $t > \tau_\varepsilon$,

$$\psi_0(t) := 1 - E\{N_S(t)\ |N_S(0) = 0\}/E\{N_S(\infty)\} < \varepsilon/100. \tag{4.1}$$

The latter would require the inversion of the L.S. transform $\sigma^*(\theta)$. This section is concerned with a different approach, which makes use of properties of phase-type distributions, cf. [8, Ch.2]. The joint distribution of the number of jobs at the various service centers will be determined in a form which allows for the calculation of the instants τ_ε. First we approximate each service time distribution by a phase-type distribution (phase-type distributions are dense in the set of all distributions with support $(0,\infty)$, cf. [8, Ch.2]). A phase-type distribution can be represented by a network with negative exponentially distributed holding times at each node. Because jobs move independently through the original network, this implies that every network with different job classes and phase-type service time distributions is probabilistically equivalent to a (larger) network with a single job class and exponential service times. To construct the latter network duplicate each center for each job class, and then replace each center by a subnetwork representing the phase-type service time distribution. Therefore, in the following discussion we consider only networks with a single job class and exponential service times. We will use the notation of section 1, deleting the indices which refer to job classes. Results for the original networks can be obtained by aggregation of the appropriate nodes.

Let $G(z_1,..,z_K;t)$ be the generating function of the joint distribution of the number of jobs at the various centers. This function satisfies the following partial differential equation, cf. [3, Ch.8]:

$$\frac{\partial}{\partial t} G + \sum_{k=1}^{K} \frac{1}{\beta_k} [z_k - u_k - \sum_{j=1}^{K} q_{kj} z_j] \frac{\partial}{\partial z_k} G = \frac{1}{\alpha} [\sum_{j=1}^{K} c_j z_j - 1] G. \tag{4.2}$$

By standard methods, cf. [3, Ch.8], it follows that if the system is empty at time 0 then the solution of (4.2) reads:

$$G_0(z_1,..,z_K;t) = \prod_{j=1}^{K} e^{-(1-z_j)\phi_j(t)}, \tag{4.3}$$

where the functions $\phi_j(t)$, $j=1,..,K$, are the solution of the following set of linear differential equations: for $j=1,..,K$,

$$\frac{d}{dt}\phi_j(t) = -[\frac{1}{\beta_j}\phi_j(t) - \sum_{k=1}^{K} q_{kj}\frac{1}{\beta_k}\phi_k(t)] + c_j/\alpha, \tag{4.4}$$

with $\phi_j(0) = 0$. Hence, if the system is empty at time 0, the joint distribution of the number of jobs at the various centers is the product of K marginal Poisson distributions with parameters $\phi_j(t)$, for every $t > 0$, cf. (4.3). Further,

$$E\{N_S(t) \mid N_S(0) = 0\} = \sum_{j=1}^{K} \phi_j(t). \tag{4.5}$$

If the network is not empty at time 0, then the solution of (4.2) is the product of G_0 as in (4.3) and the generating function G_M of the distribution of the number of jobs at the various centers which were present at time 0 and are still present at time $t > 0$. G_M is the solution of (4.2) with the right hand side replaced by zero, satisfying

$$G_M(z_1,..,z_K;0) = \prod_{j=1}^{K} z_j^{m_j}, \tag{4.6}$$

m_j being the number of jobs at center j at time 0, $j=1,..,K$. Note that we do not have to specify residual service times at time 0 as in section 3, because all service times are negative exponentially distributed. We put the general solution of (4.2) obtained in [3, Ch.8] in the following form: for $t \geq 0$,

$$G(z_1,..,z_K;t) = \prod_{j=1}^{K} [1 + \sum_{h=1}^{K} \chi_{hj}(t)\ (z_h - 1)]^{m_j}\ e^{-(1-z_j)\phi_j(t)}; \tag{4.7}$$

here the set of functions $\chi_{hj}(t)$ is the solution of the set of linear differential equations: for every $j = 1,..,K$, for $h = 1,..,K$,

$$\frac{d}{dt} \chi_{hj}(t) = -[\frac{1}{\beta_h} \chi_{hj}(t) - \sum_{k=1}^{K} q_{kj} \frac{1}{\beta_k} \chi_{kj}(t)], \tag{4.8}$$

with initial conditions, cf. (4.6),

$$\chi_{jj}(0) = 1; \qquad \chi_{hj}(0) = 0, \quad \text{if } h \neq j. \tag{4.9}$$

