Online Companion for

“Heavy Traffic Analysis of Polling Systems in Tandem”

*Operations Research*
Vol. 47, No. 4, July-August 1999

Martin I. Reiman
Bell Laboratories, Lucent Technologies
Murray Hill, NJ

Lawrence M. Wein
Sloan School of Management
Massachusetts Institute of Technology
In this online companion to our paper (Reiman and Wein, 1999), we study four natural generalizations of our basic model. One of the purposes of this investigation is to illuminate the difficulties encountered in extending our analysis.

1 Unequal Service Rates

Let us suppose that the service rates at a station differ by class, and define $\mu_{ji}$ to be the service rate for class $ji$ for $j = A, B$ and $i = 1, \ldots, K$. The main difference in the analysis is that $\overline{V}_{A,i+1}(t)$ and $\overline{V}_{B,i+1}(t)$ are no longer constant when $s_i(t) = s_{i+1}(t)$. Although a limit cycle exists (as long as we take $\rho = 1$ in the deterministic analysis) under this generalization, the analysis is much more complicated and we only summarize the main results. A more detailed statement of the results (without derivation or proof) is presented in the appendix. Several new definitions are required. Define the service rate ratios $\tau_{ji} = \mu_{ji}/\mu_{j,i+1}$ for $j = A, B$ and $i = 1, \ldots, K - 1$. It is convenient to express our results in terms of classes $ki$ and $li$, rather than $Ai$ and $Bi$, where $\tau_{ki} \geq 1 \geq \tau_{li}$ for $i = 1, \ldots, K - 1$. The first limit cycle at station $i > 1$ begins at time $\tilde{t}_i = \inf\{t \geq 0 : \overline{V}_{ki}(t) = 0, s_{i-1}(t) = k, s_i(t) = l\}$. Unlike the analysis in Section 3 of Reiman and Wein, $V_i$ does not stay constant on the time scale of the deterministic analysis, and hence $\overline{V}_{ki}(\tilde{t}_i)$ does not necessarily equal $V_i(0) - \overline{V}_{ki}(\tilde{t}_i)$. The limit cycle dynamics cannot be analyzed without knowledge of $\overline{V}_{ki}(\tilde{t}_i)$. Define $\hat{t}_i = \inf\{t > 0 : s_i(t) \neq s_i(0)\}$ to be the first time server $i$ switches class. The analysis proceeds by recursively deriving $\hat{t}_i$, and then developing an expression for $\overline{V}_{ki}(\hat{t}_i)$ in terms of $\hat{t}_i$. Given $\overline{V}_{ki}(\hat{t}_i)$, $i = 1, \ldots, K - 1$, the next step is to calculate the limit cycle equations for $\overline{V}_{ki}(t)$ and $\overline{V}_{li}(t)$ corresponding to equations (3.16)-(3.23) in Reiman and Wein. The cycle lengths $C_{ki}$ and $C_{li}$, and virtual waiting times $Z_{ki}$ and $Z_{li}$ can then be obtained. Define $n_{li} = \left[\frac{\overline{V}_{li}(\hat{t}_i)}{\tau_{ki} C_{ki}}\right]$. There are two different structures to the
limit cycle here, depending on whether

1) \( V_{i,i+1}(\bar{t}_{i+1}) - (n_{li} - 1)r_{ki}C_{ki} \in (0, C_{ki}] \)

or

2) \( V_{i,i+1}(\bar{t}_{i+1}) - (n_{li} - 1)r_{ki}C_{ki} \in (C_{ki}, r_{ki}C_{ki}] \).

In case 1) class \( li \) is exhausted while \( s_{i}(t) = k, s_{i+1}(t) = l \), and in case 2) class \( li \) is exhausted while \( s_{i}(t) = s_{i+1}(t) = l \). Note that case 2) cannot occur if \( r_{ki} = 1 \), so this is a phenomenon that does not show up in the equal service rate setting.

When viewed on the time scale where the total workload process changes (and individual workloads are moving infinitely quickly), this process has a feature not present in the equal service rate case. Because the total workload at stations 2, ..., \( K \) may change over the limit cycle, the constraint that the total workload can never go negative is imposed on the minimum workload achieved over a cycle. As a consequence, the analysis of the vector total workload process must take into account the ‘oscillations’ in each station’s total workload.

