WEAK CONVERGENCE LIMITS FOR SOJOURN TIMES IN CYCLIC QUEUES UNDER HEAVY TRAFFIC CONDITIONS

HANS DADUNA * ** AND
CHRISTIAN MALCHIN,* *Hamburg University*
RYSZARD SZEKLI,*** *Wrocław University*

Abstract

We consider sequences of closed cycles of exponential single-server nodes with a single bottleneck. We study the cycle time and the successive sojourn times of a customer when the population sizes go to infinity. Starting from old results on the mean cycle times under heavy traffic conditions, we prove a central limit theorem for the cycle time distribution. This result is then utilised to prove a weak convergence characteristic of the vector of a customer's successive sojourn times during a cycle for a sequence of networks with population sizes going to infinity. The limiting picture is a composition of a central limit theorem for the bottleneck node and an exponential limit for the unscaled sequences of sojourn times for the nonbottleneck nodes.

Keywords: Joint sojourn time distribution; cycle times; customer stationary behaviour; bottleneck node; central limit theorem

2000 Mathematics Subject Classification: Primary 60K25
Secondary 60F05

1. Introduction

We study the behaviour of cyclic networks of exponential single-server queues when a fixed number of nodes is filled with an ever increasing population. Such studies date back to the general model construction of Gordon and Newell [8]. There are essentially two different cases.

(i) All servers have the same load. In this case the total population in the system is shared equally by all nodes up to random fluctuations.

(ii) Differently loaded servers exist. In this case bottlenecks occur, which in the simplest case with exactly one slowest server means that almost the whole population is queued up at this slowest server.

In the cyclic case the load carried by the different nodes is proportional to their mean service times. This means that in case (i) (the case of balanced machines) all servers have the same service rate, and in case (ii) if there is exactly one server with smallest service rate then this node evolves as a unique bottleneck.

Our main interest is in the detailed travel-time behaviour of individual customers in case (ii). The starting point is the steady-state distribution of a customer's vector of successive sojourn times at the different nodes during a cycle. Furthermore, we are interested in the customer's cycle time distribution. These distributions are known, given in the transform domain by their respective Laplace–Stieltjes transforms (LSTs) [4], [18]. We transform these formulae in a way that allows us to prove weak convergence results for the customer's travel-time behaviour when the bottleneck dominates the travel times. This was studied as the *influence of the slowest server* in [3].

The *usual interpretation* of the results obtained by Gordon and Newell [8] is that with an increasing number of customers the bottleneck node approaches asymptotically a Poissonian source for the network, while all the other nodes eventually form an open ergodic tandem system, the behaviour of which is well understood: local geometrical queue length distribution and independence over the nodes in steady state. While this is generally understood as a statement about the queue length description of the cycle, it seems to be rather obvious that a similar property should hold for cycle times and their asymptotic behaviour, respectively, for the joint sojourn times of a customer during a cycle.

The influence of the slowest server [3] and Chow's [6] observation about the asymptotic distributional behaviour of cycle times in a two-stage cycle suggest that there should be a standard normal limit of the scaled cycle times; this is proved in Theorem 4.1. On the other hand, the usual interpretation suggests that even the unscaled sojourn times at the nonbottleneck nodes should converge in some sense to exponential distributions. This will come out from our main theorem, Theorem 5.1, as an immediate corollary and supports anew the usual interpretation.

Theorem 5.1 aggregates all sojourn times of a customer during a cycle, in the bottleneck and the nonbottleneck nodes. Owing to the above discussion, the limiting procedure should not be performed using a general scaling over all nodes. It turns out that, for the bottleneck node, a central limit scaling is appropriate, while, for the nonbottleneck nodes, no scaling is necessary.

This is different to the case of balanced machines (case (i)), where Kelly [11] proved that an overall law of large numbers scaling is the appropriate one to obtain distributional limits for sojourn times, which nevertheless are not normal.

Our approach in this paper is in line with the classical investigation of Gordon and Newell [8], where the bottleneck was considered to be exploding when the population increased, while the nonbottleneck nodes showed regular behaviour and were studied using conventional network processes. These results describe the networks' behaviour at a fixed time point in equilibrium and the networks' limiting distributions. More recent research in this direction has been carried out; see, e.g. [1] and the references therein.

We consider the exploding bottleneck and the regularly developing nodes in a joint picture, which is only possible if the behaviour of the bottleneck is described suitably rescaled. Because we are interested in sojourn and cycle times of customers, we have to observe the different networks not only at a fixed time instant or asymptotically, but in a suitable equilibrium over a (finite) random time horizon.

Owing to the normal limits which our results exhibit, our investigations are clearly related to the diffusion approximation. In the literature there are two paths in this area: the diffusion limit of the steady states for a sequence of networks via the central limit theorem and the diffusion limit of the sequence of (transient) processes via functional central limit theorems. An early survey on both problems is [13, Section 3], which dealt with closed systems of the repairman type: approximating the equilibrium distribution and the processes as a whole. More recent book sections on diffusion approximations for general closed networks via functional central

limit theorems are [5, Section 7.10] and [12, Section 6.2]; see also survey [19]. As will be seen, our results are related to both paths, because on the one hand we only consider systems in steady state, but on the other hand we observe the systems' behaviour over time.

But our approach is different from the standard diffusion approximations of networks. One important difference is that we do not have state space collapse, which means that there is convergence in distribution 'in heavy traffic (after normalization) to diffusion limits which "live" on a lower dimensional space than the original system' [17]. Such a state space collapse would happen in our systems if the nonbottleneck sojourn time sequences underwent a central limit theorem scaling. This can be concluded from the description in the cited references from the respective statements on actual workloads (which in the case of first-come–first-served (FCFS) is the actual waiting time of a customer), and will be obvious from our proofs.