From (4.7) it is seen that in the general case the numbers of jobs at the various centers are not independent. For the average total number of jobs in the network it is obtained from (4.7):

$$E\{N_S(t) \mid M\} = \sum_{j=1}^{K} \phi_j(t) + \sum_{j=1}^{K} m_j \sum_{h=1}^{K} \chi_{hj}(t). \tag{4.10}$$

From (4.10) it follows with (4.4) and (4.8) that the relaxation time T of $N_S(t)$ is equal to $1/\zeta_m$, where ζ_m is the smallest eigenvalue of the matrix (because $\theta = -\zeta_m$ is the pole with the largest real part of $\sigma^*(\theta)$, cf. (4.10),(3.1),(3.4), ζ_m must be real):

$$\begin{pmatrix} (1-q_{11})/\beta_1 & -q_{21}/\beta_2 & \cdots & -q_{K1}/\beta_K \\ -q_{12}/\beta_1 & (1-q_{22})/\beta_2 & \cdots & -q_{K2}/\beta_K \\ \vdots & \vdots & \ddots & \vdots \\ -q_{1K}/\beta_1 & -q_{2K}/\beta_2 & \cdots & (1-q_{KK})/\beta_K \end{pmatrix}. \qquad (4.11)$$

In the next section we shall discuss as an example a network with two service centers. For completeness this section is concluded with the remark that, cf. (4.4), (1.4),(1.3), and (4.8), for h,j = 1,..,K,

$$\lim_{t\to\infty} \phi_j(t) = \rho_j, \qquad \lim_{t\to\infty} \chi_{hj}(t) = 0. \qquad (4.12)$$

This implies with (4.10) that

$$\lim_{t\to\infty} E\{N_S(t)\ |M\} = \sum_{j=1}^{K} \rho_j = \rho. \qquad (4.13)$$

5. An exponential network with two service centers

As an example we shall work out the results of the preceding section for the case of two service centers (K=2). For simplicity we shall take $q_{11}=q_{22}=0$, because feedback at the centers themselves is equivalent with rescaling the parameters β_1, β_2, q_{12}, q_{21}. The eigenvalues of the matrix

$$\begin{pmatrix} 1/\beta_1 & -q_{21}/\beta_2 \\ -q_{12}/\beta_1 & 1/\beta_2 \end{pmatrix}, \qquad (5.1)$$

are

$$\zeta_{1,2} = \frac{1}{2}\left[\frac{1}{\beta_1} + \frac{1}{\beta_2} \pm \sqrt{\left(\frac{1}{\beta_1} + \frac{1}{\beta_2}\right)^2 - 4(1 - q_{12}q_{21})/(\beta_1\beta_2)}\right]. \qquad (5.2)$$

This implies that the relaxation time of the network is equal to

$$T = \frac{1}{2}\left[\beta_1 + \beta_2 + \sqrt{(\beta_1 + \beta_2)^2 - 4(1 - q_{12}q_{21})\beta_1\beta_2}\right]/(1 - q_{12}q_{21}). \qquad (5.3)$$

Note that the eigenvalues ζ_1 and ζ_2 coincide if and only if $\beta_1 = \beta_2$ and $q_{12}q_{21} = 0$, and then $\zeta_1=\zeta_2=1/\beta_1$. If $\zeta_1 \neq \zeta_2$, then the solution of (4.4) in the case $K=2$ is:

$$\phi_1(t) = \rho_1 - \frac{(1/\beta_1-\zeta_2)\rho_1 - q_{21}\rho_2/\beta_2}{\zeta_1-\zeta_2} e^{-\zeta_1 t} - \frac{(1/\beta_1-\zeta_1)\rho_1 - q_{21}\rho_2/\beta_2}{\zeta_2-\zeta_1} e^{-\zeta_2 t},$$

$$\phi_2(t) = \rho_2 - \frac{(1/\beta_2-\zeta_2)\rho_2 - q_{12}\rho_1/\beta_1}{\zeta_1-\zeta_2} e^{-\zeta_1 t} - \frac{(1/\beta_2-\zeta_1)\rho_2 - q_{12}\rho_1/\beta_1}{\zeta_2-\zeta_1} e^{-\zeta_2 t}, \tag{5.4}$$