2 Multi-Type Networks

Another natural generalization to investigate is the case of more than two customer types. The simplest example is a perfectly symmetric three-type, two-station network; that is, each type has arrival rate \( \lambda \) and all six classes have service rate \( \mu \). As before, let \( C_i \) equal the cycle length at station \( i \) and \( V_i \) be the normalized workload at station \( i \), for \( i = 1, 2 \). Also, let \( Z_2 \) denote the normalized virtual waiting time for any class at station 2, conditioned on \( V_1 \) and \( V_2 \). After analyzing numerous problem instances, we have the following conjecture. There are two cases, depending upon whether

\[
\left[ \frac{V_2}{C_1} \right] - \frac{V_2}{C_1} \geq \frac{1}{3} \quad \text{(case 1)}
\]

or

\[
\left[ \frac{V_2}{C_1} \right] - \frac{V_2}{C_1} \in \left[ 0, \frac{1}{3} \right) \quad \text{(case 2)}.
\]

In case 1, we have

\[
C_2 = \left( 3 \left[ \frac{V_2}{C_1} \right] - 2 \right) C_1
\]

and

\[
Z_2 = \begin{cases} 
V_2 - \left[ \frac{V_2}{C_1} \right] C_1 + j^{\frac{2}{3}C_1} & \text{w.p.} \quad \left( 3 \left[ \frac{V_2}{C_1} \right] - 2 \right)^{-1} \quad \text{for} \quad j = 0, \ldots, 2 \left[ \frac{V_2}{C_1} \right]; \\
V_2 - \left[ \frac{V_2}{C_1} \right] C_1 - j^{\frac{2}{3}C_1} & \text{w.p.} \quad \left( 3 \left[ \frac{V_2}{C_1} \right] - 2 \right)^{-1} \quad \text{for} \quad j = 1, \ldots, \left[ \frac{V_2}{C_1} \right].
\end{cases}
\]
In case 2, 
\[ C_2 = \left( 6 \left[ \frac{V_2}{C_1} \right] - 1 \right) C_1 \]  
(2.5) 
and 
\[ Z_2 = \begin{cases} 
V_2 - \left[ \frac{V_2}{C_1} \right] \frac{C_1}{3} + j C_1/3 & \text{m.w.p. } \left( 6 \left[ \frac{V_2}{C_1} \right] - 1 \right)^{-1} \text{ for } j = 0, \ldots, 2 \left[ \frac{V_2}{C_1} \right] - 1; \\
V_2 - \left[ \frac{V_2}{C_1} \right] \frac{C_1}{3} - j C_1/3 & \text{m.w.p. } \left( 6 \left[ \frac{V_2}{C_1} \right] - 1 \right)^{-1} \text{ for } j = 1, \ldots, \left[ \frac{V_2}{C_1} \right] - 1; \\
V_2 - \left( \left[ \frac{V_2}{C_1} \right] - 1 \right) \frac{C_1}{3} - j C_1/3 & \text{w.p. } \left( 6 \left[ \frac{V_2}{C_1} \right] - 1 \right)^{-1} \text{ for } j = 0, \ldots, 2 \left[ \frac{V_2}{C_1} \right] - 1; \\
V_2 - \left( \left[ \frac{V_2}{C_1} \right] - 1 \right) \frac{C_1}{3} + j C_1/3 & \text{w.p. } \left( 6 \left[ \frac{V_2}{C_1} \right] - 1 \right)^{-1} \text{ for } j = 1, \ldots, \left[ \frac{V_2}{C_1} \right]. 
\end{cases} \]  
(2.6) 
In case 1, the three customer classes at station 2 are served in contiguous blocks of \( C_2/3 \) time units. However, in case 2, each class is served twice in a cycle, once in the shorter subcycle described by the first two equations in (2.6) and once in the longer subcycle given by the last two equations of (2.6). There are six blocks of service in a cycle, which alternate in length between \( \left[ \frac{V_2}{C_1} \right] C_1 - C_1/3 \) and \( \left[ \frac{V_2}{C_1} \right] C_1 \). Since the server location \( s_i(t) \) dictates the dynamics of \( V_{j,i+1}(t) \), the complicated service process in case 2 would presumably make the analysis of a third station considerably more complex.