Further related results on sojourn time approximations in simple closed queueing networks with processor sharing systems are proved by Mitra and Morrison; see [15] and the references therein. Diffusion approximation for sojourn times in open Jackson networks is developed in [16].

The paper is organized as follows. In Section 2 we describe the model and collect the necessary results on sojourn time distributions. In Section 3 we recall Boxma's result on the influence of the slowest server on the total cycle time and add on a somewhat curious observation about the behaviour of the joint vector of a customer's sojourn times given that his total cycle is known. This and the asymptotic moments of the cycle times and the local sojourn times (presented here) will be utilised in the proof of Theorem 5.1. In Section 4 we prove the central limit theorem for the overall cycle time when the population size grows unboundedly. The main problem in performing the proof is to suppress the strong dependencies in the summands which constitute the cycle times. The asymptotic behaviour of the joint vector of sojourn times is derived in Section 5.

2. Closed cycle with M stations

We consider a closed cyclic queueing network with M nodes (stations), where station $Q[i]$, $i = 1, \ldots, M$, is a single server with infinite waiting room under a FCFS queueing discipline. A fixed number $N \geq 1$ of identical customers circulate in the network. If the service of a customer at node $Q[i]$ terminates, he moves instantaneously to node $Q[i + 1]$ (he moves instantaneously to node $Q[1]$ if a service ends at node $Q[M]$). The customers at station $Q[i]$ (if there are any) move one position forward and the service of the customer at the head of the queue begins immediately. If there are further customers present at queue $Q[i]$ at the arrival instant of the moving customer, he joins the tail of the queue; if there are no further customers, his service begins immediately. Jumps and reorganisation of the queues take zero time. Service times of all customers at station $Q[i]$ are $\exp(\mu_i)$-distributed and are an independent family and independent of anything else. The joint queue length process of the cycle is a strong Markov process, which is irreducible and positive recurrent with steady-state distribution [8], given by

$$\pi^{(M,N)}(n_1, \ldots, n_M) = \frac{1}{G(M, N)} \prod_{i=1}^{M} \left(\frac{1}{\mu_i} \right)^{n_i} \tag{2.1}$$

with

$$(n_1, \ldots, n_M) \in Z(M, N) = \{(n_1, \ldots, n_M) \in \mathbb{N} : n_1 + \cdots + n_M = N\}$$

and normalising constant

$$G(M, N) := \sum_{n \in Z(M,N)} \prod_{i=1}^{M} \left(\frac{1}{\mu_i}\right)^{n_i}.$$

We fix some test customer (TC) and evaluate the asymptotic and stationary characteristics of the TC's sojourn times and cycle times. The starting instants of the TC's cycles are his successive entrance times into station $Q[1]$. The limiting distribution of the vector of the TC's successive sojourn times during his cycles is determined by the LST (see [4, Theorem 1]):

$$\phi^{(M,N)}(\theta_1, \ldots, \theta_M) = \sum_{n \in Z(M,N-1)} \pi^{(M,N-1)}(n) \prod_{j=1}^{M} \left(\frac{\mu_j}{\mu_j + \theta_j}\right)^{n_j+1}, \qquad 0 \le \theta_j < 1. \tag{2.2}$$

Here $\pi^{(M,N-1)}(n_1, \ldots, n_M)$ (given by (2.1)) is the steady-state probability that at the arrival instants of the TC at $Q[1]$ there are n_1 further customers present at node $Q[1]$ (without counting the TC himself), and n_j customers at nodes $Q[j]$ for $j \in \{2, \ldots, M\}$.

The LST of the limiting distribution of the TC's cycle time is (see [18])

$$\psi^{(M,N)}(\theta) = \sum_{n \in Z(M,N-1)} \pi^{(M,N-1)}(n) \prod_{j=1}^{M} \left(\frac{\mu_j}{\mu_j + \theta}\right)^{n_j+1}. \tag{2.3}$$

The sojourn and cycle time distributions, (2.2) and (2.3), are by definition limiting distributions. They can also be considered as stationary distributions under a customer stationary regime [7]. Our results can be interpreted in both situations. We henceforth fix the following notation for a cycle with M nodes and N customers, including the TC.

An M-dimensional vector with nonnegative real coordinates

$$(S_1^{(N)}, S_2^{(N)}, \ldots, S_M^{(N)}) \sim \phi^{(M,N)}(\theta_1, \ldots, \theta_M),$$

i.e. having distribution with LST $\phi^{(M,N)}(\theta_1, \ldots, \theta_M)$, denotes the TC's successive sojourn times during a cycle under customer stationary conditions, and a nonnegative real variable

$$S(N) \sim \psi^{(M,N)}(\theta),$$

i.e. having distribution with LST $\psi^{(M,N)}(\theta)$, denotes the TC's cycle time under customer stationary conditions.

It is well known that in the case that not all service rates μ_j are the same in networks with many customers cycling, bottlenecks will occur, i.e. nodes where almost all customers will queue up in the long run. The bottleneck nodes are the stations $Q[i]$ with $\mu_i = \min\{\mu_j : j = 1, \ldots, M\}$.

We shall assume throughout (unless otherwise stated) that $Q[1]$ is the only bottleneck station and, for technical reasons, that all service rates are distinct, i.e. $\mu_1 < \mu_2 < \cdots < \mu_M$. From the product-form steady state and sojourn time distributions, this ordering assumption does not reduce generality.

