U N I V E R S I T Ä T   H A M B U R G

Weak Convergence Limits for Closed Cyclic Networks of Queues with Multiple Bottleneck Nodes

O. Stenzel, H. Daduna

Preprint–No. 2009-05   Dezember 2009

DEPARTMENT MATHEMATIK

MATHEMATISCHE STATISTIK
UND STOCHASTISCHE PROZESSE

BUNDESSTR. 55, D – 20146 HAMBURG

Weak Convergence Limits for Closed Cyclic Networks of Queues with Multiple Bottleneck Nodes

Ole Stenzel, Hans Daduna
Center of Mathematical Statistics and Stochastic Processes
University of Hamburg

December 24, 2009

Abstract: We consider a sequence of cycles of exponential single server nodes, where the number of nodes is fixed and the number of customers grows unboundedly. We prove a central limit theorem for the cycle times distribution. We investigate the idle time structure of the bottleneck nodes and the joint sojourn time distribution that a test customer observes at the non-bottleneck nodes during a cycle. Furthermore, we study the filling behaviour of the bottleneck nodes and show that there is a different asymptotic depending on having a single bottleneck or having multiple bottlenecks.

Keywords: Cyclic exponential network, cycle times, idle times, filling behaviour, bottleneck structure, central limit theorem.

Introduction

The aim of this paper is to analyze the behaviour of cyclic Gordon-Newell networks in equilibrium with single-server nodes as the number of customers in the system increases to infinity. There are two different cases to consider:
(i) All nodes have the same service rate. Then the customers are uniformly distributed over the nodes.
(ii) Different service rates exist. Then at least one bottleneck (node with the smallest service rate) exists. Almost all the customers will be queued up at the bottleneck nodes.
We will focus on the second case where bottlenecks occur. We are particularly interested in the distribution of a customer's cycle time for which a central limit theorem will be proved. Furthermore, we are interested in the differences of the filling behaviour of the nodes, in case of a single bottleneck and in case of multiple bottlenecks.

It will turn out, that an important aspect of the networks' behaviour is characterized by the mean idle time of the bottleneck nodes during the cycle of a test customer, whose cycle time distribution is investigated.

The paper is organized as follows: In Section 1, the model (a closed cyclic queueing network) will be described and some properties needed in the sequel will be referenced . In Section 2, we compute a test customer's cycle time distribution and discuss the intimate connection of this distribution to the mean idle times of the bottlenecks. This connection originates from the description of the cycle time from the viewpoint of a node: For population size N the time between N departures from some specified node is the typical cycle time of the first of the N departing customers which are observed [Box88].

Starting with the limiting distribution of the cycle time, we will prove a central limit theorem in the general setting of this paper. This result generalizes the central limit theorem for the cycle time distribution

when the number of nodes is fixed and all service rates are distinct [DMS08]. In Section 3 we analyze differences in the filling behaviour of the network in case of a single bottleneck and in case of multiple bottlenecks. We start from the observation that there is a fundamental difference with respect to the mean idle times between these two cases:
Considering the possible service rate vectors of the cyclic networks as parameter space the observed jump in the mean idle times constitutes a phase transition, when reaching a certain boundary region of the parameter space.

Motivated by this, we analyze in detail the speed with which the bottleneck nodes are filled up. It turns out that there is a fundamental difference in the speed of convergence between the cases of single and multiple bottlenecks.

In Section 4, we will prove a weak convergence theorem for the joint steady-state queue length distribution of the non-bottleneck nodes.

1 Model specification and previous results

We consider a closed cyclic queueing network consisting of M nodes $Q[1], \ldots, Q[M]$ and N indistinguishable circulating customers. Node $Q[i]$, $i = 1, \ldots, M$, is a single-server with infinite waiting room, which is organized according to FCFS (First Come, First Served) regime. The dynamic of the system is as follows: If a customer arrives at node $Q[i]$, $i = 1, \ldots, M$, and finds an idle service channel, i.e. no other customers are present at that node, his service begins immediately. Otherwise, he will join the tail of the queue. Service times requested by the customers at node $Q[i]$ are $\exp(\mu_i)$-distributed. If the service of the customer at node $Q[i]$ is finished, he will move instantenously to node $Q[i+1]$ (with $Q[M+1] := Q[1]$). The other customers waiting in the line (if there are any) move one position forward. All jumps of the customers (in the queue or to other nodes) happen without any time lag. The service times at node $Q[i]$ form a sequence of *iid* random variables which is independent of the service times at other nodes.

Denote by $X_i^{(M,N)}(t)$ the queue length at node $Q[i]$, i.e. the number of customers present at node $Q[i]$ (waiting or in service), at time $t \geq 0$. We define $X^{(M,N)}(t) := (X_1^{(M,N)}(t), \ldots, X_M^{(M,N)}(t))$ as the joint queue length vector at time t and $X^{(M,N)} := \left(X^{(M,N)}(t),\ t \geq 0\right)$ as the joint queue length process. The state space of $X^{(M,N)}$ is $Z(M,N) := \{(n_1, \ldots, n_M) \in \mathbb{N}^M \mid n_1 + \ldots + n_M = N\}$.

For more information on Gordon-Newell networks, see [Rob03][Section 4.4.].

Proposition 1.1 *The joint queue length process $X^{(M,N)} = (X^{(M,N)}(t),\ t \geq 0)$ described above is a strong Markov process, which is irreducible and positive recurrent. The limiting and stationary distribution is*

$$\pi^{(M,N)}(n) = G(M,N)^{-1} \prod_{i=1}^{M} \left(\frac{1}{\mu_i}\right)^{n_i} \tag{1.1}$$

with $n = (n_1, \ldots, n_M) \in Z(M,N)$ and normalising constant

$$G(M,N) = \sum_{n \in Z(M,N)} \prod_{i=1}^{M} \left(\frac{1}{\mu_i}\right)^{n_i}. \tag{1.2}$$

Assumption. In the following we will assume that the joint queue length process is in equilibrium. We will therefore omit the time parameter t, i.e., we will write $X^{(M,N)}$ instead of $X^{(M,N)}(t)$.

We want to investigate individual behaviour of customers in the cycle. Therefore in the following we fix one customer, called TC (the Test Customer). The random time between two entrances of the TC into node $Q[1]$ is called the *cycle time.* Formally, we get a sequence of cycle times $(c_N^{(i)},\ i \in \mathbb{N}_+)$ with $c_N^{(i)}$ being the i-th passage time through the cycle.

Theorem 1.2 (Limiting distribution of the cycle time and sojourn time vector) *The limiting distribution of the TC's cycle time is given by its the Laplace-Stieltjes Transform (LST)*

$$\psi^{(M,N)}(\theta) = \sum_{n \in Z(M,N-1)} \pi^{(M,N-1)}(n) \prod_{j=1}^{M} \left(\frac{\mu_j}{\mu_j + \theta} \right)^{n_j+1}, \quad \theta \geq 0, \tag{1.3}$$

where $\pi^{(M,N-1)}(n_1, \ldots, n_M)$ *(defined by* (1.1)*) is the steady-state probability that at the arrival instants of the TC at node* $Q[1]$*, he observes* n_1 *customers before him at node* $Q[1]$ *and* n_i *customers present at node* $Q[i]$*,* $i = 2, \ldots, M$*, without counting himself.*

The limiting joint distribution of the TC's successive sojourn times during a cycle is given by its the Laplace-Stieltjes Transform (LST)

$$\phi^{(M,N)}(\theta_1, \ldots, \theta_J) = \sum_{n \in Z(M,N-1)} \pi^{(M,N-1)}(n) \prod_{j=1}^{M} \left(\frac{\mu_j}{\mu_j + \theta_j} \right)^{n_j+1}, \quad \theta_j \geq 0,\ j = 1, \ldots, J. \tag{1.4}$$

A proof can be found in [BKK84].

Comment.

(i) Note that the limiting distribution of the TC's cycle time is a mixture of convolutions of Gamma distributions:

$$\mathcal{L}\left(c_N\right) = \sum_{n \in Z(M,N-1)} \pi^{(M,N-1)}(n) \left(*_{j=1}^{M} \Gamma_{\mu_j, n_j+1} \right). \tag{1.5}$$

We see that for large values of N the limiting distribution is rather complicated. Therefore, asymptotic expansions would be of value. E.g., we shall prove a central limit theorem for the cycle time.

(ii) Note that the steady state distribution of the joint queue length process (1.1) and the limiting distribution of the cycle time are not influenced by the ordering of the nodes.

Notation. Denote by c_N a real valued non-negative random variable with LST given by (1.3), i.e.,

$$c_N \sim \psi^{(M,N)}.$$

From now on, when recalling a cycle time, we always refer to a random variable c_N with LST $\psi^{(M,N)}$.
In any case the number M of nodes is fixed.
Similarly, an M-dimensional vector with non negative 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 $\phi^{(M,N)}(\theta_1, \ldots, \theta_M)$ is considered as the vector of TC's successive sojourn times during a cycle under customer stationary conditions.

Definition 1.3 *The node $Q[i]$, $i = 1, \ldots, M$, is said to be a bottleneck node if $Q[i]$ has the slowest service rate, i.e., $\mu_i = \min\{\mu_k,\ k = 1, \ldots, M\}$.*

Assumption. In order to keep the notation and the computations as simple as possible, we will always assume that without loss of generality $\mu_1 \leq \ldots \leq \mu_M$. This assumption is made only to keep the calculations simple.

Notation. Denote m the number of different service rates, i.e. the number of different values of $\mu_1 \leq \ldots \leq \mu_M$. The distinct service rates will be denoted by $\eta_1 < \eta_2 < \ldots < \eta_m$ and let ν_l, $1 \leq l \leq m$, be the number of $i \in \{1, \ldots, M\}$ with $\mu_i = \eta_l$. (Note that ν_1 is the number of bottleneck nodes and that $\mu_1 = \eta_1$ and $\mu_M = \eta_m$.)

It was already noticed by Gordon and Newell [GN67] that the bottlenecks have an overwhelming influence on the asymptotic and steady state behaviour of a closed network, whenever the number of customers is high compared to the number of nodes. A precise meaning of this statement was proved [GN67] in case of node 1 being the single bottleneck of the network when M is constant and $N \to \infty$. Gordon and Newell proved

$$\lim_{N\to\infty} P(X_1^{(M,N)} \geq n) = 1, \quad \forall n \in \mathbb{N}, \tag{1.6}$$

$$\lim_{N\to\infty} P(X_2^{(M,N)} = n_2, \ldots, X_M^{(M,N)} = n_M) = \prod_{j=2}^{M} \left(1 - \frac{1}{\mu_j}\right)\left(\frac{1}{\mu_j}\right)^{n_j}, \ \forall\, n_j \geq 0,\ j = 2, \ldots, M. \tag{1.7}$$

The *usual interpretation* of the results obtained by Gordon and Newell [GN67] is in case of a cyclic network that with increasing number of customers the bottleneck node approaches asymptotically a Poissonian source feeding the rest of 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. It is rather obvious that a similar property should follow for the non-bottleneck nodes in the one-bottleneck case from (1.3), resp., (1.4), for the sojourn times and partial cycle times and their asymptotic behaviour, see Theorem 5.1 in [DMS08].