3 FCFS Queues

Most manufacturing systems have some workstations that incur sizeable setups and other workstations where setups are negligible. At the latter stations, FCFS is a commonly employed queueing discipline. In this section, we show that our analysis can be extended to a tandem network that consists of any combination of FCFS queues and polling queues. First, note that a set of product-form tandem FCFS queues can be placed in front of the network analyzed in this paper without altering our results; the Poisson output from the last FCFS queue preserves the product-form stationary distribution.

Now suppose that \( m \) FCFS queues are appended to the end of the network. Because customer overtaking cannot occur, the sojourn time of all customers in the FCFS stations is given by the workload. Using the snapshot principle, the sojourn time distribution can be calculated from the vector of station level workloads.

Now suppose station \( K + m + 1 \) is a polling queue. Then stations 1, \ldots, \( K + m \) affect \( \{ V_{j,K+m+1}(t), t \geq 0 \} \) only through the server location process \( s_{K+m}(t) \). Although we omit the details, Theorem 1 can be applied directly to analyze station \( K + m + 1 \); that is, equations (3.8)- (3.11) in Reiman and Wein hold for \( i = K + m \). To prove this fact one only needs to verify
that $r_{K+m+1} = \inf\{t \geq 0 : \nabla_{A,K+m}(t) = \nabla_{K+m}, \nabla_{A,K+m+1}(t) = 0\}$ is reached within a finite amount of time.

4 Closed networks

Suppose that customers exiting station $K$ immediately return to station 1, thereby generating a closed cyclic network. Although customers never enter or leave this network, in a manufacturing context where each customer represents a “job”, it is useful to think in terms of replacing a completed job that exits station $K$ with a new job that enters station 1. Readers are referred to Dai and Harrison (1993) for an extensive discussion of multiclass closed queueing network models for manufacturing systems. Chen and Mandelbaum (1991) prove a limit theorem for single class closed networks. As in the open case, their results immediately apply when the service time distribution depends only on the station.

Two types of closed networks can be formulated, depending upon the nature in which customers or jobs are injected back into the system. If a job departing station $K$ does not change type when it returns to station 1 then we have a multichain network, where the population size of each customer type remains constant. If the customer type of each reinjected job is exogenously prespecified (in a deterministic or probabilistic fashion) then a single-chain network is obtained; in the latter case, the total population size is fixed but the number of customers in the network of each type may fluctuate over time. Manufacturing system managers need to ensure that the output mix of job types equals the exogenous demand mix. The most direct and reliable way to guarantee this match between the desired mix and the realized mix is to use a single-chain network where the mix of entering jobs equals the exogenous product mix; see Harrison and Wein (1990) for further discussion on this point.

Consequently, we focus on the single-chain formulation. Let us alter the basic model in Section 2 of Reiman and Wein by disallowing exogenous arrivals and maintaining a fixed population size of $N$ customers. We assume that the long run proportion of exiting jobs that are reinjected as type $j$ customers is $q_j$, where $q_A + q_B = 1$. Also assume that the service time distributions are associated with each station rather than each class, so that the squared coefficients of variation satisfy $c_{A,i}^2 = c_{B,i}^2$ for $i = 1, \ldots, K$.

Recall that in the deterministic analysis of the open queueing network in Section 3 of Reiman and Wein, type $j$ work arrives to station 1 at the constant rate $\rho_j$; for $j = A, B$. Observe that in a similar deterministic analysis of this closed network, station 1 receives type $j$ work at a steady
rate of \( q_j \) for \( j = A, B \); furthermore, the interactions between stations \( i \) and \( i + 1 \) are identical in the open and closed networks for \( i = 1, \ldots, K - 1 \). Hence, the results in Section 3 of Reiman and Wein, namely equations (3.3), (3.5), (3.6) and (3.8)–(3.11), all hold for the closed cyclic network with \( \rho_j \) replaced by \( q_j \) for \( j = A, B \). Moreover, the closed network results may be more robust (with respect to the heavy traffic assumptions) than the open network results because \( q_A + q_B = 1 \), whereas we needed to assume that \( \rho_A + \rho_B = 1 \) in the open network analysis. The results on unequal service rates do not carry over to the closed network case because the unequal service rates lead to a fluctuating arrival rate of work at station 1 (depending upon which customer type is being served at station \( K \)) of the closed network.