3. The influence of the slowest server

If station $Q[1]$ is the only bottleneck of the cycle, it is not surprising to the TC that in the case of a large population almost all other customers are waiting before him at $Q[1]$ when his cycle commences. Then, in particular, it follows (see [3]) that, even for $\mu_1 < \mu_2 \leq \cdots \leq \mu_M$,

$$\mathrm{E}(S(N)) = N\mu_1^{-1}\left(1 + O\left(\left(\frac{\mu_1}{\mu_2}\right)^N\right)\right),$$

$$\mathrm{var}(S(N)) = N\mu_1^{-2}\left(1 + O\left(\left(\frac{\mu_1}{\mu_2}\right)^N\right)\right), \qquad N \to \infty. \tag{3.1}$$

From (3.1), obviously in heavy traffic the slowest queue generates the main fraction of the cycle time of the TC. This clearly reflects the bottleneck behaviour with respect to the number of customers. So, it is reasonable to approximate the distribution of the cycle time for large values of N by the sum of N consecutive service times at the slowest queue. This tempting conjecture is supported by Chow's [6] observation that in a two-stage cycle a result parallel to (3.1) holds for the LST of the cycle times as well. Moreover, this suggests that there should hold some form of central limit theorem for the rescaled cycle time, when the number of customers tends to ∞, while the number of stations remains fixed. This will indeed be proved in Section 4.

Although the quantitative results in (3.1) are completely in line with intuition, there is still something behind the first- and second-order picture expressed there. This can be expressed as a conditional invariance property of the sojourn times; see [14] for a discussion. We state the result here without further comments because we shall need it for the derivations later on.

Theorem 3.1. (An invariance property for the conditional sojourn time distribution.) *For cycles with general service rates (with or without bottlenecks), the conditional distribution of the TC's successive sojourn times, given his cycle time, is independent of the number N of customers.*

For measurable and bounded $h \colon \mathbb{R}_+^M \to \mathbb{R}$, we have the conditional expectation

$$\mathrm{E}\big(h \circ \big(S_1^{(N)}, \ldots, S_M^{(N)}\big) \mid S_1^{(N)} + \cdots + S_M^{(N)} = t\big)$$

$$= \int_{\Delta_{M-1}^t} h\left(t - \sum_{j=2}^M x_j, x_2, \ldots, x_M\right)\mu_1 \exp(-\mu_1 t)$$

$$\times \prod_{j=2}^M \mu_j \exp(-x_j(\mu_j - \mu_1))\,\mathrm{d}(x_2, \ldots, x_M)$$

$$\times \left(\int_{\Delta_{M-1}^t} \mu_1 \exp(-\mu_1 t) \prod_{j=2}^M \mu_j \exp(-x_j(\mu_j - \mu_1))\,\mathrm{d}(x_2, \ldots, x_M)\right)^{-1},$$

where, for $M \in \mathbb{N}_+$ and $t \geq 0$, we define

$$\Delta_M^t = \{(x_1, \ldots, x_M) \in \mathbb{R}_+^M \colon x_1 + \cdots + x_M \leq t\}.$$

As preparation (which is of independent interest) for performing the analysis, we state the asymptotics of the moments' behaviour for cycle times (which follows directly from Boxma's results) and for local sojourn times (which is done by direct but tedious computations from (2.2)).

Lemma 3.1. *For pairwise distinct service rates, we have the following local asymptotics of moments.*

(i) *The mean cycle time and the cycle time variance:*

$$\lim_{N \to \infty} \frac{1}{N} \mathrm{E}(S(N)) = \frac{1}{\mu_1}, \qquad \lim_{N \to \infty} \frac{1}{N} \mathrm{var}(S(N)) = \frac{1}{\mu_1^2}.$$

(ii) *The bottleneck mean and the variance of the sojourn time:*

$$\lim_{N \to \infty} \frac{1}{N} \mathrm{E}\big(S_1^{(N)}\big) = \frac{1}{\mu_1}, \qquad \lim_{N \to \infty} \frac{1}{N} \mathrm{var}\big(S_1^{(N)}\big) = \frac{1}{\mu_1^2}.$$

(iii) *The nonbottleneck mean and the variance of the sojourn time:*

$$\lim_{N \to \infty} \mathrm{E}\big(S_k^{(N)}\big) = \frac{1}{\mu_k - \mu_1}, \qquad \lim_{N \to \infty} \mathrm{var}\big(S_k^{(N)}\big) = \frac{1}{(\mu_1 - \mu_k)^2}.$$

(iv) *The covariance of sojourn times at a bottleneck and some other station:*

$$\lim_{N \to \infty} \mathrm{cov}\big(S_1^{(N)}, S_l^{(N)}\big) = -\frac{1}{(\mu_1 - \mu_l)^2}.$$

(v) *The covariance of sojourn times at two nonbottleneck stations:*

$$\lim_{N \to \infty} \mathrm{cov}\big(S_k^{(N)}, S_l^{(N)}\big) = 0, \qquad k, l \neq 1.$$

A short comment may be in order here. From Lemma 3.1(iv) we see that the covariances between a customer's sojourn times at the bottleneck and the other nonbottleneck nodes do not vanish in the limit. In contrast to this, the usual interpretation, as described on page 334, states that in the limiting open tandem system the Poissonian source is independent of the service mechanism at the stations. By direct computations, it can be shown that results similar to Lemma 3.1(iv) and (v) hold for the joint queue length vector as well.

4. A central limit theorem for the cycle time

As suggested by the influence of the slowest server, we will prove in this section a central limit theorem for the TC's cycle time. The running index which tends to ∞ in this statement will be the number of customers in the cycle.

The main problem which we have to overcome are the strong dependencies of the TC's successive sojourn times during his cycle. Although the result of Boxma suggests that cycle times can be approximated by sums of independent $\exp(\mu_1)$ variables, the dependent sojourn times at the nonbottleneck nodes will come into play. To suppress these dependencies requires lengthy computations.

Theorem 4.1. *The sequence $(T(N))_{N \in \mathbb{N}}$ of the normalised and centred sojourn times*

$$T(N) := \frac{S(N) - \mathrm{E}(S(N))}{\sqrt{\mathrm{var}(S(N))}}$$

converges with an increasing number of customers weakly to the standard normal distribution $N(0, 1)$.