On the other hand it is a tempting conjecture (but not obvious) that a similar interpretation should be available for the case of several bottlenecks which are distributed over the cycle and divide the cycle into bottlenecks and (possibly empty) sequences of non-bottleneck nodes between them. This will be proved below in Section 4.

If $Q[1]$ is the only bottleneck ($\mu_1 < \mu_2 \leq \cdots \leq \mu_M$) it is not surprising to TC that in case of large population almost all other customers are waiting before him at $Q[1]$ when his cycle commences. Then in particular it follows ([Box88]) that

$$E\big(c_N\big) = N\mu_1^{-1}\big\{1 + O\big([\frac{\mu_1}{\mu_2}]^N\big)\big\}, \quad \mathrm{Var}\big(c_N\big) = N\mu_1^{-2}\big\{1 + O\big([\frac{\mu_1}{\mu_2}]^N\big)\big\}, \quad N \to \infty\ . \tag{1.8}$$

From (1.8) obviously in heavy traffic the slowest queue generates the main fraction of the cycle time of 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 observation that in a two-stage cycle a result parallel to (1.8) holds for the LST of the cycle times as well [Cho80]. This suggests that there should hold a central limit theorem for the rescaled cycle time, when the number of customers tends to

infinity, while the number of stations remains fixed:
In case of a single bottleneck $Q[1]$, the TC finds almost all other customers waiting before him at the bottleneck node, which in precise terms is

$$\lim_{N\to\infty} \frac{E\left[X_1^{(M,N)}\right]}{N} = 1.$$

As a result, the TC's cycle time is mainly the time needed to serve all customers before him at the bottleneck. These times are i.i.d., with finite mean and variance.

The central limit theorem for the rescaled cycle time was proved by Daduna, Malchin and Szekli [DMS08][Theorem 4.1] for the case of pairwise distinct service rates ($\mu_1 < \ldots < \mu_M$), i.e., in case of a single bottleneck. Furthermore, they proved a convergence property of the joint sojourn times vectors, with the bottleneck under central limit scaling, the non-bottlenecks without scaling.

On the other hand the *usual interpretation* suggests that even the unscaled sojourn times at the non-bottleneck nodes should converge in some sense to exponential distributions. This will be proved in Theorem 4.1 and supports anew the *usual interpretation*.

We shall investigate in the following the similar problems without restrictions on the number of bottlenecks. It turns out, that the proofs are much more involved.

2 Cycle times and idle times of bottlenecks

Our approach resumes the description of cycle times given by Boxma [Box88][p. 19.]

The cycle time is defined as the random time between two successive entrances of TC into $Q[1]$. In the following, we analyze the cycle time from the viewpoint of a server: We start the observation of the server $Q[1]$ at a time instant, when TC leaves $Q[1]$. We denote by i_1 the random idle time (which may be zero) until the next customer starts his service at node $Q[1]$. Afterwards, we denote by τ_1 the random service time of the next customer in service. When this customers leaves $Q[1]$, we denote by i_2 the next idle time (which may be zero again) of the service channel and so on. By and by, the customers pass $Q[1]$ and we denote the idle times by i_j and the service times by τ_j, $j = 1, \ldots, N$. (Idle times are zero if there are customers waiting in the queue of $Q[1]$.)
After TC leaves $Q[1]$ again, we sum up the service times and the idle times to get the cycle time. Therefore, the cycle time can be expressed by

$$c_N = i_1 + \ldots + i_N + \tau_1 + \ldots + \tau_N \tag{2.1}$$

Defining $\delta_N := \tau_1 + \ldots + \tau_N$ as the sum of all service times and $\rho_N := i_1 + \ldots + i_N$ as the cumulative idle time, we have

$$c_N = \rho_N + \delta_N \tag{2.2}$$

where δ_N is the sum of N *iid* $\exp(\mu_1)$-distributed service times.

A first simple observation yields a bound for $E[\rho_N]$ which is independent on N. We have

$$\begin{aligned} E[c_N] &\overset{(1.5)}{=} \sum_{n\in Z(M,N-1)} \pi^{(M,N-1)}(n) \sum_{j=1}^{M} (n_j+1) \underbrace{\mu_j^{-1}}_{\le \mu_1^{-1}} \le \sum_{n\in Z(M,N-1)} \pi^{(M,N-1)}(n) \mu_1^{-1} \underbrace{\sum_{j=1}^{M} (n_j+1)}_{=N+M-1} \\ &= (N+M-1)\mu_1^{-1} \sum_{n\in Z(M,N-1)} \pi^{(M,N-1)}(n) = (N+M-1)\mu_1^{-1}. \end{aligned}$$

Since $E[\rho_N] = E[c_N] - E[\delta_N] = E[c_N] - N\mu_1^{-1}$, it follows

$$E[\rho_N] \le (M-1)\mu_1^{-1}. \tag{2.3}$$

Boxma [Box88] observed that (in his setting with exactly one bottleneck) holds

$$E[\rho_N] \to 0, \qquad n \to \infty.$$

The observation (2.3) is of importance because the cycle time consists of two components: δ_N, the sum of N *iid* $\exp(\mu_1)$-distributed random service times and the cumulative idle time ρ_N, whose expected values remain bounded, whereares the expected values of δ_N grow to infinity as $N \to \infty$. One would therefore assume that for large values of N, the influence of ρ_N on the cycle time is not significant compared to the influence of δ_N. This suggests

$$\frac{c_N - E[c_N]}{\sqrt{Var(c_N)}} \approx \frac{\delta_N - E[\delta_N]}{\sqrt{Var(\delta_N)}}.$$

For the right side holds an elementary central limit theorem. The main problem will be to sharpen and to extend (2.3).

2.1 Idle times at the bottlenecks

Recall that node $Q[1]$ is always a bottleneck, and that $\nu_1 \ge 1$ is the number of bottlenecks. For the expected cumulative idle times at $Q[1]$ during TC's cycle in steady state regime we have a precise asymptotic. This is obtained from moment properties of the cycle time, which will be used to prove the central limit theorem.

Theorem 2.1 *It holds*

$$\lim_{N\to\infty} \left(E[c_N] - \mu_1^{-1}N\right) = \mu_1^{-1}(\nu_1 - 1). \tag{2.4}$$

Theorem 2.2 *It holds*

$$\lim_{N\to\infty} \left(Var(c_N) - \mu_1^{-2}N\right) = \mu_1^{-2}(\nu_1 - 1). \tag{2.5}$$

An immediate corollary is now

Corollary 2.3 *It holds*

$$\lim_{N\to\infty} E[\rho_N] = \mu_1^{-1}(\nu_1 - 1). \tag{2.6}$$

and

$$\lim_{N\to\infty} \left(Var(\rho_N) + Cov(\rho_N, \delta_N)\right) = \mu_1^{-2}(\nu_1 - 1). \tag{2.7}$$

While (2.7) does not admit a direct interpretation, (2.6) has a surprising interpretation.
Assume that we have a single bottleneck which is $Q[1]$. Then

$$\lim_{N\to\infty} E[\rho_N] = 0.$$

Assume further, that for nodes $Q[2], Q[3], \ldots, Q[k]$, with $k < M$ we let converge $\mu_j \to \mu_1, j = 2, \ldots, k$, then we have a discontinuity of the asymptotic expected idle times in this limiting procedure. Saying it in the simplest setting the other way round:
Consider our sequence of networks to be dependent on the parameter vector $(\mu_1, \ldots, \mu_M)$, and assume that the parameters vary in a way that we start with a single bottleneck $Q[1]$. Consider the function which maps

any sequence of networks to the value $\lim_{N\to\infty} E[\rho_N]$. This function starts at zero and stays there until at least one other μ_j reaches $\mu_j = \mu_1$. Then the mean value function $E[\rho_N]$ immediately jumps to a value $\geq \mu_1^{-1} > 0$, a phase transition occurs, when the parameter vector reaches the boundary region defined by $\{\mu_1 = \mu_2\}$.

We shall discuss this behaviour in more detail in Section 3.

The proofs of Theorem 2.1 and 2.2 need several preparatory steps which will be given now.

The following interchange formula for sums and products is a direct consequence of the most general form of Harrison's formula [Har85].

Lemma 2.4

$$\sum_{n\in Z(M,N)} \prod_{j=1}^{M} \left(\frac{1}{\mu_j}\right)^{n_j} = \sum_{l=1}^{m} \sum_{a\in Z(m,\nu_l-1)} (-1)^{a_l-\nu_l+1} \left(\frac{1}{\eta_l}\right)^{N-\nu_l+a_l+1} \binom{N+a_l}{N} \cdot \left(\prod_{u=1,u\neq l}^{m} \eta_u^{\nu_u} \binom{\nu_u+a_u-1}{\nu_u-1} \left(\frac{1}{\eta_u-\eta_l}\right)^{\nu_u+a_u}\right). \tag{2.8}$$

Proof.

$$\begin{aligned}
&\sum_{n\in Z(M,N)} \prod_{j=1}^{M} \left(\frac{1}{\mu_j}\right)^{n_j} \\
&\overset{(1)}{=} \sum_{l=1}^{m} \sum_{a\in Z(m,\nu_l-1)} (-1)^{a_l-\nu_l+1} \left(\frac{1}{\eta_l}\right)^{N+M-\nu_l} \binom{N+a_l}{N} \left(\prod_{u=1,\ u\neq l}^{m} \eta_u^{-a_u} \binom{\nu_u+a_u-1}{\nu_u-1} \left(\frac{1}{\eta_l}-\frac{1}{\eta_u}\right)^{-\nu_u-a_u}\right) \\
&= \sum_{l=1}^{m} \sum_{a\in Z(m,\nu_l-1)} (-1)^{a_l-\nu_l+1} \left(\frac{1}{\eta_l}\right)^{N+M-\nu_l} \binom{N+a_l}{N} \left(\prod_{u=1,\ u\neq l}^{m} \eta_u^{-a_u} \binom{\nu_u+a_u-1}{\nu_u-1} \left(\frac{\eta_l\eta_u}{\eta_u-\eta_l}\right)^{\nu_u+a_u}\right) \\
&= \sum_{l=1}^{m} \sum_{a\in Z(m,\nu_l-1)} (-1)^{a_l-\nu_l+1} \left(\frac{1}{\eta_l}\right)^{N+M-\nu_l} \binom{N+a_l}{N} \eta_l^{\sum_{u\neq l}(\nu_u+a_u)} \\
&\qquad \cdot \left(\prod_{u=1,\ u\neq l}^{m} \eta_u^{\nu_u} \binom{\nu_u+a_u-1}{\nu_u-1} \left(\frac{1}{\eta_u-\eta_l}\right)^{\nu_u+a_u}\right) \\
&= \sum_{l=1}^{m} \sum_{a\in Z(m,\nu_l-1)} (-1)^{a_l-\nu_l+1} \left(\frac{1}{\eta_l}\right)^{N-\nu_l+a_l+1} \binom{N+a_l}{N} \left(\prod_{u=1,\ u\neq l}^{m} \eta_u^{\nu_u} \binom{\nu_u+a_u-1}{\nu_u-1} \left(\frac{1}{\eta_u-\eta_l}\right)^{\nu_u+a_u}\right)
\end{aligned}$$

$\overset{(1)}{=}$ is the generalization of Harrison's formula, see [Ser99][Proposition 1.32]. Note that in the expression given there, typos occur, see [Mal08][Equation (1.19)]. □

The following constants will help to keep the computations with sometimes lengthy expressions shorter.