If we further assume that setup times are zero and all \( K \) service time distributions have identical squared coefficients of variation, then the steady state heavy traffic analysis yields remarkably simple results, as shown in Section 11 of Dai and Harrison. The throughput rate of our network is identical to the throughput rate of the FCFS network considered by Dai and Harrison, and is given by equation (11.2) of their paper. Their approximating heavy traffic analysis implies that the stationary workload process \( (L_1, \ldots, L_K) \) is uniformly distributed on the simplex \( \{ x \in R^K_+: \sum_{k=1}^K x_k = \mu^{-1} N \} \). This result holds regardless of the values of \( q_j \) and the precise nature (i.e., deterministic or random) of the reinjection mechanism. Moreover, this result holds for \( K = 2 \) stations even if the squared coefficients of variation differ at the various stations. For the two-station case (whose analysis reduces to reflected Brownian motion on the interval \([0, \mu^{-1}]\)), the virtual waiting time distribution at station 1 is given by

\[
P(W_{j1} \leq t) = \begin{cases} \frac{\nu_{j1} t}{N} \left[ 1 + \ln \left( \frac{N}{\nu_{j1}} \right) \right] & \text{for } 0 \leq t \leq \frac{N}{\nu_{j1}} \\ 1 & \text{for } t \geq \frac{N}{\nu_{j1}}. \end{cases} \tag{4.1} \]

We also obtain \( E[W_{j1}] = N/4\mu q_j \) and \( Var[W_{j1}] = 7N^2/144\mu^2 q_j^2 \). If in addition \( q_A = q_B = 1/2 \), then \( E[W_{j1}] = E[W_{j2}] = N/2\mu \) and by (4.12) of Reiman and Wein,

\[
Var[W_{j2}] = \frac{4}{3} \left( \frac{N}{\mu} \right)^{-1} \int_0^{\frac{N}{\mu}} x^2 \left[ r_{acj} N \mu - x^2 \right] \left( \left[ \frac{N}{\mu} - x \right] - 1 \right) dx \tag{4.2}
\]

\[
= \frac{8N^2}{9\mu^2} \sum_{n=1}^\infty \frac{n(n-1)(12n^2+1)}{(4n^2-1)^3}. \tag{4.3}
\]

Calculation of the covariance of \( W_{j1} \) and \( W_{j2} \) leads to the variance for the steady state sojourn time of a type \( j \) customer, which is

\[
\frac{N^2}{36\mu^2} + \frac{8N^2}{9\mu^2} \sum_{n=1}^\infty \frac{n(n-1)(12n^2+1)}{(4n^2-1)^3} + \frac{(c_{j1}^2 + c_{j2}^2)}{\mu^2}. \tag{4.4}
\]
In contrast, under FCFS, the corresponding heavy traffic approximation for the variance of the virtual waiting time at each station is only $N^2/12\mu^2$ and the variance of the total sojourn time at both stations is simply $(c_{j1}^2 + c_{j2}^2)/\mu^2$ (due to the snapshot principle and the deterministic relationship between workloads and queue lengths).

Finally, let us consider the two-station cyclic network with setup times. The analysis in Coffman, Puhalskii and Reiman (1998) suggests that the normalized station 1 workload $V_1$ is a Bessel process on the interval $[0, \mu^{-1}]$ with instantaneous reflection, drift $\alpha(x) = q_A q_B s / x$, and variance $\sigma^2 = \mu^{-2}(c_{j1}^2 + c_{j2}^2)$, where the total population size is $N = \sqrt{n}$. From standard results on the stationary distribution of one-dimensional diffusions (c.f. Mandl 1968 or Karlin and Taylor 1981), the stationary density of this process solves the ordinary differential equation

$$\frac{\sigma^2}{2} \frac{d^2}{dx^2} \pi(x) - \frac{d}{dx}(\alpha(x)\pi(x)) = 0, \quad 0 < x < 1/\mu,$$

with boundary conditions

$$\frac{\sigma^2}{2} \frac{d\pi}{dx}(x) = \alpha(x)\pi(x), \quad x = 0, \frac{1}{\mu},$$

and normalization condition

$$\int_0^{1/\mu} \pi(x) dx = 1.$$

The solution of (4.5)–(4.7) is

$$\pi(x) = \frac{2\sigma^2}{(2\sigma^2 + s)\mu} \left(\frac{x}{\mu}\right)^{s/2\sigma^2}, \quad 0 \leq x \leq \mu^{-1}.$$ 

Note that if $s = 0$ (zero setup times) this reduces to a uniform density.

Appendix

In this appendix we provide some more details of the analysis of the unequal service rates case, considered in Section 1. For the sake of brevity we only state results, omitting derivations and proofs.