Before proving the theorem, we prepare some preliminary steps. For pairwise distinct service rates $\mu_i \neq \mu_j$ for $i, j \in \{1, \dots, M\}$ and $i \neq j$, we can write the normalising constant in closed form [9] as

$$G(M, N) = \sum_{i=1}^{M} \left(\frac{1}{\mu_i} \right)^N \prod_{\substack{j=1 \\ j \neq i}}^{M} \left(\frac{\mu_j}{\mu_j - \mu_i} \right). \tag{4.1}$$

We define coefficients $C_{i,N+1}$, $i \in \{1, \dots, M\}$ and $N \in \mathbb{N}_0$, related to $G(M, N)$ by

$$C_{i,N+1} := \left(\frac{1}{\mu_i} \right)^N \prod_{\substack{j=1 \\ j \neq i}}^{M} \frac{\mu_j}{\mu_j - \mu_i} \bigg/ \sum_{i=1}^{M} \left(\frac{1}{\mu_i} \right)^N \prod_{\substack{j=1 \\ j \neq i}}^{M} \frac{\mu_j}{\mu_j - \mu_i}. \tag{4.2}$$

Lemma 4.1. *For $\mu_1 < \mu_2 < \cdots < \mu_M$, the coefficients $C_{i,N}$ have the following properties:*

(i) $\sum_{i=1}^{M} C_{i,N} = 1$,

(ii) $\lim_{N \to \infty} C_{1,N} = 1$,

(iii) $\lim_{N \to \infty} N^r C_{k,N+1} = 0$, $k \neq 1$, $r \in \mathbb{R}_+$,

(iv) $C_{k,N} > 0$ *for k odd and $C_{k,N} < 0$ for k even.*

Proof. From $1 > \mu_1/\mu_2 > \mu_1/\mu_3 > \cdots > \mu_1/\mu_M > 0$ we directly conclude (ii):

$$\lim_{N \to \infty} C_{1,N+1} = \lim_{N \to \infty} \left(\prod_{\substack{j=1 \\ j \neq 1}}^{M} \frac{\mu_j}{\mu_j - \mu_1} \bigg/ \left(\prod_{j=2}^{M} \frac{\mu_j}{\mu_j - \mu_1} + \sum_{i=2}^{M} \left(\frac{\mu_1}{\mu_i} \right)^N \prod_{\substack{j=1 \\ j \neq i}}^{M} \frac{\mu_j}{\mu_j - \mu_i} \right) \right)$$

$$= 1.$$

In order to prove (iii), we take into account the fact that $\mu_1/\mu_k < 1$ for every $i = 2, \dots, M$, and write

$$N^r C_{k,N+1} = N^r \left(\frac{1}{\mu_k} \right)^N \prod_{\substack{j=1 \\ j \neq k}}^{M} \frac{\mu_j}{\mu_j - \mu_k} \bigg/ \sum_{i=1}^{M} \left(\frac{1}{\mu_i} \right)^N \prod_{\substack{j=1 \\ j \neq i}}^{M} \frac{\mu_j}{\mu_j - \mu_i}$$

$$= N^r \left(\frac{\mu_1}{\mu_k} \right)^N \prod_{\substack{j=1 \\ j \neq k}}^{M} \frac{\mu_j}{\mu_j - \mu_k} \bigg/ \left(\prod_{j=2}^{M} \frac{\mu_j}{\mu_j - \mu_1} + \sum_{i=2}^{M} \left(\frac{\mu_1}{\mu_i} \right)^N \prod_{\substack{j=1 \\ j \neq i}}^{M} \frac{\mu_j}{\mu_j - \mu_i} \right).$$

Finally, the alternating signs of $(1/\mu_i)^N \prod_{j=1, j \neq i}^{M} \mu_j/(\mu_j - \mu_i)$ prove property (iv) because, by (4.1), the denominator of the right-hand side of (4.2) is the normalising constant $G(M, N)$ and therefore positive.

From Harrison's formula [9] we obtain the following theorem.

Theorem 4.2. *The Laplace transform $\psi^{(M,N)}$ of the cycle time $S(N)$ under steady-state conditions can in the case of pairwise distinct service rates be written as*

$$\psi^{(M,N)}(\theta) = \sum_{i=1}^{M} C_{i,N} \left(\frac{\mu_i}{\mu_i + \theta} \right)^N. \tag{4.3}$$

Proof. To apply Harrison's formula, we have to ensure that $\mu_k + \theta \neq \mu_l + \theta$ for $k \neq l$ and $\theta \in [0, 1)$. Therefore,

$$\psi^{(M,N)}(\theta) = \frac{1}{G(M, N-1)} \sum_{i=1}^{M} \left(\prod_{\substack{j=1 \\ j \neq i}}^{M} \frac{\mu_j}{\mu_j - \mu_i} \right) \left(\frac{1}{\mu_i} \right)^{N-1} \left(\frac{\mu_i}{\mu_i + \theta} \right)^{N}$$

$$= \sum_{i=1}^{M} C_{i,N} \left(\frac{\mu_i}{\mu_i + \theta} \right)^{N}.$$

The Laplace transform $\psi^{(M,N)}$ in (4.3) is not a mixture distribution because some coefficients are negative, but, nevertheless, the nth moment of the cycle time can be expressed as

$$E(S(N)^n) = \left(\prod_{j=1}^{n} (N + j - 1) \right) \sum_{i=1}^{M} C_{i,N} \left(\frac{1}{\mu_i} \right)^{n}.$$

Hence, the cycle time for N customers circulating satisfies

$$E(S(N)) = \sum_{i=1}^{M} C_{i,N} N \left(\frac{1}{\mu_i} \right) \tag{4.4}$$

and

$$\text{var}(S(N)) = N \sum_{i=1}^{M} C_{i,N} \left(\frac{1}{\mu_i} \right)^{2} + N^2 \sum_{k<l} C_{k,N} C_{l,N} \left(\frac{1}{\mu_k} - \frac{1}{\mu_l} \right)^{2}.$$