Definition 2.5 *For the general cycle we set*

$$(i) \qquad K_1 := 1_{\{\nu_1>1\}} \sum_{u=2}^{m} \frac{\nu_u}{\eta_u - \eta_1},$$

$$(ii) \qquad K_2 := 1_{\{\nu_1>2\}} \sum_{a_2+\ldots+a_M=2} \prod_{u=2}^{m} \left(\frac{1}{\eta_u-\eta_1}\right)^{a_u} \binom{\nu_u+a_u-1}{\nu_u-1},$$

$$(iii) \qquad \tilde{K_1} := (\nu_1-1)K_1,$$

$$(iv) \qquad \tilde{K_2} := (\nu_1-1)(\nu_1-2)K_2,$$

(v) *by* $(a_N)_{N\in\mathbb{N}}$ *we denote sequences with the property*

$$a_N := 1 - \eta_1 \frac{1}{N+\nu_1-1}\tilde{K_1} + \eta_1^2 \frac{1}{(N+\nu_1-1)(N+\nu_1-2)}\tilde{K_2} + O\left(\left(\frac{1}{N}\right)^3\right).$$

Note that $K_1 = \tilde{K_1} = 0$ if $\nu_1 = 1$ and that $K_2 = \tilde{K_2} = 0$ if $\nu_1 \le 2$. The following fact is a direct consequence of the definitions.

Lemma 2.6 *Let* $(a_N)_{N\in\mathbb{N}}$ *denote a sequence with*

$$a_N := 1 - \eta_1 \frac{1}{N+\nu_1-1}\tilde{K_1} + \eta_1^2 \frac{1}{(N+\nu_1-1)(N+\nu_1-2)}\tilde{K_2} + O\left(\left(\frac{1}{N}\right)^3\right)$$

Then

$$\lim_{N\to\infty} N(a_N - 1) = -\eta_1 \tilde{K_1}. \tag{2.9}$$

Proposition 2.7 *The norming constant in the general cycle is*

$$G(M,N) \tag{2.10}$$
$$= \sum_{l=1}^{m} \sum_{a\in Z(m,\nu_l-1)} (-1)^{a_l-\nu_l+1} \left(\frac{1}{\eta_l}\right)^{N-\nu_l+a_l+1} \binom{N+a_l}{N} \cdot \left(\prod_{u=1,u\neq l}^{m} \eta_u^{\nu_u} \binom{\nu_u+a_u-1}{\nu_u-1} \left(\frac{1}{\eta_u-\eta_l}\right)^{\nu_u+a_u} \right)$$

The norming constants obey the following asymptotic expansion:

$$G(M,N) = \left(\frac{1}{\eta_1}\right)^N \binom{N+\nu_1-1}{N} \left(\prod_{u=2}^{m} \left(\frac{\eta_u}{\eta_u-\eta_1}\right)^{\nu_u} \right) a_N, \tag{2.11}$$

where $(a_N)_{N\in\mathbb{N}}$ *is a sequence according to (v) in Definition 2.5.*

(2.11) tells us that the partition function $G(M,N)$ is dominated by the term

$$\left(\frac{1}{\eta_1}\right)^N \binom{N+\nu_1-1}{N} \prod_{u=2}^{m} \left(\frac{\eta_u}{\eta_u-\eta_1}\right)^{\nu_u}.$$

It turns out that the proof of (2.11) is of a prototype structure for many of our later arguments. We therefore give the details here.

Proof. (i) Equation (2.10) follows immediately from (2.8).

(ii) We start with the norming constant in the form (2.10) and split the term where $l = 1$, i.e., the summand representing the bottlenecks. As we will show later on, only this summand is persistent and therefore of importance for computing asymptotic distributions.

$$
\begin{aligned}
& G(M,N) \\
= & \sum_{l=2}^{m} \sum_{a\in Z(m,\nu_l-1)} (-1)^{a_l-\nu_l+1} \left(\frac{1}{\eta_l}\right)^{N-\nu_l+a_l+1} \binom{N+a_l}{N} \left(\prod_{u=1,u\neq l}^{m} \eta_u^{\nu_u} \binom{\nu_u+a_u-1}{\nu_u-1} \left(\frac{1}{\eta_u-\eta_l}\right)^{\nu_u+a_u} \right) \\
& + \sum_{a\in Z(m,\nu_1-1)} (-1)^{a_1-\nu_1+1} \left(\frac{1}{\eta_1}\right)^{N-\nu_1+a_1+1} \binom{N+a_1}{N} \left(\prod_{u=2}^{m} \eta_u^{\nu_u} \binom{\nu_u+a_u-1}{\nu_u-1} \left(\frac{1}{\eta_u-\eta_1}\right)^{\nu_u+a_u} \right) \\
= & \sum_{l=2}^{m} \sum_{a\in Z(m,\nu_l-1)} (-1)^{a_l-\nu_l+1} \left(\frac{1}{\eta_l}\right)^{N-\nu_l+a_l+1} \binom{N+a_l}{N} \left(\prod_{u=1,u\neq l}^{m} \eta_u^{\nu_u} \binom{\nu_u+a_u-1}{\nu_u-1} \left(\frac{1}{\eta_u-\eta_l}\right)^{\nu_u+a_u} \right) \\
& + \sum_{a\in Z(m,\nu_1-1),a_1<\nu_1-3} (-1)^{a_1-\nu_1+1} \left(\frac{1}{\eta_1}\right)^{N-\nu_1+a_1+1} \binom{N+a_1}{N} \left(\prod_{u=2}^{m} \eta_u^{\nu_u} \binom{\nu_u+a_u-1}{\nu_u-1} \left(\frac{1}{\eta_u-\eta_1}\right)^{\nu_u+a_u} \right) \\
& + \sum_{a\in Z(m,\nu_1-1),a_1\geq\nu_1-3} (-1)^{a_1-\nu_1+1} \left(\frac{1}{\eta_1}\right)^{N-\nu_1+a_1+1} \binom{N+a_1}{N} \left(\prod_{u=2}^{m} \eta_u^{\nu_u} \binom{\nu_u+a_u-1}{\nu_u-1} \left(\frac{1}{\eta_u-\eta_1}\right)^{\nu_u+a_u} \right)
\end{aligned}
$$

Splitting the last expressions for $a_1 \geq \nu_1 - 3$, i.e. $a_1 = \nu_1 - 1,\ a_1 = \nu_1 - 2$ and $a_1 = \nu_1 - 3$, yields

$$
\begin{aligned}
& G(M,N) \\
= & \sum_{l=2}^{m} \sum_{a\in Z(m,\nu_l-1)} (-1)^{a_l-\nu_l+1} \left(\frac{1}{\eta_l}\right)^{N-\nu_l+a_l+1} \binom{N+a_l}{N} \left(\prod_{u=1,u\neq l}^{m} \eta_u^{\nu_u} \binom{\nu_u+a_u-1}{\nu_u-1} \left(\frac{1}{\eta_u-\eta_l}\right)^{\nu_u+a_u} \right) \\
& + \sum_{a\in Z(m,\nu_1-1),a_1<\nu_1-3} (-1)^{a_1-\nu_1+1} \left(\frac{1}{\eta_1}\right)^{N-\nu_1+a_1+1} \binom{N+a_1}{N} \left(\prod_{u=2}^{m} \eta_u^{\nu_u} \binom{\nu_u+a_u-1}{\nu_u-1} \left(\frac{1}{\eta_u-\eta_1}\right)^{\nu_u+a_u} \right) \\
& + \left(\frac{1}{\eta_1}\right)^{N-2} \binom{N+\nu_1-3}{N} \left(\prod_{u=2}^{m} \left(\frac{\eta_u}{\eta_u-\eta_1}\right)^{\nu_u} \right) K_2 \\
& - \left(\frac{1}{\eta_1}\right)^{N-1} \binom{N+\nu_1-2}{N} \left(\prod_{u=2}^{m} \left(\frac{\eta_u}{\eta_u-\eta_1}\right)^{\nu_u} \right) K_1 \\
& + \left(\frac{1}{\eta_1}\right)^{N} \binom{N+\nu_1-1}{N} \prod_{u=2}^{m} \left(\frac{\eta_u}{\eta_u-\eta_1}\right)^{\nu_u}.
\end{aligned}
$$

A short comment may be in order here: So far, we only have split the sum. First, we extracted the summand with $l = 1$, then we extracted the summands with $a_1 \geq \nu_1 - 3$. Finally, we wrote down the explicit expressions for $a_1 = \nu_1 - 1,\ a_1 = \nu_1 - 2$ and $a_1 = \nu_1 - 3$. Now we factor out $\left(\frac{1}{\eta_1}\right)^N \binom{N+\nu_1-1}{N} \prod_{u=2}^m \left(\frac{\eta_u}{\eta_u-\eta_1}\right)^{\nu_u}$.
Since

$$
\lim_{N\to\infty} N^p \left(\frac{\eta_1}{\eta_l}\right)^N = 0,\ p \in \mathbb{N},\ 2 \leq l \leq m,
$$

it follows that

$$
\begin{aligned}
& \left(\left(\frac{1}{\eta_1}\right)^N \binom{N+\nu_1-1}{N} \prod_{u=2}^{m} \left(\frac{\eta_u}{\eta_u-\eta_1}\right)^{\nu_u} \right)^{-1} \\
& \cdot \sum_{l=2}^{m} \sum_{a\in Z(m,\nu_l-1)} (-1)^{a_l-\nu_l+1} \left(\frac{1}{\eta_l}\right)^{N-\nu_l+a_l+1} \binom{N+a_l}{N} \cdot \left(\prod_{u=1,u\neq l}^{m} \eta_u^{\nu_u} \binom{\nu_u+a_u-1}{\nu_u-1} \left(\frac{1}{\eta_u-\eta_l}\right)^{\nu_u+a_u} \right) \\
= & O\left(\left(\frac{1}{N}\right)^r\right), \qquad \text{for all } r \in \mathbb{N}.
\end{aligned}
$$

We choose $r = 3$ and receive $O\left(\left(\frac{1}{N}\right)^3\right)$. Since

$$\frac{\binom{N+a_1}{N}}{\binom{N+\nu_1-1}{N}} = O\left(\left(\frac{1}{N}\right)^3\right) \text{ for } a_1 < \nu_1 - 3,$$

it follows that

$$\begin{aligned}
G(M,N) &= \left(\frac{1}{\eta_1}\right)^N \binom{N+\nu_1-1}{N} \prod_{u=2}^{m} \left(\frac{\eta_u}{\eta_u-\eta_1}\right)^{\nu_u} \\
&\quad \cdot \left[1 - \eta_1 \frac{\nu_1-1}{N+\nu_1-1} K_1 + \eta_1^2 \frac{(\nu_1-1)(\nu_1-2)}{(N+\nu_1-1)(N+\nu_1-2)} K_2 + O\left(\left(\frac{1}{N}\right)^3\right)\right] \\
&= \left(\frac{1}{\eta_1}\right)^N \binom{N+\nu_1-1}{N} \prod_{u=2}^{m} \left(\frac{\eta_u}{\eta_u-\eta_1}\right)^{\nu_u} \\
&\quad \cdot \left[1 - \eta_1 \frac{1}{N+\nu_1-1} \tilde{K_1} + \eta_1^2 \frac{1}{(N+\nu_1-1)(N+\nu_1-2)} \tilde{K_2} + O\left(\left(\frac{1}{N}\right)^3\right)\right] \\
&= \left(\frac{1}{\eta_1}\right)^N \binom{N+\nu_1-1}{N} \prod_{u=2}^{m} \left(\frac{\eta_u}{\eta_u-\eta_1}\right)^{\nu_u} a_N . \square
\end{aligned}$$

□

The next calculations prepare to obtain asymptotic expansions of the first two cycle time moments.