The first step in the analysis is a recursive determination of $\hat{t}_i$, from which an expression for $\sum \hat{t}_i$ can be obtained. In particular, $\hat{t}_1, \ldots, \hat{t}_{K-1}$ are defined by

$$\hat{t}_1 = \frac{\sum_{j=1}^{2}(0)}{1 - \frac{\nu_{j1}}{\mu_{j1}}} \text{ if } s_1(0) = j \text{ for } j = A, B,$$

(A.1)
and for \( i = 1, \ldots, K - 2 \),

\[
\hat{t}_{i+1} = \nabla_{k,i+1}(0) + r_k \hat{t}_i + \left( \max \left\{ 1, \frac{\nabla_{k,i+1}(0) + (r_k - 1)(\hat{t}_i - C_{ki})}{r_k C_{ki}} \right\} \right) - 1 \) \( r_k C_{ki} \) \hspace{1cm} (A.2) \\
\text{if } s_i(0) = s_{i+1}(0) = k; \\
\hat{t}_{i+1} = \nabla_{k,i+1}(0) + \left( \frac{\nabla_{k,i+1}(0) - \hat{t}_i + (r_k - 1)C_{ki}}{r_k C_{ki}} \right) I_{\{\nabla_{k,i+1}(0) > \hat{t}_i \}} - I_{\{\nabla_{k,i+1}(0) \leq \hat{t}_i \}} + 1 \) \( r_k C_{ki} \) \hspace{1cm} (A.3) \\
\text{if } s_i(0) = l \text{ and } s_{i+1}(0) = k; \\
\hat{t}_{i+1} = \nabla_{l,i+1}(0) + (m_k + 1) r_k C_{li} \text{ if } s_i(0) = k, s_{i+1}(0) = l \\
\text{and } \nabla_{l,i+1}(0) - \hat{t}_i + (r_i - 1)C_{li} - m_k r_k C_{ki} \in (-\infty I_{[m_k=1]}, C_{ki}], \hspace{1cm} (A.4) \\
\text{where}
\[
m_{ki} = \left[ \frac{\nabla_{l,i+1}(0) - \hat{t}_i + (r_i - 1)C_{li}}{r_k C_{ki}} \right] I_{\{\nabla_{l,i+1}(0) > \hat{t}_i - (r_i - 1)C_{li} \}} - I_{\{\nabla_{l,i+1}(0) \leq \hat{t}_i - (r_i - 1)C_{li} \}}; \hspace{1cm} (A.5) \\
\hat{t}_i = \hat{t}_i + (m_k + 1) C_i + (r_i - 1)^{-1}[(m_k + 1) r_k C_{ki} + \hat{t}_i - \nabla_{l,i+1}(0)] \\
\text{if } s_i(0) = k, s_{i+1}(0) = l \text{ and } \nabla_{l,i+1}(0) - \hat{t}_i + (r_i - 1)C_{li} - m_k r_k C_{ki} \in (C_{ki}, r_k C_{ki}], \hspace{1cm} (A.6) \\
\hat{t}_{i+1} = \nabla_{l,i+1}(0) + r_k \hat{t}_i + m_k r_k C_{li} \text{ if } s_i(0) = s_{i+1}(0) = l \\
\text{and } \nabla_{l,i+1}(0) + (r_i - 1) \hat{t}_i - m_k r_k C_{ki} \in (0, C_{ki}], \hspace{1cm} (A.7) \\
\text{where}
\[
m_{li} = \left[ \frac{\nabla_{l,i+1}(0) + (r_i - 1) \hat{t}_i}{r_k C_{ki}} \right] I_{\{\nabla_{l,i+1}(0) > (1-r_i) \hat{t}_i \}} - I_{\{\nabla_{l,i+1}(0) \leq (1-r_i) \hat{t}_i \}}; \hspace{1cm} (A.8) \\
\text{and}
\hat{t}_i = m_{li} C_i + C_{ki} + (r_i - 1)^{-1}[m_k r_k C_{ki} + C_{ki} - \nabla_{l,i+1}(0)] \text{ if } s_i(0) = s_{i+1}(0) = l \\
\text{and } \nabla_{l,i+1}(0) + (r_i - 1) \hat{t}_i - m_k r_k C_{ki} \in (C_{ki} - \infty I_{[m_k=1]}, r_k C_{ki}], \hspace{1cm} (A.9) \\
\text{The quantities } \nabla_{l2}(\bar{t}_2), \ldots, \nabla_{lK}(\bar{t}_K) \text{ are given by}
\[
\nabla_{l,i+1}(\bar{t}_{i+1}) = \nabla_{i+1}(0) + (r_k - 1) \hat{t}_i + (r_i - 1)C_{li} \text{ if } s_i(0) = k \hspace{1cm} (A.10) \\
\text{and}
\nabla_{l,i+1}(\bar{t}_{i+1}) = \nabla_{i+1}(0) + (r_k - 1) \hat{t}_i \text{ if } s_i(0) = l \hspace{1cm} (A.11)
for $i = 1, \ldots, K - 1$. Notice that, as expected, $\nabla_{l,i+1}(f_{i+1}) = \nabla_{i+1}(0)$ if $r_{ki} = r_{li} = 1$. With the vector of $\nabla_{li}(f_i)$’s in hand, we can calculate the limit cycle equations for $\nabla_{ki}(t)$ and $\nabla_{li}(t)$ corresponding to (3.16)-(3.23) of Reiman and Wein (they will not be written out here), and then derive the cycle lengths $C_{ki}$ and $C_{li}$ and virtual waiting times $Z_{ki}$ and $Z_{li}$ as in Section 3.