with, cf. (1.3),

$$\rho_1 = \beta_1/\alpha\ [c_1+c_2q_{21}]/[1-q_{12}q_{21}], \qquad \rho_2 = \beta_2/\alpha\ [c_1q_{12}+c_2]/[1-q_{12}q_{21}]; \tag{5.5}$$

and the solution of (4.8) is:

$$\chi_{jj}(t) = [(1/\beta_j-\zeta_2)e^{-\zeta_1 t} - (1/\beta_j-\zeta_1)e^{-\zeta_2 t}]/(\zeta_1-\zeta_2), \quad j=1,2,$$

$$\chi_{21}(t) = \beta_2/q_{21}\ (1/\beta_1-\zeta_1)(1/\beta_1-\zeta_2)(e^{-\zeta_1 t} - e^{-\zeta_2 t})/(\zeta_1-\zeta_2),$$

$$\chi_{12}(t) = \beta_1/q_{12}\ (1/\beta_2-\zeta_1)(1/\beta_2-\zeta_2)(e^{-\zeta_1 t} - e^{-\zeta_2 t})/(\zeta_1-\zeta_2). \tag{5.6}$$

With (4.10) it follows after some rearrangements that in the case $K=2$, $\zeta_1 \neq \zeta_2$,

$$E\{N_S(t)\,|M\} = \rho_1+\rho_2 - \frac{1/\alpha - (\rho_1+\rho_2)\zeta_2}{\zeta_1-\zeta_2} e^{-\zeta_1 t} - \frac{1/\alpha - (\rho_1+\rho_2)\zeta_1}{\zeta_2-\zeta_1} e^{-\zeta_2 t} +$$
$$+ m_1\ [\{(1-q_{12})/\beta_1 - \zeta_2\}e^{-\zeta_1 t} - \{(1-q_{12})/\beta_1 - \zeta_1\}e^{-\zeta_2 t}]/(\zeta_1-\zeta_2) +$$
$$+ m_2\ [\{(1-q_{21})/\beta_2 - \zeta_2\}e^{-\zeta_1 t} - \{(1-q_{21})/\beta_2 - \zeta_1\}e^{-\zeta_2 t}]/(\zeta_1-\zeta_2). \tag{5.7}$$

In the case $\zeta_1 = \zeta_2$, i.e. two service centers with the same service time distribution in series, we have (choosing $q_{21}=0$):

$$E\{N_S(t)\,|M\} = (\rho_1+\rho_2)[1 - \{1 + t/\beta_1\ c_1q_{12}/(1+c_1q_{12})\}e^{-t/\beta_1}] +$$
$$+ m_1\ \{1 + q_{12}t/\beta_1\}e^{-t/\beta_1} + m_2\ e^{-t/\beta_1}. \tag{5.8}$$

Next consider the function $\psi_0(t)$ defined in (4.1) for the case $K=2$, $\zeta_1 < \zeta_2$. From (5.7) we obtain ($m_1=m_2=0$):

$$\psi_0(t) = e^{-\zeta_1 t}\ [1 + \{\zeta_1 - 1/[\alpha(\rho_1+\rho_2)]\}\{1 - e^{-(\zeta_2-\zeta_1)t}\}/(\zeta_2-\zeta_1)]. \tag{5.9}$$

This expression may be used to calculate the instants τ_ε for any ε, cf. section 4. By using the inequalities $0 < 1-e^{-x} < x$, $x > 0$, it follows from (5.9) that

$$e^{-\zeta_1 t} < \psi_0(t) < e^{-\zeta_1 t}[1 + \{\zeta_1 - 1/[\alpha(\rho_1+\rho_2)]\}t], \quad \text{if } \zeta_1(\rho_1+\rho_2) > \frac{1}{\alpha},$$

$$\{\zeta_2 - 1/[\alpha(\rho_1+\rho_2)]\}(\zeta_2-\zeta_1)^{-1}e^{-\zeta_1 t} < \psi_0(t) < e^{-\zeta_1 t}, \quad \text{if } \zeta_1(\rho_1+\rho_2) < \frac{1}{\alpha}. \tag{5.10}$$

With the aid of (5.2) and (5.5) it is straight forward to prove that if $\zeta_1 < \zeta_2$,

$$\zeta_1(\rho_1+\rho_2) < \frac{2}{\alpha}. \tag{5.11}$$

From (5.10) and (5.11) it follows that in general, if $K=2$, $\zeta_1 \le \zeta_2$,

$$\psi_0(t) \le e^{-\zeta_1 t}[1 + \tfrac{1}{2}\zeta_1 t], \quad t \ge 0. \tag{5.12}$$