Proposition 2.8 *It holds*

(i)

$$\begin{aligned}
\psi^{(M,N+1)}(\theta) &= G(M,N)^{-1} \sum_{l=1}^{m} \sum_{a \in Z(m,\nu_l-1)} (-1)^{a_l-\nu_l+1} \eta_l^{\nu_l} \left(\frac{1}{\eta_l+\theta}\right)^{N+a_l+1} \\
&\quad \cdot \binom{N+a_l}{N} \left(\prod_{u=1,u\neq l}^{m} \eta_u^{\nu_u} \binom{\nu_u+a_u-1}{\nu_u-1} \left(\frac{1}{\eta_u-\eta_l}\right)^{\nu_u+a_u}\right)
\end{aligned} \tag{2.12}$$

(ii)

$$E[c_{N+1}] = \eta_1^{-1}(N+\nu_1)\frac{1}{a_N} - \frac{1}{a_N}\tilde{K_1} + \eta_1 \frac{1}{a_N}\frac{1}{(N+\nu_1-1)}\tilde{K_2} + O\left(\left(\frac{1}{N}\right)^2\right) \tag{2.13}$$

(iii)

$$E[c_{N+1}] = \eta_1^{-1}(N+\nu_1)\frac{1}{a_N} - \frac{1}{a_N}\tilde{K_1} + O\left(\frac{1}{N}\right) \tag{2.14}$$

(iv)

$$E[c_{N+1}^2] = \eta_1^{-2}(N+\nu_1)(N+\nu_1+1)\frac{1}{a_N} - \eta_1^{-1}(N+\nu_1)\frac{1}{a_N}\tilde{K_1} + \frac{1}{a_N}\tilde{K_2} + O\left(\frac{1}{N}\right) \tag{2.15}$$

Proof. (i)

$$\psi^{(M,N+1)}(\theta) G(M,N) \stackrel{(1.3)}{=} G(M,N) \sum_{n \in Z(M,N)} \pi^{(M,N)}(n) \prod_{j=1}^{M} \left(\frac{\mu_j}{\mu_j+\theta}\right)^{n_j+1}$$

$$
\begin{aligned}
&= \left(\prod_{j=1}^{M}\frac{\mu_j}{\mu_j+\theta}\right)\sum_{n\in Z(M,N)}\prod_{j=1}^{M}\left(\frac{1}{\mu_j}\right)^{n_j}\prod_{j=1}^{M}\left(\frac{\mu_j}{\mu_j+\theta}\right)^{n_j} \\
&= \prod_{l=1}^{m}\left(\frac{\eta_l}{\eta_l+\theta}\right)^{\nu_l}\cdot\sum_{n\in Z(M,N)}\prod_{j=1}^{M}\left(\frac{1}{\mu_j+\theta}\right)^{n_j} \\
&\overset{(2.8)}{=} \prod_{l=1}^{m}\left(\frac{\eta_l}{\eta_l+\theta}\right)^{\nu_l}\sum_{l=1}^{m}\sum_{a\in Z(m,\nu_l-1)}(-1)^{a_l-\nu_l+1}\left(\frac{1}{\eta_l+\theta}\right)^{N-\nu_l+a_l+1}\binom{N+a_l}{N} \\
&\quad\cdot\left(\prod_{u=1,u\neq l}^{m}(\eta_u+\theta)^{\nu_u}\binom{\nu_u+a_u-1}{\nu_u-1}\left(\frac{1}{\eta_u-\eta_l}\right)^{\nu_u+a_u}\right) \\
&= \sum_{l=1}^{m}\sum_{a\in Z(m,\nu_l-1)}(-1)^{a_l-\nu_l+1}\eta_l^{\nu_l}\left(\frac{1}{\eta_l+\theta}\right)^{N+a_l+1}\binom{N+a_l}{N} \\
&\quad\cdot\left(\prod_{u=1,u\neq l}^{m}\eta_u^{\nu_u}\binom{\nu_u+a_u-1}{\nu_u-1}\left(\frac{1}{\eta_u-\eta_l}\right)^{\nu_u+a_u}\right).
\end{aligned}
$$

(ii) Since $E[c_{N+1}] = -\frac{\partial}{\partial\theta}\psi^{(M,N+1)}(\theta)|_{\theta=0}$, we first calculate $\frac{\partial}{\partial\theta}\psi^{(M,N+1)}(\theta)$:

$$
\begin{aligned}
\frac{\partial}{\partial\theta}\psi^{(M,N+1)}(\theta) \quad &\overset{(2.12)}{=} \quad -G(M,N)^{-1}\sum_{l=1}^{m}\sum_{a\in Z(m,\nu_l-1)}(-1)^{a_l-\nu_l+1}\eta_l^{\nu_l}(N+a_l+1) \\
&\cdot\left(\frac{1}{\eta_l+\theta}\right)^{N+a_l+2}\binom{N+a_l}{N}\left(\prod_{u=1,u\neq l}^{m}\eta_u^{\nu_u}\binom{\nu_u+a_u-1}{\nu_u-1}\left(\frac{1}{\eta_u-\eta_l}\right)^{\nu_u+a_u}\right).
\end{aligned}
$$

So,

$$
\begin{aligned}
& E[c_{N+1}] = -\frac{\partial}{\partial\theta}\psi^{(M,N+1)}(\theta)|_{\theta=0} \\
\overset{(2.16)}{=}\quad & G(M,N)^{-1}\sum_{l=1}^{m}\sum_{a\in Z(m,\nu_l-1)}(-1)^{a_l-\nu_l+1}(N+a_l+1)\left(\frac{1}{\eta_l}\right)^{N-\nu_l+a_l+2}\binom{N+a_l}{N} \\
& \left(\prod_{u=1,u\neq l}^{m}\eta_u^{\nu_u}\binom{\nu_u+a_u-1}{\nu_u-1}\left(\frac{1}{\eta_u-\eta_l}\right)^{\nu_u+a_u}\right) \\
=\quad & G(M,N)^{-1}\sum_{l=2}^{m}\sum_{a\in Z(m,\nu_l-1)}(-1)^{a_l-\nu_l+1}(N+a_l+1)\left(\frac{1}{\eta_l}\right)^{N-\nu_l+a_l+2}\binom{N+a_l}{N} \\
& \left(\prod_{u=1,u\neq l}^{m}\eta_u^{\nu_u}\binom{\nu_u+a_u-1}{\nu_u-1}\left(\frac{1}{\eta_u-\eta_l}\right)^{\nu_u+a_u}\right) \\
& +G(M,N)^{-1}\sum_{a\in Z(m,\nu_1-1),a_1<\nu_1-3}(-1)^{a_1-\nu_1+1}(N+a_1+1)\left(\frac{1}{\eta_1}\right)^{N-\nu_1+a_1+2}\binom{N+a_1}{N} \\
& \left(\prod_{u=2}^{m}\eta_u^{\nu_u}\binom{\nu_u+a_u-1}{\nu_u-1}\left(\frac{1}{\eta_u-\eta_1}\right)^{\nu_u+a_u}\right) \\
& +G(M,N)^{-1}\sum_{a\in Z(m,\nu_1-1),a_1\geq\nu_1-3)}(-1)^{a_1-\nu_1+1}(N+a_1+1)\left(\frac{1}{\eta_1}\right)^{N-\nu_1+a_1+2}\binom{N+a_1}{N} \\
& \left(\prod_{u=2}^{m}\eta_u^{\nu_u}\binom{\nu_u+a_u-1}{\nu_u-1}\left(\frac{1}{\eta_u-\eta_1}\right)^{\nu_u+a_u}\right).
\end{aligned}
$$

Applying equation (2.11) leads to the result that the first two sums are of order $O\left(\left(\frac{1}{N}\right)^2\right)$, therefore

$$\begin{aligned}
E[c_{N+1}] &= G(M,N)^{-1}(N+\nu_1)\left(\frac{1}{\eta_1}\right)^{N+1}\binom{N+\nu_1-1}{N}\prod_{u=2}^{m}\left(\frac{\eta_u}{\eta_u-\eta_1}\right)^{\nu_u}\\
&\quad -G(M,N)^{-1}(N+\nu_1-1)\left(\frac{1}{\eta_1}\right)^{N}\binom{N+\nu_1-2}{N}\left(\prod_{u=2}^{m}\left(\frac{\eta_u}{\eta_u-\eta_1}\right)^{\nu_u}\right)K_1\\
&\quad +G(M,N)^{-1}(N+\nu_1-2)\left(\frac{1}{\eta_1}\right)^{N-1}\binom{N+\nu_1-3}{N}\left(\prod_{u=2}^{m}\left(\frac{\eta_u}{\eta_u-\eta_1}\right)^{\nu_u}\right)K_2\\
&\quad +O\left(\left(\frac{1}{N}\right)^2\right)\\
&\overset{(2.11)}{=} (N+\nu_1)\eta_1^{-1}\frac{1}{a_N}-\frac{1}{a_N}\frac{(N+\nu_1-1)(\nu_1-1)}{N+\nu_1-1}K_1\\
&\quad +\eta_1\frac{1}{a_N}\frac{(N+\nu_1-2)(\nu_1-1)(\nu_1-2)}{(N+\nu_1-1)(N+\nu_1-2)}K_2+O\left(\left(\frac{1}{N}\right)^2\right)\\
&= (N+\nu_1)\eta_1^{-1}\frac{1}{a_N}-\frac{1}{a_N}\tilde{K_1}+\eta_1\frac{1}{a_N}\frac{1}{(N+\nu_1-1)}\tilde{K_2}+O\left(\left(\frac{1}{N}\right)^2\right).
\end{aligned}$$

(iii) This is an immediate corollary of (2.13).