For the cycle with M nodes fixed and N customers, we abbreviate the expectation of the cycle time and the square root of the variance by

$$\nu_N := E(S(N)) \quad \text{and} \quad \sigma_N := \sqrt{\text{var}(S(N))}.$$

We have now prepared the ground for the proof of Theorem 4.1

Proof of Theorem 4.1. The Fourier transform (FT) $\mathcal{F}_{S(N)}$ of $S(N)$ is

$$\mathcal{F}_{S(N)}(x) := \int_{\mathbb{R}} e^{ixy} \, dP^{S(N)} = \sum_{j=1}^{M} C_{j,N} \left(\frac{\mu_j}{\mu_j - ix} \right)^{N}, \qquad x \in \mathbb{R}.$$

Therefore, the FT $\mathcal{F}_{T(N)}$ of the normalised and centred sojourn time $T(N)$ is

$$\mathcal{F}_{T(N)}(x) = \exp\left(-\frac{ix\nu_N}{\sigma_N} \right) \mathcal{F}_{S(N)} \left(\frac{x}{\sigma_N} \right)$$

$$= \exp\left(-\frac{ix\nu_N}{\sigma_N} \right) \sum_{j=1}^{M} C_{j,N} \left(\frac{\mu_j}{\mu_j - ix/\sigma_N} \right)^{N}, \qquad x \in \mathbb{R}.$$

We show that

$$(i) \quad \lim_{N \to \infty} C_{1,N} \left(\exp\left(-\frac{ix}{\sqrt{N}} \frac{\nu_N/N}{\sigma_N/\sqrt{N}} \right) \bigg/ \left(1 - \frac{ix}{\sqrt{N}} \frac{1}{\mu_1 \sigma_N/\sqrt{N}} \right) \right)^{N} = \exp\left(-\frac{x^2}{2} \right),$$

(ii)
$$\lim_{N \to \infty} \exp\left(-\frac{\mathrm{i}x\nu_N}{\sigma_N}\right) \sum_{j=2}^{M} C_{j,N} \left(\frac{\mu_j}{\mu_j - \mathrm{i}x/\sigma_N}\right)^N = 0.$$

To prove (i), it suffices to show that

$$\left| \left(\exp\left(-\frac{\mathrm{i}x}{\sqrt{N}}\frac{\nu_N/N}{\sigma_N/\sqrt{N}}\right) \bigg/ \left(1 - \frac{\mathrm{i}x}{\sqrt{N}}\frac{1}{\mu_1 \sigma_N/\sqrt{N}}\right)\right)^N - \left(1 - \frac{x^2}{2N}\right)^N \right|$$
$$\longrightarrow 0 \quad \text{as } N \to \infty.$$

For every $x \in \mathbb{R}$ and sufficiently large N,

$$\left| \exp\left(-\frac{\mathrm{i}x}{\sqrt{N}}\frac{\nu_N/N}{\sigma_N/\sqrt{N}}\right) \bigg/ \left(1 - \frac{\mathrm{i}x}{\sqrt{N}}\frac{1}{\mu_1 \sigma_N/\sqrt{N}}\right)\right| = \left(1 + \frac{x^2}{N}\frac{1}{\mu_1^2 \sigma_N^2/N}\right)^{-1/2} \leq 1$$

and

$$\left| 1 - \frac{x^2}{2N}\right| \leq 1.$$

From $|u^n - v^n| \leq n\,|u - v|$ for $u, v \in \mathbb{C}$ with $|u| \leq 1$, $|v| \leq 1$, and $n \in \mathbb{N}$ [10, Lemma 4.13], we have

$$\left| \left(\exp\left(-\frac{\mathrm{i}x}{\sqrt{N}}\frac{\nu_N/N}{\sigma_N/\sqrt{N}}\right)\left(1 - \frac{\mathrm{i}x}{\sqrt{N}}\frac{1}{\mu_1 \sigma_N/\sqrt{N}}\right)^{-1}\right)^N - \left(1 - \frac{x^2}{2N}\right)^N \right|$$

$$\leq N\left| \exp\left(-\frac{\mathrm{i}x}{\sqrt{N}}\frac{\nu_N/N}{\sigma_N/\sqrt{N}}\right)\left(1 - \frac{\mathrm{i}x}{\sqrt{N}}\frac{1}{\mu_1 \sigma_N/\sqrt{N}}\right)^{-1} - \left(1 - \frac{x^2}{2N}\right)\right|$$

$$= N\left| \left(1 - \frac{\mathrm{i}x}{\sqrt{N}}\frac{\nu_N/N}{\sigma_N/\sqrt{N}} - \frac{x^2}{2N}\left(\frac{\nu_N/N}{\sigma_N/\sqrt{N}}\right)^2 + \sum_{k=3}^{\infty}\left(\frac{-\mathrm{i}x}{\sqrt{N}}\frac{\nu_N/N}{\sigma_N/\sqrt{N}}\right)^k \frac{1}{k!}\right) \right.$$

$$\left. \times \left(1 - \frac{\mathrm{i}x}{\sqrt{N}}\frac{1}{\mu_1 \sigma_N/\sqrt{N}}\right)^{-1} - 1 + \frac{x^2}{2N}\right|$$

$$= \left| \mathrm{i}x\frac{\sqrt{N}}{\mu_1 \sigma_N}\sqrt{N}\left(1 - \frac{\mu_1 \nu_N}{N}\right)\left(1 - \frac{\mathrm{i}x}{\sqrt{N}}\frac{1}{\mu_1 \sigma_N/\sqrt{N}}\right)^{-1} \right.$$