The right hand side of this inequality is equal to the function $\psi_0(t)$ for the network with $c_1=1$, $c_2=0$, $q_{12}=1$, $q_{21}=0$, $\beta_1=\beta_2=1/\zeta_1$, cf. (5.8). Hence, (5.12) implies that of all networks with two centers which have relaxation time $T = 1/\zeta_1$ the just mentioned tandem system reaches stochastic equilibrium the most slowly. Or, to put it in another way, of all $M/G/\infty$ systems for which the L.S. transform of the service time distribution is rational with a denominator of degree two and with $-\zeta_1$ as the pole closest to the imaginary axis, the $M/E_2/\infty$ system with mean service time $2/\zeta_1$ reaches steady-state in the slowest way. Note that we consider here τ_ε for systems with the same relaxation time T. If on the other hand we consider the above class of $M/G/\infty$ systems but with the same mean service time, then the relaxation time of the $M/E_2/\infty$ system is minimal.

It is not difficult to see that similar statements hold for exponential networks with K centers and for $M/G/\infty$ systems for which the L.S. transform of the service time distribution is rational with a denominator of degree K, $K > 2$. In the next section these properties will be used to derive an upper bound for the instants τ_ε which depends only on the number of nodes K and on the relaxation time T.

From (5.7) it follows further that for $m_1, m_2 = 0,1,2,..$, $\zeta_1 < \zeta_2$,

$$\psi_M(t) := 1 - E\{N_S(t)\,|M\}/E\{N_S(\infty)\} = e^{-\zeta_1 t}[1 - (m_1+m_2)/(\rho_1+\rho_2) +$$

$$+ \{1 - e^{-(\zeta_2-\zeta_1)t}\} \frac{\zeta_1(\rho_1+\rho_2-m_1-m_2) - 1/\alpha + m_1(1-q_{12})/\beta_1 + m_2(1-q_{21})/\beta_2}{(\zeta_2 - \zeta_1)(\rho_1 + \rho_2)}]. \tag{5.13}$$

Because $\psi_M(t)$ has not a fixed sign, it is difficult to give a good general bound for this function. By using the equality, cf. (5.5),

$$1/\alpha = \rho_1(1 - q_{12})/\beta_1 + \rho_2(1 - q_{21})/\beta_2, \tag{5.14}$$

the following inequality can be obtained from (5.13):

$$|\psi_M(t)| \le e^{-\zeta_1 t} \{|1 - \frac{m_1+m_2}{\rho_1+\rho_2}| + t|\frac{m_1-\rho_1}{\rho_1+\rho_2}||\zeta_1 - \frac{1-q_{12}}{\beta_1}| + t|\frac{m_2-\rho_2}{\rho_1+\rho_2}||\zeta_1 - \frac{1-q_{21}}{\beta_2}|\}. \tag{5.15}$$

This bound may be useful for m_1 and m_2 large with respect to ρ_1 and ρ_2. Because it lacks an interpretation like the bound (5.12) for $\psi_0(t)$ it is not clear how (5.15) can be generalized to the case of more than two service centers.

6. An upper bound for the instants τ_ε when the network is initially empty

Because the tandem system in which all jobs visit subsequently the centers 1 until K and in which all mean service times are equal reaches its equilibrium the most slowly of all exponential networks with K service centers and fixed relaxation time T, cf. section 5, we will determine the instants $\tau_\varepsilon = \tilde{\tau}_\varepsilon(K,\beta)$ for such tandem systems and use them as an upper bound for the instants τ_ε for general networks. For $c_1 = 1$, $\beta_j = \beta$ $(j=1,..,K)$, $q_{k,k+1} = 1$ $(k=1,..,K-1)$, $u_K = 1$, the differential equations (4.4) become

$$\frac{d}{dt}\phi_1(t) = -\frac{1}{\beta}\phi_1(t) + \frac{1}{\alpha}, \qquad \phi_1(0) = 0,$$

$$\frac{d}{dt}\phi_j(t) = -\frac{1}{\beta}\phi_j(t) + \frac{1}{\beta}\phi_{j-1}(t), \qquad \phi_j(0) = 0, \quad j = 2,..,K. \tag{6.1}$$