(iv) Since $E[c_{N+1}^2]=\frac{\partial^2}{\partial\theta^2}\psi^{(M,N+1)}(\theta)|_{\theta=0}$, we first have to calculate $\frac{\partial^2}{\partial\theta^2}\psi^{(M,N+1)}(\theta)$:

$$\begin{aligned}
\frac{\partial^2}{\partial\theta^2}\psi^{(M,N+1)}(\theta) &\overset{(2.16)}{=} G(M,N)^{-1}\sum_{l=1}^{m}\sum_{a\in Z(m,\nu_l-1)}(-1)^{a_l-\nu_l+1}\eta_l^{\nu_l}(N+a_l+1)(N+a_l+2)\\
&\quad\cdot\left(\frac{1}{\eta_l+\theta}\right)^{N+a_l+3}\binom{N+a_l}{N}\left(\prod_{u=1,u\neq l}^{m}\eta_u^{\nu_u}\binom{\nu_u+a_u-1}{\nu_u-1}\left(\frac{1}{\eta_u-\eta_l}\right)^{\nu_u+a_u}\right).
\end{aligned}$$

Applying this result yields

$$\begin{aligned}
E[c_{N+1}^2] &= \frac{\partial^2}{\partial\theta^2}\psi^{(M,N+1)}(\theta)|_{\theta=0}\\
&= G(M,N)^{-1}\sum_{l=1}^{m}\sum_{a\in Z(m,\nu_l-1)}(-1)^{a_l-\nu_l+1}(N+a_l+1)(N+a_l+2)\\
&\quad\cdot\left(\frac{1}{\eta_l}\right)^{N-\nu_l+a_l+3}\binom{N+a_l}{N}\left(\prod_{u=1,u\neq l}^{m}\eta_u^{\nu_u}\binom{\nu_u+a_u-1}{\nu_u-1}\left(\frac{1}{\eta_u-\eta_l}\right)^{\nu_u+a_u}\right)\\
&= (N+\nu_1)(N+\nu_1+1)\eta_1^{-2}\frac{1}{a_N}-(N+\nu_1-1)(N+\nu_1)\eta_1^{-1}\frac{1}{a_N}\frac{1}{N+\nu_1-1}\tilde{K_1}\\
&\quad +(N+\nu_1-2)(N+\nu_1-1)\frac{1}{a_N}\frac{1}{(N+\nu_1-1)(N+\nu_1-2)}\tilde{K_2}+O\left(\frac{1}{N}\right)\\
&= (N+\nu_1)(N+\nu_1+1)\eta_1^{-2}\frac{1}{a_N}-(N+\nu_1)\eta_1^{-1}\frac{1}{a_N}\tilde{K_1}+\frac{1}{a_N}\tilde{K_2}+O\left(\frac{1}{N}\right).
\end{aligned}$$

□

After all these preparations, we are now in a position to prove our Theorems 2.1 and 2.2.

Proof of Theorem 2.1. Note that $\mu_1 = \eta_1$. It follows

$$\lim_{N\to\infty}\left(E[c_{N+1}] - \eta_1^{-1}(N+\nu_1)\right) \overset{(2.14)}{=} \lim_{N\to\infty}\left(\eta_1^{-1}(N+\nu_1)\frac{1}{a_N} - \frac{1}{a_N}\tilde{K_1} + O\left(\frac{1}{N}\right) - \eta_1^{-1}(N+\nu_1)\right)$$
$$= \eta_1^{-1}\lim_{N\to\infty}\left(\frac{N(1-a_N)}{a_N}\right) - \tilde{K_1} \overset{(2.9)}{=} 0.$$

The theorem follows. □

Proof of Theorem 2.2. Since $\mu_1 = \eta_1$, it is sufficient to show that

$$\lim_{N\to\infty}\left(Var(c_{N+1}) - \eta_1^{-2}(N+\nu_1)\right) = 0.$$

$$\begin{aligned}
E^2[c_{N+1}] &\overset{(2.13)}{=} \left(\eta_1^{-1}(N+\nu_1)\frac{1}{a_N} - \frac{1}{a_N}\tilde{K_1} + \eta_1\frac{1}{a_N}\frac{1}{(N+\nu_1-1)}\tilde{K_2} + O\left(\left(\frac{1}{N}\right)^2\right)\right)^2 \\
&= \left(\eta_1^{-1}(N+\nu_1)\frac{1}{a_N} - \frac{1}{a_N}\tilde{K_1} + \eta_1\frac{1}{a_N}\frac{1}{(N+\nu_1-1)}\tilde{K_2}\right)^2 + O\left(\frac{1}{N}\right) \\
&= \left(\eta_1^{-1}(N+\nu_1)\frac{1}{a_N} - \frac{1}{a_N}\tilde{K_1}\right)^2 + 2\left(\eta_1^{-1}(N+\nu_1)\frac{1}{a_N} - \frac{1}{a_N}\tilde{K_1}\right)\left(\eta_1\frac{1}{a_N}\frac{1}{(N+\nu_1-1)}\tilde{K_2}\right) \\
&\quad + \left(\eta_1\frac{1}{a_N}\frac{1}{(N+\nu_1-1)}\tilde{K_2}\right)^2 + O\left(\frac{1}{N}\right) \\
&= \left(\eta_1^{-1}(N+\nu_1)\frac{1}{a_N} - \frac{1}{a_N}\tilde{K_1}\right)^2 + 2\frac{1}{a_N^2}\frac{N+\nu_1}{N+\nu_1-1}\tilde{K_2} + O\left(\frac{1}{N}\right) \\
&= \eta_1^{-2}(N+\nu_1)^2\frac{1}{a_N^2} - 2\eta_1^{-1}(N+\nu_1)\frac{1}{a_N^2}\tilde{K_1} + \frac{1}{a_N^2}\tilde{K_1}^2 + 2\frac{1}{a_N^2}\frac{N+\nu_1}{N+\nu_1-1}\tilde{K_2} + O\left(\frac{1}{N}\right)
\end{aligned}$$

$$\begin{aligned}
&\lim_{N\to\infty}\left(Var(c_{N+1}) - \eta_1^{-2}(N+\nu_1)\right) = \lim_{N\to\infty}\left(E[c_{N+1}^2] - E^2[c_{N+2}] - \eta_1^{-2}(N+\nu_1)\right) \\
\overset{(2.15)}{=}\ & \lim_{N\to\infty}\Big[\eta_1^{-2}(N+\nu_1)(N+\nu_1+1)\frac{1}{a_N} - \eta_1^{-1}(N+\nu_1)\frac{1}{a_N}\tilde{K_1} + \frac{1}{a_N}\tilde{K_2} \\
& -\eta_1^{-2}(N+\nu_1)^2\frac{1}{a_N^2} + 2\eta_1^{-1}(N+\nu_1)\frac{1}{a_N^2}\tilde{K_1} \\
& -\frac{1}{a_N^2}\tilde{K_1}^2 - 2\frac{1}{a_N^2}\frac{N+\nu_1}{N+\nu_1-1}\tilde{K_2} - \eta_1^{-2}(N+\nu_1)\Big] \\
=\ & \lim_{N\to\infty}\Big[\eta_1^{-2}(N+\nu_1)^2\frac{1}{a_N} - \eta_1^{-2}(N+\nu_1)^2\frac{1}{a_N^2} \\
& -\eta_1^{-1}(N+\nu_1)\frac{1}{a_N}\tilde{K_1} + 2\eta_1^{-1}(N+\nu_1)\frac{1}{a_N^2}\tilde{K_1}\Big] \\
& + \lim_{N\to\infty}\left(\eta_1^{-2}(N+\nu_1)\frac{1}{a_N} - \eta_1^{-2}(N+\nu_1)\right) - \tilde{K_1}^2 - \tilde{K_2}
\end{aligned} \tag{2.16}$$

Note that

$$\lim_{N\to\infty}\left(\eta_1^{-2}(N+\nu_1)\frac{1}{a_N} - \eta_1^{-2}(N+\nu_1)\right) = \eta_1^{-2}\lim_{N\to\infty}\left(\frac{N(1-a_N)}{a_N}\right) \overset{(2.5.1)}{=} \eta_1^{-1}\tilde{K_1}. \tag{2.17}$$

Also note that with using Lemma 2.6 at $\overset{(1)}{=}$

$$\begin{aligned}
& \eta_1^{-2}(N+\nu_1)^2\frac{1}{a_N} - \eta_1^{-2}(N+\nu_1)^2\frac{1}{a_N^2} = \eta_1^{-2}\frac{1}{a_N^2}(N+\nu_1)^2(a_N-1) \\
\overset{(1)}{=} \; & \eta_1^{-2}\frac{1}{a_N^2}(N^2+2N\nu_1+\nu_1^2)\left(-\eta_1\frac{1}{N+\nu_1-1}\tilde{K_1}+\eta_1^2\frac{1}{(N+\nu_1-1)(N+\nu_1-2)}\tilde{K_2}+O\left(\left(\frac{1}{N}\right)^3\right)\right) \\
= \; & -\eta_1^{-1}\frac{1}{a_N^2}\frac{N^2}{N+\nu_1-1}\tilde{K_1}+\frac{1}{a_N^2}\frac{N^2}{(N+\nu_1-1)(N+\nu_1-2)}\tilde{K_2} \\
& \qquad -2\eta_1^{-1}\nu_1\frac{N}{N+\nu_1-1}\tilde{K_1}+O\left(\frac{1}{N}\right).
\end{aligned} \tag{2.18}$$

Inserting (2.17) and (2.18) into (2.16) leads to

$$\begin{aligned}
& \lim_{N\to\infty}\left(Var(c_{N+1})-\eta_1^{-2}(N+\nu_1)\right) \\
= \; & \lim_{N\to\infty}\left[-\eta_1^{-1}\frac{1}{a_N^2}\frac{N^2}{N+\nu_1-1}\tilde{K_1}-\eta_1^{-1}(N+\nu_1)\frac{1}{a_N}\tilde{K_1}+2\eta_1^{-1}(N+\nu_1)\frac{1}{a_N^2}\tilde{K_1}\right] \\
& \qquad +\eta_1^{-1}\tilde{K_1}-2\eta_1^{-1}\nu_1\tilde{K_1}-\tilde{K_1}^2+\tilde{K_2}-\tilde{K_2} \\
= \; & \eta_1^{-1}\tilde{K_1}\lim_{N\to\infty}\left[\frac{1}{a_N^2}\underbrace{\frac{(N+\nu_1)(N+\nu_1-1)-N^2}{N+\nu_1-1}}_{\to 2\nu_1-1}+\underbrace{\frac{1}{a_N^2}(N+\nu_1)(1-a_N)}_{\to\eta_1\tilde{K_1}}\right]+\eta_1^{-1}\tilde{K_1}-2\eta_1^{-1}\nu_1\tilde{K_1}-\tilde{K_1}^2 \\
= \; & 0.
\end{aligned}$$

□

2.2 Central limit theorem for cycle times

The following theorem removes the requirement of distinct service rates at all nodes in Theorem 4.1 of [DMS08]. Our preparatory derivations revealed that the methods needed for the proof are completely different.