$$+ \frac{x^2}{2}\left(1 - \left(\frac{\nu_N/N}{\sigma_N/\sqrt{N}}\right)^2\left(1 - \frac{\mathrm{i}x}{\sqrt{N}}\frac{1}{\mu_1 \sigma_N/\sqrt{N}}\right)^{-1}\right)$$

$$\left. + N\sum_{k=3}^{\infty}\left(\frac{-\mathrm{i}x}{\sqrt{N}}\frac{\nu_N/N}{\sigma_N/\sqrt{N}}\right)^k \frac{1}{k!}\left(1 - \frac{\mathrm{i}x}{\sqrt{N}}\frac{1}{\mu_1 \sigma_N/\sqrt{N}}\right)^{-1}\right|$$

$$\longrightarrow 0 \quad \text{as } N \to \infty.$$

For the last step, we have used the facts that, by (4.4) and Lemma 4.1(i) and (iii),

$$\lim_{N \to \infty} \sqrt{N}\left(1 - \frac{\mu_1 \nu_N}{N}\right) = \lim_{N \to \infty} \sqrt{N}\sum_{j=2}^{M} C_{j,N}\left(1 - \frac{\mu_1}{\mu_j}\right) = 0,$$

and that, by Lemma 3.1(i) and [10, Lemma 4.14],

$$\lim_{N \to \infty}\left| N\sum_{k=3}^{\infty}\left(\frac{-\mathrm{i}x}{\sqrt{N}}\frac{\nu_N/N}{\sigma_N/\sqrt{N}}\right)^k \frac{1}{k!}\right| \leq \lim_{N \to \infty} \frac{1}{\sqrt{N}}\left|\frac{\nu_N/N}{\sigma_N/\sqrt{N}}\right|^3 \frac{|x|^3}{3!} = 0.$$

To prove (ii), note that

$$\left| \exp\left(-\frac{\mathrm{i}x\, \mathrm{E}(S_{(N)})}{\sqrt{\mathrm{var}(S_{(N)})}} \right) \sum_{j=2}^{M} C_{j,N} \left(\frac{\mu_j}{\mu_j - \mathrm{i}x/\sqrt{\mathrm{var}(S_{(N)})}} \right)^N \right|$$

$$= \left| \sum_{j=2}^{M} C_{j,N} \left(\frac{\mu_j}{\mu_j - \mathrm{i}x/\sqrt{\mathrm{var}(S_{(N)})}} \right)^N \right|$$

$$\leq \sum_{j=2}^{M} |C_{j,N}| \left| \frac{\mu_j}{\mu_j - \mathrm{i}x/\sqrt{\mathrm{var}(S_{(N)})}} \right|^N$$

$$= \sum_{j=2}^{M} |C_{j,N}| \left| \frac{1}{\sqrt{1 + x^2/\mu_j^2\, \mathrm{var}(S_{(N)})}} \right|^N$$

$$\leq \sum_{j=2}^{M} |C_{j,N}| \longrightarrow 0 \quad \text{as } N \to \infty.$$

5. Weak convergence limits for sojourn times

Theorem 5.1. *Consider a cycle of M nodes with N customers and with rates $\mu_1 < \cdots < \mu_M$. For the TC's sojourn times $(S_1^{(N)}, S_2^{(N)}, \ldots, S_M^{(N)})$, consider the partly rescaled sequence*

$$\widetilde{S(N)} = \left(\frac{S_1^{(N)} - \mathrm{E}(S_1^{(N)})}{\sqrt{\mathrm{var}(S_1^{(N)})}}, S_2^{(N)}, \ldots, S_M^{(N)} \right) \quad \text{as } N \to \infty.$$

The sequence $\widetilde{S(N)}$ converges for $N \to \infty$ to a random vector which is distributed as

$$\mathcal{N}(0, 1) \otimes \exp(\mu_2 - \mu_1) \otimes \cdots \otimes \exp(\mu_M - \mu_1).$$

Proof. We shall compute, for $\theta_j \in \mathbb{C}$, $j = 1, \ldots, M$,

$$\lim_{N \to \infty} \mathrm{E}\left(\exp\left(\mathrm{i}\theta_1 \frac{S_1^{(N)} - \mathrm{E}(S_1^{(N)})}{\sqrt{\mathrm{var}(S_1^{(N)})}} \right) \prod_{j=2}^{M} \exp\left(\mathrm{i}\theta_j S_j^{(N)} \right) \right) = \exp\left(-\frac{\theta_1^2}{2} \right) \prod_{j=2}^{J} \frac{\mu_j - \mu_1}{\mu_j - \mu_1 - \mathrm{i}\theta_j}.$$

Applying Theorem 3.1, we obtain (recall that $\Delta_k^t := \{(x_1, \ldots, x_k) \in \mathbb{R}_+^k \mid x_1 + \cdots + x_k \leq t\}$)

$$\mathrm{E}\left(\exp\left(\mathrm{i}\theta_1 \frac{S_1^{(N)} - \mathrm{E}(S_1^{(N)})}{\sqrt{\mathrm{var}(S_1^{(N)})}} \right) \prod_{j=2}^{M} \exp\left(\mathrm{i}\theta_j S_j^{(N)} \right) \right)$$

$$= \int_{\mathbb{R}} \mathrm{P}^{T(N)}(\mathrm{d}t)\, \mathrm{E}\left(\exp\left(\mathrm{i}\theta_1 \frac{S_1^{(N)} - \mathrm{E}(S_1^{(N)})}{\sqrt{\mathrm{var}(S_1^{(N)})}} \right) \prod_{j=2}^{M} \exp\left(\mathrm{i}\theta_j S_j^{(N)} \right) \,\bigg|\, T(N) = t \right)$$