By induction it can be proved that the solution of this set of differential equations is given by: for $j = 1,..,K$, for $t \ge 0$,

$$\phi_j(t) = (\rho/K)[1 - e^{-t/\beta} \sum_{k=0}^{j-1} (t/\beta)^k/k!], \tag{6.2}$$

here $\rho = K\beta/\alpha$, cf. section 1. This implies that, cf. (4.1),(4.5), for $t \ge 0$,

$$\psi_0(t) = e^{-t/\beta} \sum_{j=0}^{K-1} (1-j/K)(t/\beta)^j/j! = 1 - \frac{t}{K\beta} - e^{-t/\beta} \sum_{j=K}^{\infty} (1-j/K)(t/\beta)^j/j!. \quad (6.3)$$

Note that $\psi_0(t)$ is strictly decreasing for $t > 0$ for every $K \geq 1$, because

$$\psi_0'(t) = \frac{-1}{K\beta} e^{-t/\beta} \sum_{j=0}^{K-1} (t/\beta)^j/j! < 0. \quad (6.4)$$

For $K = 1$ we have $\psi_0(t) = e^{-t/\beta}$, so that, cf. (4.1),

$$\tilde{\tau}_\varepsilon(1,\beta) = \beta \ln(100/\varepsilon). \quad (6.5)$$

From (6.3) it is seen that for every $t > 0$ as $K \to \infty$,

$$\psi_0(t) = 1 - \frac{1}{K} t/\beta + o(\frac{1}{K}), \quad (6.6)$$

so that

$$\tilde{\tau}_\varepsilon(K,\beta) \simeq (1 - \varepsilon/100)K\beta, \quad K \to \infty. \quad (6.7)$$

In table 6.1 the values of $\tilde{\tau}_\varepsilon(K,\beta)$ are listed as function of K for $\beta = 1$, $\varepsilon = 1,..,5$. These values are obtained by numerically solving the equations $\psi_0(t) = \varepsilon/100$, cf. (6.3), (4.1).

Let $\tau_\varepsilon = \tau_\varepsilon(K,T)$ be the instants defined by (4.1) for an exponential network with K service centers and relaxation time T, then, cf. section 5,

$$\tau_\varepsilon(K,T) \leq \tilde{\tau}_\varepsilon(K,T) = \tilde{\tau}_\varepsilon(K,1)\, T. \quad (6.8)$$

From (6.7) it is seen that the upper bound $\tilde{\tau}_\varepsilon(K,1)$ increases approximately linearly with K. There are, however, networks which reach their steady-state much faster (with respect to their relaxation time). As an example consider the symmetric network with $c_j = 1/K$, $\beta_j = \beta$, $q_{kj} = q/(K-1)$, $k \neq j$, $q < 1$, $q_{jj} = 0$ $(k,j = 1,..,K)$. For this network it follows from (4.4) that (here $\rho = \beta/[\alpha(1-q)]$)

$$\phi_j(t) = (\rho/K)\,[1 - e^{-(1-q)t/\beta}], \quad j = 1,..,K, \quad (6.9)$$

so that, independent of K,

$$\psi_0(t) = e^{-t/T}, \qquad \tau_\varepsilon(K,T) = T \ln(100/\varepsilon), \qquad T = \beta/(1-q). \quad (6.10)$$

In this case the upper bound $\tilde{\tau}_\varepsilon(K,1)T$, cf. (6.8), is very rude. But it seems to be

TABLE 6.1. The instants $\tilde{\tau}_\varepsilon(K,1)$ after which the mean number of jobs in an initially empty system with K M/M/∞ service centers in series stays within ε% of its limiting value

K	$\tilde{\tau}_1(K,1)$	$\tilde{\tau}_2(K,1)$	$\tilde{\tau}_3(K,1)$	$\tilde{\tau}_4(K,1)$	$\tilde{\tau}_5(K,1)$
1	4.61	3.91	3.51	3.22	3.00
2	5.99	5.19	4.72	4.38	4.11
3	7.30	6.41	5.87	5.49	5.19
4	8.55	7.58	7.00	6.57	6.23
5	9.77	8.73	8.09	7.63	7.26
10	15.6	14.2	13.4	12.8	12.3
25	32.0	29.9	28.6	27.6	26.8
50	58.3	55.3	53.4	51.9	50.7
100	109	105	102	100	98.1
150	160	155	151	148	145
200	210	204	199	196	193
250	260	253	248	244	240
300	310	302	296	291	287