Define

$$n_{li} = \left[\frac{\nabla_{l,i+1}(f_{i+1})}{r_{ki}C_{ki}}\right] \quad \text{and} \quad n_{li} = \left[\frac{\nabla_{l,i+1}(f_{i+1})}{r_{li}C_{li}}\right].$$

(\text{A.12})

There are two possible limit cycles: In case 1, $\nabla_{l,i+1}(f_{i+1}) - (n_{li} - 1)r_{li}C_{li}, [0, C_{li}]$ and class $l$, $i+1$ is exhausted while $s_i(t) = k, s_{i+1}(t) = l$. In case 2, $\nabla_{l,i+1}(f_{i+1}) - (n_{ki} - 1)r_{ki}C_{ki} \in (C_{ki}, r_{ki}C_{ki})$ and class $l$, $i+1$ is exhausted while $s_i(t) = s_{i+1}(t) = l$. Let $U[a, b]$ denote a uniform random variable on the interval $[a, b]$, and let “w.p.” be shorthand for “with probability”. Under case 1, our results corresponding to Theorem 1 are as follows: For $i = 1, \ldots, K - 1$,

$$C_{k,i+1} = (n_{ki} - 1)C_i + n_{ki}r_{ki}C_{ki} - (n_{li} - 1)r_{li}C_{li},$$

(\text{A.13})

$$C_{l,i+1} = n_{ki}C_i + (n_{li} - 1)r_{li}C_{li} - n_{ki}r_{ki}C_{ki},$$

(\text{A.14})

$$C_{i+1} = (n_{ki} + n_{li} - 1)C_i,$$

(\text{A.15})

$$Z_{k,i+1} = (r_{ki} - 1)U[0, C_{ki}] + \nabla_{l,i+1}(f_{i+1}) + jr_{ki}C_{ki} \quad \text{w.p.} \quad \frac{1}{n_{ki} + n_{li} - 1} \quad \text{for } j = 1, \ldots, n_{li} - 1,$$

(\text{A.16})

$$Z_{k,i+1} = (r_{ki} - 1)U[0, C_{ki}] + \nabla_{l,i+1}(f_{i+1}) - jr_{ki}C_{ki} \quad \text{w.p.} \quad \frac{1}{n_{ki} + n_{li} - 1} \quad \text{for } j = 1, \ldots, n_{li} - 1,$$

(\text{A.17})

$$Z_{k,i+1} = (r_{ki} - 1)U[0, \nabla_{l,i+1}(f_{i+1}) - (n_{li} - 1)r_{li}C_{li}] + \nabla_{l,i+1}(f_{i+1}) \quad \text{w.p.} \quad \frac{\nabla_{l,i+1}(f_{i+1}) - (n_{ki} - 1)r_{ki}C_{ki}}{(n_{ki} + n_{li} - 1)C_{ki}}.$$