$$= \int_{\mathbb{R}} P^{T(N)}(dt) \, \mathrm{E}\left(\exp\left(i\theta_1 \frac{S_1^{(N)} - \mathrm{E}(S_1^{(N)})}{\sqrt{\mathrm{var}(S_1^{(N)})}} \right) \prod_{j=2}^{M} \exp\left(i\theta_j S_j^{(N)} \right) \,\bigg|\, S(N) = t\sigma_N + v_N \right)$$

$$= \int_{\mathbb{R}} P^{T(N)}(dt)$$

$$\times \int_{\Delta_{M-1}^{t\sigma_N + v_N}} \exp\left(i\theta_1 \frac{t\sigma_N + v_N - \sum_{j=2}^{M} x_j - \mathrm{E}(S_1^{(N)})}{\sqrt{\mathrm{var}(S_1^{(N)})}} \right) \prod_{j=2}^{M} \exp\left(i\theta_j x_j \right)$$

$$\times \mu_1 \exp(-\mu_1(t\sigma_N + v_N)) \prod_{j=2}^{M} \mu_j \exp\left(-x_j(\mu_j - \mu_1) \right) d(x_2, \ldots, x_M)$$

$$\times \left(\int_{\Delta_{M-1}^{t\sigma_N + v_N}} \mu_1 \exp(-\mu_1(t\sigma_N + v_N)) \right.$$

$$\left. \times \prod_{j=2}^{M} \mu_j \exp\left(-x_j(\mu_j - \mu_1) \right) d(x_2, \ldots, x_M) \right)^{-1}$$

$$= \int_{\mathbb{R}} P^{T(N)}(dt)$$

$$\times \int_{\Delta_{M-1}^{t\sigma_N + v_N}} \exp\left(i\theta_1 \frac{t\sigma_N + v_N - \sum_{j=2}^{M} x_j - \mathrm{E}(S_1^{(N)})}{\sqrt{\mathrm{var}(S_1^{(N)})}} \right) \prod_{j=2}^{M} \exp\left(i\theta_j x_j \right)$$

$$\times \prod_{j=2}^{M} (\mu_j - \mu_1) \exp\left(-x_j(\mu_j - \mu_1) \right) d(x_2, \ldots, x_M)$$

$$\times \left(\int_{\Delta_{M-1}^{t\sigma_N + v_N}} \prod_{j=2}^{M} (\mu_j - \mu_1) \exp\left(-x_j(\mu_j - \mu_1) \right) d(x_2, \ldots, x_M) \right)^{-1}.$$

Let $(Y_2, \ldots, Y_M)$ denote a vector with nonnegative coordinates which are independent and let the Y_j be $\exp(\mu_j - \mu_1)$ distributed. Then the last expression can be written as

$$\int_{\mathbb{R}} P^{T(N)}(dt) \int_{\mathbb{R}_+^{M-1}} P^{(Y_2, \ldots, Y_M)} \, d(x_2, \ldots, x_M)$$

$$\times \mathbf{1}_{\Delta_{M-1}^{t\sigma_N + v_N}}(x_2, \ldots, x_M)(P((Y_2, \ldots, Y_M) \in \Delta_{M-1}^{t\sigma_N + v_N}))^{-1}$$

$$\times \exp\left(i\theta_1 \frac{t\sigma_N + v_N - \sum_{j=2}^{M} x_j - \mathrm{E}(S_1^{(N)})}{\sqrt{\mathrm{var}(S_1^{(N)})}} \right) \prod_{j=2}^{M} \exp\left(i\theta_j x_j \right).$$

For fixed $\theta_j \in \mathbb{C}$, $j = 1, \ldots, M$, we define the random variables

$$H(N) = H(N)(\theta_j : j = 1, \ldots, M) : \mathbb{R} \to \mathbb{C}$$

by

$$H(N)(t) = \exp\left(\frac{i\theta_1 t\sigma_N}{\sqrt{\text{var}(S_1^{(N)})}}\right) \exp\left(\frac{i\theta_1 (\nu_N - E(S_1^{(N)}))}{\sqrt{\text{var}(S_1^{(N)})}}\right)$$
$$\times \int_{\mathbb{R}_+^{M-1}} P^{(Y_2,\ldots,Y_M)} \, d(x_2, \ldots, x_M) \, \mathbf{1}_{\Delta_{M-1}^{t\sigma_N+\nu_N}}(x_2, \ldots, x_M)$$
$$\times \left(P((Y_2, \ldots, Y_M) \in \Delta_{M-1}^{t\sigma_N+\nu_N})\right)^{-1} \exp\left(\frac{i\theta_1 \sum_{j=2}^{M} x_j}{\sqrt{\text{var}(S_1^{(N)})}}\right) \prod_{j=2}^{M} \exp(i\theta_j x_j)$$

and

$$H = H(\theta_j : j = 1, \ldots, M) : \mathbb{R} \to \mathbb{C}$$

by

$$H(t) = \exp(i\theta_1 t) \int_{\mathbb{R}_+^{M-1}} P^{(Y_2,\ldots,Y_M)} \, d(x_2, \ldots, x_M) \prod_{j=2}^{M} \exp(i\theta_j x_j).$$

By direct inspection we see that the set of points $t \in \mathbb{R}$, such that there exists a sequence $(t_N \in \mathbb{R} : N \in \mathbb{N})$ with $(t_N \to t)$ such that $(H(N)(t_N) \to H(t))$ fails to hold, has under $N(0,1)$, probability 0. Here we used the facts that $\nu_N - E(S_1^{(N)}) \to 0$ and $\text{var}(S_1^{(N)}) \to \infty$ according to the influence of the slowest server.

Applying Theorem 5.5 of [2], we conclude from weak convergence of the cycle time distributions $P^{T(N)}$ to $N(0,1)$ weak convergence of the sequence of image measures $P^{T(N)} H(N)^{-1}$ to $N(0,1)H^{-1}$.