difficult to find a better upper bound without getting involved into a detailed analysis of the networks. For instance, in the case K = 2, cf. (5.10), the condition $\zeta_1(\rho_1+\rho_2) \leq 1/\alpha$ is sufficient for $\tau_\varepsilon(2,T) \leq \tilde{\tau}_\varepsilon(1,1)T$ to hold, but it is not clear how this condition should be generalized to obtain lower upper bounds for specific networks with more than two service centers. An exception forms the class of feedforward networks in which jobs visit each center at most once. In such networks there are only a finite number of possible routes for the jobs and the length of each route is smaller than or equal to the size K of the network. Because the matrix (4.11) is triangular (possibly after renumbering the nodes), it follows readily that the relaxation time of a feedforward network is equal to

$$T = \max\{\beta_k;\ 1 \leq k \leq K\}. \tag{6.11}$$

Further, the above implies the following upper bound for feedforward networks, where $\kappa \leq K$ is the maximum number of service centers visited by any job:

$$\tau_\varepsilon(K,T) \leq \tilde{\tau}_\varepsilon(\kappa,1)T. \tag{6.12}$$

7. On networks with single server centers

In this section some analogies between networks with infinitely many servers and those with a single server at each center will be indicated. The notation of section 1 will be used, with deletion of indices referring to job classes (only the case of one job class will be considered). The discussion will be restricted to stable networks, which implies that for networks with single server centers it will be assumed, cf. [5], that $\rho_j < 1$, for $j = 1,..,K$, cf. (1.3).

For the M/M/1 queueing system it was shown in [4, §II.2.1] that as $t \to \infty$,

$$\psi_0(t) = \frac{1}{2}\pi^{-\frac{1}{2}}\rho^{-\frac{3}{4}}(1+\sqrt{\rho})\,(t/T)^{-3/2}e^{-t/T}[1+O(\tfrac{1}{t})], \tag{7.1}$$

where the relaxation time T is given by

$$T = \beta\,[1-\sqrt{\rho}]^{-2}. \tag{7.2}$$

By numerical solution of the Kolmogorov differential equations for the state probabilities it can be found that τ_ε/T for the M/M/1 system is smaller than $\tilde{\tau}_\varepsilon(1,1)$ for the M/M/∞ system; for instance $\tau_1/T = 3.1$ for $\rho = 0.1$ and $\tau_1/T = 2.1$ for $\rho = 0.9$, while $\tilde{\tau}_1(1,1) = 4.61$, cf. table 6.1. This can be explained by the factor $(t/T)^{-3/2}$ in (7.1), whereas for the M/M/∞ system we have $\psi_0(t) = e^{-t/T}$, cf. (6.3).

Recently, it was found in [1] that for two M/M/1 systems in series the function

$$\Pi_0(t) := \Pr\{N_S(t)=0\,|\,N_S(0)=0\}/\Pr\{N_S(\infty)=0\} - 1, \tag{7.3}$$

possesses the following asymptotic expansions as $t \to \infty$:

$$\Pi_0(t) = C(\rho_1,\rho_2)\,(\rho_1-\rho_2)^{-1}(t/T_1)^{-3/2}e^{-t/T_1}[1+O(\tfrac{1}{t})], \quad \text{if } \rho_2 < \rho_1,$$

$$\Pi_0(t) = C(\rho_2,\rho_1)\,(\rho_2-\rho_1)^{-1}(t/T_2)^{-3/2}e^{-t/T_2}[1+O(\tfrac{1}{t})], \quad \text{if } \rho_1 < \rho_2,$$

$$\Pi_0(t) = \frac{1}{2}\pi^{-\frac{1}{2}}\rho_1^{\frac{1}{4}}(1+\sqrt{\rho_1})^{-2}(t/T_1)^{-1/2}e^{-t/T_1}[1+O(\tfrac{1}{t})], \quad \text{if } \rho_1 = \rho_2, \tag{7.4}$$

here $C(\rho_1,\rho_2)$ is a bounded, non-vanishing constant, and

$$T_j = \beta_j\,[1-\sqrt{\rho_j}]^{-2}, \quad j = 1,2. \tag{7.5}$$

Hence, the relaxation time for this tandem system is equal to

$$T = \max\{T_j;\ j=1,2\} = \max\{\beta_j[1-\sqrt{\rho_j}]^{-2};\ j=1,2\}. \tag{7.6}$$