(\text{A.18})

$$Z_{l,i+1} = (r_{li} - 1)U[0, (n_{li} - 1)r_{ki}C_{ki} + C_{ki} - \nabla_{l,i+1}(f_{i+1})] + r_{ki}\nabla_{l,i+1}(f_{i+1}) + r_{ki}(n_{li} - 1)(r_{li} - 1)C_{li} \quad \text{w.p.} \quad \frac{(n_{li} - 1)r_{ki}C_{ki} + C_{ki} - \nabla_{l,i+1}(f_{i+1})}{(n_{ki} + n_{li} - 1)C_{ki}}.$$

(\text{A.19})

$$Z_{l,i+1} = (r_{li} - 1)U[0, C_{li}] + \nabla_{l,i+1}(f_{i+1}) + jr_{ki}C_{ki} \quad \text{w.p.} \quad \frac{1}{n_{ki} + n_{li} - 1} \quad \text{for } j = 0, \ldots, n_{li} - 2,$$

(\text{A.20})

$$Z_{l,i+1} = (r_{li} - 1)U[0, C_{li}] + \nabla_{l,i+1}(f_{i+1}) - C_{ki} + jr_{ki}C_{ki} \quad \text{w.p.} \quad \frac{1}{n_{ki} + n_{li} - 1} \quad \text{for } j = 2, \ldots, n_{ki},$$

(\text{A.21})

$$Z_{l,i+1} = (r_{li} - 1)U[0, C_{li}] + \nabla_{l,i+1}(f_{i+1}) - n_{ki}r_{li}C_{li} + \nabla_{l,i+1}(f_{i+1}) + (r_{ki} - 1)C_{ki} \quad \text{w.p.} \quad \frac{C_{li} + \nabla_{l,i+1}(f_{i+1}) - n_{ki}r_{li}C_{li}}{(n_{ki} + n_{li} - 1)C_{li}}.$$

(\text{A.22})
and
\[ Z_{l,i+1} = (r_{i} - 1)U[0, n_{k}r_{i}C_{li} - \nabla_{l,i+1}(\bar{t}_{i+1})] + r_{\bar{u}}\nabla_{l,i+1}(\bar{t}_{i+1}) - (r_{\bar{u}} - 1)n_{k}r_{\bar{u}}C_{\bar{u}} \]
\[ \text{w.p.} \quad \frac{n_{k}r_{\bar{u}}C_{\bar{u}} - \nabla_{l,i+1}(\bar{t}_{i+1})}{(n_{k} + n_{\bar{u}} - 1)C_{\bar{u}}} . \]  
(A.23)

Notice that (A.13)-(A.23) reduce to Theorem 1 if \( r_{ki} = r_{\bar{u}} \) for \( i = 1, \ldots, K - 1 \).

For case 2, we have, for \( i = 1, \ldots, K - 1 \),
\[ C_{k,i+1} = (n_{k} + n_{\bar{u}})r_{ki}C_{ki}, \]  
(A.24)

\[ C_{l,i+1} = (n_{k} + n_{li})r_{li}C_{li}, \]  
(A.25)

\[ C_{i+1} = (n_{k} + n_{\bar{u}})C_{i}, \]  
(A.26)

\[ Z_{k,i+1} = (r_{ki} - 1)U[0, C_{ki}] + (r_{\bar{u}} - 1)^{-1}[n_{\bar{u}}r_{li}C_{i} - \nabla_{l,i+1}(\bar{t}_{i+1})] - j_{\bar{u}}r_{k}C_{li} \]
\[ \text{w.p.} \quad \frac{1}{n_{ki} + n_{\bar{u}}} \quad \text{for} \quad j = 0, \ldots, n_{k} + n_{\bar{u}} - 1, \]  
(A.27)

\[ Z_{l,i+1} = (r_{i} - 1)U[0, C_{li}] + \nabla_{l,i+1}(\bar{t}_{i+1}) - C_{ki} - j_{k}r_{ki}C_{ki} \]
\[ \text{w.p.} \quad \frac{1}{n_{ki} + n_{\bar{u}}} \quad \text{for} \quad j = 0, \ldots, n_{k} - 2, \]  
(A.28)

\[ Z_{l,i+1} = (r_{i} - 1)U[0, n_{k}C_{i}] + \nabla_{l,i+1}(\bar{t}_{i+1}) - C_{ki} + (n_{k} - j)r_{ki}C_{ki} \]
\[ \text{w.p.} \quad \frac{1}{n_{ki} + n_{\bar{u}}} \quad \text{for} \quad j = 0, \ldots, n_{k} - 2, \]  
(A.29)