Theorem 2.9 (Central limit theorem for the cycle time) *Let* $\mu_1 \le \ldots \le \mu_M$. *Then*

$$\frac{c_N - E[c_N]}{\sqrt{Var(c_N)}} \xrightarrow{\mathcal{D}} X \sim \mathcal{N}(0,1) \text{ as } N\to\infty. \tag{2.19}$$

Proof. We shall utilize from Slutsky's theorem [BD77][p. 461] the following facts: Let X, X_n, Y_n, $n \ge 1$, be real-valued random variables so that $X_n \xrightarrow{\mathcal{D}} X$ and $Y_n \xrightarrow{\mathcal{P}} c$ for some $c \in \mathbb{R}$. Then it holds

$$X_n + Y_n \xrightarrow{\mathcal{D}} X + c \tag{2.20}$$

$$X_n Y_n \xrightarrow{\mathcal{D}} Xc \tag{2.21}$$

Recall now from (2.2) that c_N can be expressed as $c_N = \delta_N + \rho_N$, with δ_N being the sum of N *iid* $\exp(\mu_1)$-distributed service times and ρ_N being the cumulative idle time. Therefore the normalized cycle time can be written as

$$\frac{c_N - E[c_N]}{\sqrt{Var(c_N)}} = \frac{\delta_N - E[\delta_N] + \rho_N - E[\rho_N]}{\sqrt{Var(c_N)}} = \frac{\delta_N - E[\delta_N]}{\sqrt{Var(\delta_N)}}\frac{\sqrt{Var(\delta_N)}}{\sqrt{Var(c_N)}} + \frac{\rho_N - E[\rho_N]}{\sqrt{Var(c_N)}}$$

First, note that because of the central limit theorem for *iid* random variables

$$\frac{\delta_N - E[\delta_N]}{\sqrt{Var(\delta_N)}} \xrightarrow{\mathcal{D}} X \sim \mathcal{N}(0,1) \text{ as } N \to \infty. \tag{2.22}$$

Since $Var(\delta_N) = \mu_1^{-2} N \to \infty$ as $N \to \infty$ and $\lim_{N\to\infty} |Var(c_N) - Var(\delta_N)| = const.$ (see (2.5)), it holds

$$\lim_{N\to\infty} \frac{\sqrt{Var(\delta_N)}}{\sqrt{Var(c_N)}} = 1$$

and therefore

$$\frac{\sqrt{Var(\delta_N)}}{\sqrt{Var(c_N)}} \xrightarrow{P} 1 \text{ as } N \to \infty. \tag{2.23}$$

Combining (2.22) and (2.23) and applying (2.21), we get

$$\frac{\delta_N - E[\delta_N]}{\sqrt{Var(\delta_N)}} \frac{\sqrt{Var(\delta_N)}}{\sqrt{Var(c_N)}} \xrightarrow{\mathcal{D}} X \sim \mathcal{N}(0,1) \text{ as } N \to \infty. \tag{2.24}$$

Finally, let us analyze the last term $\frac{\rho_N - E[\rho_N]}{\sqrt{Var(c_N)}}$. It holds

$$E\left[\left|\frac{\rho_N - E[\rho_N]}{\sqrt{Var(c_N)}}\right|\right] = \frac{E\left[|\rho_N - E[\rho_N]|\right]}{\sqrt{Var(c_N)}} \leq \frac{2E[\rho_N]}{\sqrt{Var(c_N)}} \to 0 \text{ as } N \to \infty \tag{2.25}$$

since $Var(c_N) \to \infty$ as $N \to \infty$ and

$$\lim_{N\to\infty} E[\rho_N] = \lim_{N\to\infty} (E[c_N] - E[\delta_N]) = \lim_{N\to\infty} \left(E[c_N] - N\mu_1^{-1}\right) \overset{(2.4)}{=} const.$$

Applying the Markov inequality

$$P\left(\left|\frac{\rho_N - E[\rho_N]}{\sqrt{Var(c_N)}}\right| \geq \varepsilon\right) \leq \frac{E\left[\left|\frac{\rho_N - E[\rho_N]}{\sqrt{Var(c_N)}}\right|\right]}{\varepsilon} \to 0 \text{ as } N \to \infty \ \forall \varepsilon > 0,$$

it follows that

$$\frac{\rho_N - E[\rho_N]}{\sqrt{Var(c_N)}} \xrightarrow{P} 0 \text{ as } N \to \infty. \tag{2.26}$$

Combining (2.24) and (2.26) and applying (2.20) yields

$$\frac{c_N - E[c_N]}{\sqrt{Var(c_N)}} \xrightarrow{\mathcal{D}} X \sim \mathcal{N}(0,1) \text{ as } N \to \infty.$$

□

3 Single bottleneck case versus multi bottleneck case

In this section, we analyze differences in the filling behaviour of the network in case of a single bottleneck and in case of multiple bottlenecks: How does the dynamic of the networks depend on the number of bottleneck nodes. Recall that we already noticed in Corollary 2.1 a surprising discontinuity of the asymptotic idle time behaviour when varying the service rates.

$$\lim_{N\to\infty} E[\rho_N] = \begin{cases} (\nu_1 - 1)\mu_1^{-1} > 0 & \text{for } \nu_1 > 1 \\ 0 & \text{for } \nu_1 = 1 \end{cases}$$

In case of pairwise distinct service rates ($\mu_1 < \ldots < \mu_M$), it even holds (cf. [Box88])

$$E[\rho_N] = N\mu_1^{-1} O\left(\left(\frac{\mu_1}{\mu_2}\right)^N\right).$$

For the cumulative idle time ρ_N, it makes a major difference whether the network has a single bottleneck node or multiple bottleneck nodes: In case of a single bottleneck, the cumulative idle time of a bottleneck node during a customer's cycle does not only converge stochastically to zero as the number of customers in the network goes to infinity, but the rate of convergence is very high. Thus, for large values of N, the bottleneck node will almost never be empty. Also, since $c_N = \delta_N + \rho_N$ and ρ_N is converging stochastically to zero, for large values of N the cycle time c_N can be approximated by δ_N, i.e. by the sum of N *iid* $\exp(\mu_1)$-distributed random variables. In contrast, in case of multiple bottleneck nodes, the expected values of the cumulative idle times converges towards a positive constant, i.e. during one cycle, the bottleneck node $Q[1]$ will on average be empty for a positive time period. By symmetry this also holds for the other bottleneck nodes.

We now consider the filling behaviour of bottleneck nodes and observe a similar dichotomy.

Theorem 3.1 *Let $Q[i]$ be a bottleneck node. Then it holds*

$$P\left(X_i^{(M,N)} \geq r\right) = \begin{cases} 1 - O\left(\frac{1}{N}\right) & \text{for } \nu_1 > 1 \\ 1 - N^{\nu_2 - 1} O\left(\left(\frac{\mu_1}{\mu_2}\right)^N\right) & \text{for } \nu_1 = 1 \end{cases} \tag{3.1}$$

and the equation is sharp with respect to speed of convergence.

Proof. Without loss of generality, we assume that $Q[i] = Q[1]$.

$$\begin{aligned} P\left(X_1^{(M,N)} < r\right) &= \sum_{n \in Z(M,N)} \pi(n) P\left(X_1^{(M,N)} < r \;\middle|\; X^{(M,N)} = n\right) \\ &\overset{(1.1)}{=} \sum_{n_1=0}^{r-1} \left[\sum_{n_2+\ldots+n_M = N-n_1} G(M,N)^{-1} \prod_{j=1}^{M} \left(\frac{1}{\mu_j}\right)^{n_j}\right] \\ &= \sum_{n_1=0}^{r-1} \left(\frac{1}{\mu_1}\right)^{n_1} \left[\frac{\sum_{n_2+\ldots+n_M=N-n_1} \prod_{j=2}^{M} \left(\frac{1}{\mu_j}\right)^{n_j}}{G(M,N)}\right] \end{aligned} \tag{3.2}$$

Let us first consider the case of multiple bottlenecks, i.e. $\nu_1 > 1$. Note that

$$\sum_{n_2+\ldots+n_M=N-n_1} \prod_{j=2}^{M} \left(\frac{1}{\mu_j}\right)^{n_j}$$

is the normalizing constant of a network with the same service rates, but with one bottleneck node and n_1 customers less. We therefore define for $j = 1, \ldots, m$

$$\tilde{\nu_j} = \begin{cases} \nu_1 - 1 & \text{for } j = 1 \\ \nu_j & \text{for } j \neq 1 \end{cases}.$$

It follows

$$\begin{aligned} \sum_{n_2+\ldots+n_M=N-n_1} \prod_{j=2}^{M} \left(\frac{1}{\mu_j}\right)^{n_j} &\overset{(2.8)}{=} \sum_{l=1}^{m} \sum_{a \in Z(m, \tilde{\nu_l}-1)} (-1)^{a_l - \tilde{\nu_l} + 1} \left(\frac{1}{\eta_l}\right)^{N - n_1 - \tilde{\nu_l} + a_l + 1} \binom{N - n_1 + a_l}{N - n_1} \\ &\cdot \left(\prod_{u=1, u\neq l}^{m} \eta_u^{\tilde{\nu_u}} \binom{\tilde{\nu_u} + a_u - 1}{\tilde{\nu_u} - 1} \left(\frac{1}{\eta_u - \eta_l}\right)^{\tilde{\nu_u} + a_u}\right). \end{aligned}$$

From (2.11) we know that

$$G(M,N) = \left(\frac{1}{\eta_1}\right)^N \binom{N+\nu_1-1}{N} \left(\prod_{u=2}^{m} \left(\frac{\eta_u}{\eta_u-\eta_1}\right)^{\nu_u}\right) a_N,$$

with $a_N = 1 + O\left(\frac{1}{N}\right)$. Since

$$\lim_{N\to\infty} N^p \left(\frac{\eta_1}{\eta_l}\right)^N = 0,\ p \in \mathbb{N},\ 2 \le l \le m,$$

it follows that

$$\begin{aligned}
& G(M,N)^{-1} \sum_{n_2+\ldots+n_M=N-n_1} \prod_{j=2}^{M} \left(\frac{1}{\mu_j}\right)^{n_j} \\
= \ & G(M,N)^{-1} \sum_{a\in Z(m,\tilde{\nu_1}-1)} (-1)^{a_1-\tilde{\nu_1}+1} \left(\frac{1}{\eta_1}\right)^{N-n_1-\tilde{\nu_1}+a_1+1} \\
& \cdot \binom{N-n_1+a_1}{N-n_1} \left(\prod_{u=2}^{m} \eta_u^{\tilde{\nu_u}} \binom{\tilde{\nu_u}+a_u-1}{\tilde{\nu_u}-1} \left(\frac{1}{\eta_u-\eta_1}\right)^{\tilde{\nu_u}+a_u}\right) + O\left(\left(\frac{1}{N}\right)^2\right).
\end{aligned}$$