The function $\psi_0(t)$ has an asymptotic expansion similar to (7.4). Note that in the case $\rho_1 \neq \rho_2$ a factor $(t/T)^{-3/2}$ occurs as in (7.1), but that in the case $\rho_1 = \rho_2$ an additional factor t/T appears. Numerical evaluation of the instants τ_ε for this tandem system shows that the quotient τ_ε/T is larger than that for the ordinary M/M/1 system, but smaller than $\tilde{\tau}_\varepsilon(2.1)$ for the tandem system with infinitely many servers at the two nodes. In view of the present data for small networks it seems likely that in the worst case the function $\psi_0(t)$ for a Jackson network with K single server centers will have an asymptotic expansion of the form

$$\psi_0(t) = C\ (t/T)^{K-1}\ (t/T)^{-3/2}\ e^{-t/T}\ [1 + O(\tfrac{1}{t})], \quad t \to \infty, \tag{7.7}$$

and that the instants $\tau_\varepsilon = \tau_\varepsilon(K,T)$ for such a network are bounded by (see section 6 for the definition of $\tilde{\tau}_\varepsilon(K,1)$, see also table 6.1):

$$\tau_\varepsilon(K,T) \leq H\ T\ \tilde{\tau}_\varepsilon(K,1) \leq T\ \tilde{\tau}_\varepsilon(K,1); \tag{7.8}$$

here H is some constant between zero and one which takes into account the factor $(t/T)^{-3/2}$ for the queueing systems. In future research we hope to obtain general insights concerning the constant H as function of the parameters of the network. As in (6.8), cf. (6.12), $\tilde{\tau}_\varepsilon(K,1)$ may possibly be replaced by $\tilde{\tau}_\varepsilon(\kappa,1)$ in (7.8) in the case of feedforward networks, where κ is the maximum number of nodes visited by any job, $\kappa \leq K$.

We remark that in [2, §4.3] we discussed a conjecture on the relaxation time for Jackson networks with K single server centers, namely

$$T = \max\{\frac{|I-Q|_{jj}}{|I-Q|}\ \beta_j\ [1-\sqrt{\rho_j}]^{-2}\ ;\ j=1,..,K\}, \tag{7.9}$$

where I is the $K \times K$ identity matrix, $Q = (q_{kj})_{1\leq k,j\leq K}$, $|I-Q|$ 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 the j^{th} column.

In the near future we intend to check the validity of the inequality (7.8) and of the conjecture (7.9) by determining the function $\psi_0(t)$ numerically for small values of K. If (7.8) holds for networks with single server centers, then it will probably also hold for networks with an arbitrary number of servers at the various nodes, with a value of H closer to one, since their behaviour will be in between that of single server networks and that of infinite server networks.

Conclusion

An upper bound has been derived for the instants after which the mean number of jobs in an initially empty network with infinitely many servers at each node stays within ε% of its steady-state value. This upper bound is asymptotically a linear function of the size of the network. From small networks there is evidence that the same upper bound applies to networks with arbitrary (finite) numbers of servers at the various nodes. The results are useful for deciding when a network may be considered to be in equilibrium.

References

[1] Blanc, J.P.C. (1983) *The relaxation time of two queueing systems in series*, Report BW 192/83, Mathematical Centre, Amsterdam.

[2] Blanc, J.P.C. & E.A. van Doorn (1984) *Relaxation times for queueing systems*, to 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.

[3] Chiang, C.L. (1968) *Introduction to Stochastic Processes in Biostatistics*, Wiley, New York.

[4] Cohen, J.W. (1982) *The Single Server Queue*, North-Holland, Amsterdam, 2nd ed.

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

[6] Jackson, R.R.P. & P. Aspden (1980) *A transient solution to the multi-stage Poisson queueing system with infinite servers*, Oper. Res. 28, 618-622.

[7] Kleinrock, L. (1976) *Queueing Systems II: Computer Applications*, Wiley, New York.

[8] Neuts, M.F. (1981) *Matrix-geometric Solutions in Stochastic Models*, Hopkins Univ. Press, Baltimore.

[9] Takács, L. (1962) *Introduction to the Theory of Queues*, Oxford Univ. Press, New York.