\[ Z_{l,i+1} = (r_{i} - 1)U[0, n_{k}C_{i} - (r_{\bar{u}} - 1)^{-1}[n_{\bar{u}}r_{li}C_{i} - \nabla_{l,i+1}(\bar{t}_{i+1})]] + (n_{k} + n_{li})r_{ki}C_{ki} \]
\[ \text{w.p.} \quad \frac{n_{k}C_{i} - (r_{\bar{u}} - 1)^{-1}[n_{\bar{u}}r_{li}C_{i} - \nabla_{l,i+1}(\bar{t}_{i+1})]}{(n_{k} + n_{\bar{u}})C_{li}} \]  
(A.30)

\[ Z_{l,i+1} = (r_{i} - 1)U[0, (r_{\bar{u}} - 1)^{-1}[n_{\bar{u}}r_{li}C_{i} - \nabla_{l,i+1}(\bar{t}_{i+1})]] - (n_{\bar{u}} - 1)C_{i} - C_{ki} \]
\[ - (n_{k} - 1)r_{ki}C_{ki} - C_{ki} \quad \text{w.p.} \quad \frac{(r_{\bar{u}} - 1)^{-1}[n_{\bar{u}}r_{li}C_{i} - \nabla_{l,i+1}(\bar{t}_{i+1})]}{(n_{k} + n_{\bar{u}})C_{li}} \]  
(A.31)

\[ Z_{l,i+1} = (r_{i} - 1)U[0, (r_{ki} - 1)C_{ki} - (n_{k} + n_{\bar{u}} - 1)r_{li}C_{ki} + (r_{\bar{u}} - 1)^{-1}[n_{\bar{u}}r_{li}C_{i} - \nabla_{l,i+1}(\bar{t}_{i+1})]] \]
\[ + \nabla_{l,i+1}(\bar{t}_{i+1}) + (r_{\bar{u}} - 1)C_{ki} \quad \text{w.p.} \quad \frac{(r_{ki} - 1)C_{ki} - (n_{k} + n_{\bar{u}} - 1)r_{li}C_{ki} + (r_{\bar{u}} - 1)^{-1}[n_{\bar{u}}r_{li}C_{i} - \nabla_{l,i+1}(\bar{t}_{i+1})]}{(n_{k} + n_{\bar{u}})C_{li}} \]  
(A.32)

and
\[ Z_{l_{ii+1}} = (r_{ii} - 1)U[0, (n_{ki} + n_{li})r_{li}C_{li} - (r_{ii} - 1)^{-1}[n_{ii}r_{ii}C_{ki} - b_{li+1}(\bar{r}_{i+1})]] \\
+ r_{ii}[(n_{ki} + n_{li})r_{li}C_{li} - n_{ki}C_{ki}] \\
\quad \text{w.p.} \quad \frac{(n_{ki} + n_{li})r_{li}C_{li} - (r_{ii} - 1)^{-1}[n_{ii}r_{ii}C_{ki} - \bar{V}_{l_{ii+1}}]}{(n_{ki} + n_{li})C_{li}} \tag{A.33} \]
References

[1] Chen, H. and A. Mandelbaum. 1991. Stochastic Discrete Flow Networks: Diffusion Approximation and Bottlenecks. *Annals of Probability* **19**, 1463-1519.

[2] Coffman, E. G., Jr., A. A. Puhalskii and M. I. Reiman. 1998. Polling Systems in Heavy-Traffic: A Bessel Process Limit. To appear in *Mathematics of Operations Research*.

[3] Dai, J. G. and J. M. Harrison. 1993. The QNET Method for Two-Moment Analysis of Closed Manufacturing Systems. *Annals of Applied Probability* **3**, 968-1012.

[4] Harrison, J. M. and L. M. Wein. 1990. Scheduling Networks of Queues: Heavy Traffic Analysis of a Two-Station Closed Network. *Operations Research* **38**, 1052-1064.

[5] Karlin, S. and H. M. Taylor. 1981. *A Second Course in Stochastic Processes*. Academic Press, New York.

[6] Mandl, P. 1968. *Analytical Treatment of One-Dimensional Markov Processes*. Springer-verlag, New York.

[7] Reiman, M. I. and L. M. Wein. 1999. Heavy Traffic Analysis of Polling Systems in Tandem. To appear in *Operations Research*. 