Only the term with $a_1 = \tilde{\nu_1} - 1 = \nu_1 - 2$ needs to be considered, the rest is of order $O\left(\left(\frac{1}{N}\right)^2\right)$. Therefore

$$\begin{aligned}
& G(M,N)^{-1} \sum_{n_2+\ldots+n_M=N-n_1} \prod_{j=2}^{M} \left(\frac{1}{\mu_j}\right)^{n_j} \\
= \ & \frac{\left(\frac{1}{\eta_1}\right)^{N-n_1} \binom{N-n_1+\nu_1-2}{N-n_1} \prod_{u=2}^{m} \left(\frac{\eta_u}{\eta_u-\eta_1}\right)^{\nu_u}}{\left(\frac{1}{\eta_1}\right)^{N} \binom{N+\nu_1-1}{N} \left(\prod_{u=2}^{m} \left(\frac{\eta_u}{\eta_u-\eta_1}\right)^{\nu_u}\right)\left(1+O\left(\frac{1}{N}\right)\right)} + O\left(\left(\frac{1}{N}\right)^2\right) \\
= \ & \frac{\left(\frac{1}{\eta_1}\right)^{-n_1} \binom{N-n_1+\nu_1-2}{N-n_1}}{\binom{N+\nu_1-1}{N}\left(1+O\left(\frac{1}{N}\right)\right)} + O\left(\left(\frac{1}{N}\right)^2\right) \qquad (3.3) \\
= \ & \frac{O\left(\frac{1}{N}\right)}{1+O\left(\frac{1}{N}\right)} + O\left(\left(\frac{1}{N}\right)^2\right) = O\left(\frac{1}{N}\right). \qquad (3.4)
\end{aligned}$$

Substituting (3.4) into (3.2) leads to the first part of (3.1). From (3.3), we see that

$$\lim_{N\to\infty} \left(N \cdot G(M,N)^{-1} \sum_{n_2+\ldots+n_M=N-n_1} \prod_{j=2}^{M} \left(\frac{1}{\mu_j}\right)^{n_j}\right) = \begin{cases} > 0 & \text{for } n_1 < N \\ \ge 0 & \text{for } n_1 = N \end{cases}.$$

This shows that for $\nu_1 > 1$ it holds $P\left(X_i^{(M,N)} \ge r\right) = 1 - O\left(\frac{1}{N}\right)$ and $P\left(X_i^{(M,N)} \ge r\right) \ne 1 - o\left(\frac{1}{N}\right)$.

For the case of a single bottleneck, i.e. $\nu_1 = 1$, we have

$$\begin{aligned}
&\sum_{n_2+\ldots+n_M=N-n_1} \prod_{j=2}^{M} \left(\frac{1}{\mu_j}\right)^{n_j} \\
\stackrel{(2.8)}{=} &\sum_{l=2}^{m} \sum_{a\in Z(m,\nu_l-1)} (-1)^{a_l-\nu_l+1} \left(\frac{1}{\eta_l}\right)^{N-n_1-\nu_l+a_l+1} \binom{N-n_1+a_l}{N-n_1} \\
&\cdot \left(\prod_{u=1,u\neq l}^{m} \eta_u^{\nu_u} \binom{\nu_u+a_u-1}{\nu_u-1} \left(\frac{1}{\eta_u-\eta_l}\right)^{\nu_u+a_u}\right) \\
= &\sum_{l=2}^{m} \left(\frac{1}{\eta_l}\right)^{N} \sum_{a\in Z(m,\nu_l-1)} (-1)^{a_l-\nu_l+1} \left(\frac{1}{\eta_l}\right)^{-n_1-\nu_l+a_l+1} \binom{N-n_1+a_l}{N-n_1} \\
&\cdot \left(\prod_{u=1,u\neq l}^{m} \eta_u^{\nu_u} \binom{\nu_u+a_u-1}{\nu_u-1} \left(\frac{1}{\eta_u-\eta_l}\right)^{\nu_u+a_u}\right)
\end{aligned}$$

It follows with $G(M,N) = \left(\frac{1}{\eta_1}\right)^N \left(\prod_{u=2}^{m} \left(\frac{\eta_u}{\eta_u-\eta_1}\right)^{\nu_u}\right) a_N$ (cf. (2.11)):

$$\begin{aligned}
&G(M,N)^{-1} \sum_{n_2+\ldots+n_M=N-n_1} \prod_{j=2}^{M} \left(\frac{1}{\mu_j}\right)^{n_j} \\
= &\frac{1}{a_N} \sum_{l=2}^{m} \left(\frac{\eta_1}{\eta_l}\right)^{N} \sum_{a\in Z(m,\nu_l-1)} (-1)^{a_l-\nu_l+1} \left(\frac{1}{\eta_l}\right)^{-n_1-\nu_l+a_l+1} \binom{N-n_1+a_l}{N-n_1} \left(\prod_{u=2}^{m} \left(\frac{\eta_u}{\eta_u-\eta_1}\right)^{\nu_u}\right)^{-1} \\
&\cdot \left(\prod_{u=1,u\neq l}^{m} \eta_u^{\nu_u} \binom{\nu_u+a_u-1}{\nu_u-1} \left(\frac{1}{\eta_u-\eta_l}\right)^{\nu_u+a_u}\right) \\
= &\frac{1}{a_N} \left(\frac{\eta_1}{\eta_2}\right)^{N} \sum_{l=2}^{m} \left(\frac{\eta_2}{\eta_l}\right)^{N} \sum_{a\in Z(m,\nu_l-1)} (-1)^{a_l-\nu_l+1} \left(\frac{1}{\eta_l}\right)^{-n_1-\nu_l+a_l+1} \binom{N-n_1+a_l}{N-n_1} \left(\prod_{u=2}^{m} \left(\frac{\eta_u}{\eta_u-\eta_1}\right)^{\nu_u}\right)^{-1} \\
&\cdot \left(\prod_{u=1,u\neq l}^{m} \eta_u^{\nu_u} \binom{\nu_u+a_u-1}{\nu_u-1} \left(\frac{1}{\eta_u-\eta_l}\right)^{\nu_u+a_u}\right) \qquad (3.5)
\end{aligned}$$

Since $\lim_{N\to\infty} N^p \left(\frac{\eta_2}{\eta_l}\right)^N = 0$, $3 \le l \le m$, for all $p \in \mathbb{N}$, only the sum for $l = 2$ is of importance and only the summand with $a_2 = \nu_2 - 1$ needs to be considered. This leads to

$$\begin{aligned}
G(M,N)^{-1} \sum_{n_2+\ldots+n_M=N-n_1} \prod_{j=2}^{M} \left(\frac{1}{\mu_j}\right)^{n_j} &= \frac{1}{a_N} \binom{N-n_1+\nu_2-1}{N-n_1} \left(\frac{\eta_1}{\eta_2}\right)^N \left[1 + O\left(\frac{1}{N}\right)\right] \\
&= N^{\nu_2-1} O\left(\left(\frac{\eta_1}{\eta_2}\right)^N\right) \stackrel{\mu_2=\eta_2}{=} N^{\nu_2-1} O\left(\left(\frac{\mu_1}{\mu_2}\right)^N\right) \qquad (3.6)
\end{aligned}$$

Substituting (3.6) into (3.2) leads to the second part of (3.1).We can show that (3.5) implies

$$\lim_{N\to\infty} \left(\left\{N^{\nu_2-1} \left(\frac{\eta_1}{\eta_2}\right)^N\right\}^{-1} G(M,N)^{-1} \sum_{n_2+\ldots+n_M=N-n_1} \prod_{j=2}^{M} \left(\frac{1}{\mu_j}\right)^{n_j}\right) = \begin{cases} > 0 & \text{for } n_1 < N \\ \ge 0 & \text{for } n_1 = N \end{cases}.$$

This shows that the second equation of (3.1) is sharp according to the speed. □

Comment.

(i) From (3.1), we see that $\lim_{N\to\infty} P\left(X_1^{(M,N)} \geq r\right) = 1$ for every $r \in \mathbb{N}$. This means that the bottlenecks are filled up asymptotically. But the way they get filled up is essentially different: In case of a single bottleneck, we have a very high rate of convergence, in case of multiple bottlenecks, the rate of convergence is quite low ($O\left(\frac{1}{N}\right)$). The fact that in case of multiple bottlenecks, the bottleneck nodes are filled up slower, is comprehensible since the customers will spread over the bottleneck nodes. But the size of the difference as described by (3.1) is surprising.

(ii) Let us consider the filling behaviour in case of a single bottleneck. We see that the rate of convergence is mainly determined by the quotient of the slowest service rate and the second slowest service rate ($\left(\frac{\mu_1}{\mu_2}\right)^N$). This is well known [Box88] and not surprising. But, the speed of convergence is also influenced by the number of second slowest servers (N^{ν_2-1}), i.e. the more second slowest servers there are in the network, the more customers will be present at non-bottleneck nodes. This is an interesting result, an interpretation of which can be given as follows: The second slowest servers resist more against fast passing by customers than all other non-bottleneck nodes. Calling these nodes "semi-bottlenecks " would make this phenomenon intuitive: The semi-bottlenecks are the most important hills which customers have to climb up on their way to the single bottleneck. The more hills, the longer the time to reach the bottleneck for almost all customers.

(iii) Since the bottlenecks are filled up asymptotically, one might assume that the bottleneck nodes approach asymptotically Poissonian sources as the number of customers in the network increases to infinity and that the rest of the network forms an open ergodic tandem system. To be more precise: One might assume that the sections between two bottleneck nodes form an open ergodic tandem system. This is the topic of the following section.

4 Weak convergence of sojourn times at non-bottleneck nodes

In case of a single bottleneck the joint distribution of the queue lengths at the non-bottleneck nodes converges for $N \to \infty$ to the joint distribution of the queue lengths in an open tandem of nodes $Q[2]$, $Q[3]$,..., $Q[J]$ fed by a Poisson-μ_1 stream. This suggests that the joint sojourn time distribution at the non-bottleneck nodes for TC during his cycle converges for $N \to \infty$ to the joint distribution of the sojourn times of a customer passing this tandem. For the case of distinct service rates this follows from Theorem 5.1 in [DMS08].

Consequently, in case of multiple bottlenecks the conjecture is:

Consider two subsequent bottlenecks with some non-bottleneck nodes between them. Then the joint sojourn time distribution for TC at these non-bottleneck nodes converges weakly to the joint sojourn time distribution of a customer in an open ergodic tandem system with exactly these nodes, fed by a Poisson-μ_1 stream.

We can show more:

The limiting vectors of the joint sojourn time distribution for TC at the successive sequences of neighboured bottleneck nodes are independent.

So the limiting picture is:

The joint sojourn time distribution for TC at the non-bottleneck nodes during his cycle converges for $N \to \infty$ to the joint distribution of the sojourn times of a customer passing a sequence of independent tandems, where each of these tandem networks consists of a sequence of non-bottleneck nodes (which have in the original cycle at both sides of their boundary a bottleneck node).

Note: For fixing this interpretation of the following result the position of the nodes in the cycle matters and we incorporate this into the statement of our theorem.