Finally, we define the bounded continuous function

$$f : \mathbb{C} \to \mathbb{C}, \qquad x \to \begin{cases} x & \text{if } |x| \leq 1, \\ \dfrac{x}{|x|} & \text{if } |x| > 1, \end{cases}$$

and conclude that

$$\int_{\mathbb{C}} P^{T(N)} H(N)^{-1}(dx) f(x) \to \int_{\mathbb{C}} N(0,1)H^{-1}(dx) f(x) \quad \text{for } N \to \infty.$$

But,

$$\int_{\mathbb{C}} P^{T(N)} H(N)^{-1}(dx) f(x)$$
$$= \int_{\mathbb{R}} P^{T(N)}(dt) f(H(N)(t))$$
$$= \int_{\mathbb{R}} P^{T(N)}(dt) \int_{\mathbb{R}_+^{M-1}} P^{(Y_2,\ldots,Y_M)} \, d(x_2, \ldots, x_M) \, \mathbf{1}_{\Delta_{M-1}^{t\sigma_N+\nu_N}}(x_2, \ldots, x_M)$$
$$\times (P((Y_2, \ldots, Y_M) \in \Delta_{M-1}^{t\sigma_N+\nu_N}))^{-1}$$
$$\times \exp\left(i\theta_1 \frac{t\sigma_N + \nu_N - \sum_{j=2}^{M} x_j - E(S_1^{(N)})}{\sqrt{\text{var}(S_1^{(N)})}}\right) \prod_{j=2}^{M} \exp(i\theta_j x_j)$$

and

$$\int_{\mathbb{C}} N(0,1)H^{-1}(\mathrm{d}x)f(x) = \int_{\mathbb{R}} N(0,1)(\mathrm{d}t)f(H(t))$$

$$= \int_{\mathbb{R}} N(0,1)(\mathrm{d}t)\exp(\mathrm{i}\theta_1 t)\int_{\mathbb{R}_+^{M-1}} \mathrm{P}^{(Y_2,\ldots,Y_M)}\,\mathrm{d}(x_2,\ldots,x_M)$$

$$\times \prod_{j=2}^{M} \exp\left(\mathrm{i}\theta_j x_j\right)$$

$$= \exp\left(-\frac{\theta_1^2}{2}\right)\prod_{j=2}^{J} \frac{\mu_j - \mu_1}{\mu_j - \mu_1 - \mathrm{i}\theta_j}.$$

This completes the proof.

Acknowledgements

References

[1] BERGER, A., BREGMAN, L. AND KOGAN, Y. (1999). Bottleneck analysis in multiclass closed queueing networks and its application. *Queueing Systems* **31,** 217–237.

[2] BILLINGSLEY, P. (1968). *Convergence of Probability Measures*. John Wiley, New York.

[3] BOXMA, O. J. (1988). Sojourn times in cyclic queues – the influence of the slowest server. *Comput. Performance Reliab.* 13–24.

[4] BOXMA, O. J., KELLY, F. P. AND KONHEIM, A. G. (1984). The product form for sojourn time distributions in cyclic exponential queues. *J. Assoc. Comput. Mach.* **31,** 128–133.

[5] CHEN, H. AND YAO, D. D. (2001). *Fundamentals of Queueing Networks*. Springer, New York.

[6] CHOW, W. M. (1980). The cycle time distribution of exponential cyclic queues. *J. Assoc. Comput. Mach.* **27,** 281–286.

[7] DADUNA, H. AND SZEKLI, R. (2002). Conditional job observer property for multitype closed queueing networks. *J. Appl. Prob.* **39,** 865–881.

[8] GORDON, W. J. AND NEWELL, G. F. (1967). Closed queueing networks with exponential servers. *Operat. Res.* **15,** 254–265.

[9] HARRISON, P. G. (1985). On normalizing constants in queueing networks. *Operat. Res.* **33,** 464–468.

[10] KALLENBERG, O. (2002). *Foundations of Modern Probability*, 2nd edn. Springer, New York.

[11] KELLY, F. P. (1984). The dependence of sojourn times in closed queueing networks. In *Mathematical Computer Performance and Reliability*, eds G. Iazeolla *et al.*, North-Holland, Amsterdam, pp. 111–121.

[12] KUSHNER, H. J. (2001). *Heavy Traffic Analysis of Controlled Queueing Networks and Communication Networks*. Springer, New York.

[13] LEMOINE, A. J. (1978). Networks of queues – a survey of weak convergence results. *Manag. Sci.* **24,** 1175–1193.

[14] MALCHIN, C. AND DADUNA, H. (2005). An invariance property of sojourn times in cyclic networks. *Operat. Res. Lett.* **33,** 1–8.

[15] MORRISON, J. A. (1987). Conditioned response-time distribution for a large closed processor-sharing system in very heavy usage. *SIAM J. Appl. Math.* **47,** 1117–1129.

[16] REIMAN, M. I. (1982). The heavy traffic diffusion approximation for sojourn times in Jackson networks. In *Applied Probability – Computer Sciences: The Interface*, eds R. L. Disney and T. J. Ott, Vol. II, Birkhäuser, Boston, MA, pp. 409–421.

[17] REIMAN, M. I. (1984). Some diffusion approximations with state space collapse. In *Proc. Internat. Sem. Modeling Performance Eval. Methodol.*, eds F. Baccelli and G. Fayolle, Springer, New York, pp. 20–240.

[18] SCHASSBERGER, R. AND DADUNA, H. (1983). The time for a roundtrip in a cycle of exponential queues. *J. Assoc. Comput. Mach.* **30,** 146–150.

[19] WILLIAMS, R. J. (1996). On the approximation of queueing networks in heavy traffic. In *Stochastic Networks – Theory and Applications*, eds F. P. Kelly *et al.* , Clarendon Press, Oxford, pp. 35–56.