Theorem 4.1 *Let* $(S_1^{(N)}, S_2^{(N)}, \ldots, S_M^{(N)})$ *denote a vector distributed according to* $\phi^{(M,N)}(\theta_1, \ldots, \theta_M)$ *from* (1.4)*. Let* $A := \{l \in \{1, \ldots, M\} : \ \mu_l \neq \mu_1\}$ *be the set of indices of the non-bottleneck nodes.*

Then the sequence $(S_j^{(N)}, j \in A)$ *converges for* $N \to \infty$ *to a vector with distribution*

$$\bigotimes_{j \in A} \exp(\mu_j - \mu_1). \tag{4.1}$$

Proof. The proof is a direct consequence of the formula (1.4), the fact that the mapping from $\mathbb{N}^A$ to $\mathbb{R}$

$$(n_j : j \in A) \to \prod_{j \in A} \left(\frac{\mu_j}{\mu_j + \theta_j} \right)^{n_j + 1}$$

is continuous for any fixed $(\theta_j, j \in A)$, and the result of the following Theorem 4.2.

Theorem 4.2 *The joint steady-state queue length distribution of the non-bottleneck nodes converges weakly to an independent product of geometrical distributions, which can be represented as the respective steady-state queue length distribution of an open ergodic tandem system with a Poisson-*μ_1 *arrival stream, i.e., it holds*

$$P^{(X_s^{(M,N)}, s \in A)} \xrightarrow{\mathcal{D}} \bigotimes_{s \in A} Geo(1 - \frac{\mu_1}{\mu_s}) \ \text{ as } N \to \infty. \tag{4.2}$$

Proof. To keep notation in the proof concise, we assume (without loss of generality) $\mu_1 \leq \mu_2 \leq \ldots \leq \mu_M$. (Removing this assumption makes the notation more involved but the same arguments apply.)

Then $A = \{\nu_1 + 1, \nu_1 + 2, \ldots, M\}$. We have to show that for all $n_s \in \mathbb{N}$, $\nu_1 + 1 \leq s \leq M$, with $n_0 := n_{\nu_1+1} + \ldots + n_M \leq N$:

$$\lim_{N \to \infty} P\left(X_{\nu_1+1}^{(M,N)} = n_{\nu_1+1}, \ldots, X_M^{(M,N)} = n_M\right) = \prod_{s=\nu_1+1}^{M} \left(1 - \frac{\mu_1}{\mu_s}\right) \left(\frac{\mu_1}{\mu_s}\right)^{n_s}.$$

Let therefore $n_s \in \mathbb{N}$, $\nu_1 + 1 \leq s \leq M$, with $n_0 := n_{\nu_1+1} + \ldots + n_M \leq N$ be arbitrary, but fixed.

$$\begin{aligned}
& P\left(X_{\nu_1+1}^{(M,N)} = n_{\nu_1+1}, \ldots, X_M^{(M,N)} = n_M\right) \\
= & \sum_{n_1 + \ldots + n_{\nu_1} = N - n_0} \prod_{j=1}^{\nu_1} \left(\frac{1}{\mu_1}\right)^{n_j} \prod_{j=\nu_1+1}^{M} \left(\frac{1}{\mu_j}\right)^{n_j} G(M,N)^{-1} \\
= & \sum_{n_1 + \ldots + n_{\nu_1} = N - n_0} \left(\frac{1}{\mu_1}\right)^{N - n_{\nu_1+1} - \ldots - n_M} \prod_{j=\nu_1+1}^{M} \left(\frac{1}{\mu_j}\right)^{n_j} G(M,N)^{-1} \\
= & \sum_{n_1 + \ldots + n_{\nu_1} = N - n_0} \left(\frac{1}{\mu_1}\right)^{N} \prod_{j=\nu_1+1}^{M} \left(\frac{\mu_1}{\mu_j}\right)^{n_j} G(M,N)^{-1} \\
= & \frac{\binom{N - n_0 + \nu_1 - 1}{N - n_0} \left(\frac{1}{\mu_1}\right)^{N} \prod_{j=\nu_1+1}^{M} \left(\frac{\mu_1}{\mu_j}\right)^{n_j}}{\sum_{k_0=0}^{N} \binom{N - k_0 + \nu_1 - 1}{N - k_0} \left(\frac{1}{\mu_1}\right)^{N} \sum_{m_{\nu_1+1} + \ldots + m_M = k_0} \prod_{j=\nu_1+1}^{M} \left(\frac{\mu_1}{\mu_j}\right)^{m_j}} \\
= & \frac{\binom{N - n_0 + \nu_1 - 1}{N - n_0} \prod_{j=\nu_1+1}^{M} \left(\frac{\mu_1}{\mu_j}\right)^{n_j}}{\sum_{k_0=0}^{N} \binom{N - k_0 + \nu_1 - 1}{N - k_0} \sum_{m_{\nu_1+1} + \ldots + m_M = k_0} \prod_{j=\nu_1+1}^{M} \left(\frac{\mu_1}{\mu_j}\right)^{m_j}}
\end{aligned}$$

In the following, we will assume that $\nu_1 \geq 2$. For $\nu_1 = 1$, (the setting of [GN67]) the proposition follows immediately from the above equation. For $\nu_1 \geq 2$, it follows

$$P\left(X_{\nu_1+1}^{(M,N)} = n_{\nu_1+1}, \ldots, X_M^{(M,N)} = n_M\right)$$

$$= \frac{\prod_{j=\nu_1+1}^M \left(\frac{\mu_1}{\mu_j}\right)^{n_j} \overbrace{(N-n_0+\nu_1-1)(N-n_0+\nu_1-2)\cdot\ldots\cdot(N-n_0+1)}^{\nu_1-1 \text{ factors}}}{\sum_{k_0=0}^N \sum_{m_{\nu_1+1}+\ldots+m_M=k_0} \prod_{j=\nu_1+1}^M \left(\frac{\mu_1}{\mu_j}\right)^{m_j} \cdot \underbrace{(N-k_0+\nu_1-1)(N-k_0+\nu_1-2)\cdot\ldots\cdot(N-k_0+1)}_{\nu_1-1 \text{ factors}}}$$

$$= \frac{\prod_{j=\nu_1+1}^M \left(\frac{\mu_1}{\mu_j}\right)^{n_j}}{\sum_{k_0=0}^N \sum_{m_{\nu_1+1}+\ldots+m_M=k_0} \prod_{j=\nu_1+1}^M \left(\frac{\mu_1}{\mu_j}\right)^{m_j} \cdot \frac{(N-k_0+\nu_1-1)(N-k_0+\nu_1-2)\cdot\ldots\cdot(N-k_0+1)}{(N-n_0+\nu_1-1)(N-n_0+\nu_1-2)\cdot\ldots\cdot(N-n_0+1)}}$$

The proposition will be proven if we can show that the denominator converges to $\left(\prod_{s=\nu_1+1}^M \left(1-\frac{\mu_1}{\mu_s}\right)\right)^{-1}$. Setting

$$a(k_0) := \sum_{m_{\nu_1+1}+\ldots+m_M=k_0} \prod_{j=\nu_1+1}^M \left(\frac{\mu_1}{\mu_j}\right)^{m_j},$$

$$b_N(k_0) := \frac{(N-k_0+\nu_1-1)(N-k_0+\nu_1-2)\cdot\ldots\cdot(N-k_0+1)}{(N-n_0+\nu_1-1)(N-n_0+\nu_1-2)\cdot\ldots\cdot(N-n_0+1)},$$

the denominator can be written as

$$\sum_{k_0=0}^N a(k_0)b_N(k_0).$$

Note that $b_N(k_0) > 1$ for $k_0 < n_0$ and is decreasing towards 1 for $N \to \infty$ and that $b_N(k_0) < 1$ for $k_0 > n_0$ and is increasing towards 1 for $N \to \infty$. With

$$f_N(k_0) := \begin{cases} a(k_0)b_N(k_0) & \text{for } k_0 \leq N \\ 0 & \text{for } k_0 > N \end{cases},$$

the denominator is

$$\sum_{k_0=0}^\infty f_N(k_0) = \sum_{k_0=0}^{n_0} f_N(k_0) + \sum_{k_0=n_0+1}^\infty f_N(k_0).$$

Clearly, $\lim_{N\to\infty} \sum_{k_0=0}^{n_0} f_N(k_0) = \sum_{k_0=0}^{n_0} a(k_0)$ because every $f_N(k_0)$ converges to $a(k_0)$.
Let us now consider the second sum $\sum_{k_0=n_0+1}^\infty f_N(k_0)$. Because $f_N(\cdot)$ converges point wise to $a(\cdot)$, it follows by monotone convergence that

$$\lim_{N\to\infty} \sum_{k_0=n_0+1}^\infty f_N(k_0) = \sum_{k_0=n_0+1}^\infty \lim_{N\to\infty} f_N(k_0) = \sum_{k_0=n_0+1}^\infty a(k_0).$$

Summarizing, we have shown

$$\sum_{k_0=0}^\infty a(k_0) = \sum_{k_0=0}^\infty \sum_{m_{\nu_1+1}+\ldots+m_M=k_0} \prod_{j=\nu_1+1}^M \left(\frac{\mu_1}{\mu_j}\right)^{m_j} = \prod_{j=\nu_1+1}^M \sum_{m_j=0}^\infty \left(\frac{\mu_1}{\mu_j}\right)^{m_j} = \left(\prod_{j=\nu_1+1}^M \left(1-\frac{\mu_1}{\mu_j}\right)\right)^{-1}.$$

□

References

[BD77] P. J. Bickel and K. Doksum. *Mathematical Statistics: Basic Ideas and Selected Topics*. Holden-Day, San Francisco, 1977.

[BKK84] O. J. Boxma, F. P. Kelly, and A. G. Konheim. The product form for sojourn time distributions in cyclic exponential queues. *Journal of the Association for Computing Machinery*, 31:128–133, 1984.

[Box88] O. J. Boxma. Sojourn times in cyclic queues - the influence of the slowest server. In O.J. Boxma, P.J. Courtois, and G. Iazeolla, editors, *Computer Performance and Reliability*, pages 13–24. North - Holland, Amsterdam, 1988.

[Cho80] W. M. Chow. The cycle time distribution of exponential cyclic queues. *Journal of the Association for Computing Machinery*, 27:281–286, 1980.

[DMS08] H. Daduna, C. Malchin, and R. Szekli. Weak convergence limits for sojourn times in cyclic queues under heavy traffic conditions. *Journal of Applied Probability*, 45:1–14, 2008.

[GN67] W.J. Gordon and G.F. Newell. Closed queueing networks with exponential servers. *Operations Research*, 15:254–265, 1967.

[Har85] P. G. Harrison. On normalizing constants in queueing networks. *Operations Research*, 33:464–468, 1985.

[Mal08] C. Malchin. *Stochastic Networks: Discrete and Continuous Time Models*. PhD thesis, Department of Mathematics, University of Hamburg, 2008.

[Rob03] P. Robert. *Stochastic Networks and Queues*. Springer, New York, 2003.

[Ser99] R. F. Serfozo. *Introduction to Stochastic Networks*, volume 44 of *Applications of Mathematics*. Springer, New York, 1999.