Gaussian Approximations for Nonstationary Loss Networks with Abandonment

Jamol Pender
School of Operations Research and Information Engineering
Cornell University
jjp274@cornell.edu

October 15, 2013

Abstract

In this paper, we analyze the performance and dynamics of time varying loss networks with possibility of customer abandonment while in the waiting spaces. To this end we develop new approximations that can estimate the mean, covariance matrix, and blocking probabilities of the entire network. The one dimensional loss network with abandonment has been studied by Pender [24], however, the multi-dimensional version presents new challenges. Thus, we appeal to the functional forward equations and make an independence assumption on pairwise stations of the loss network to simplify the computations. We illustrate that our approximation methods are quite accurate by comparing them with Monte Carlo simulation of some simple loss networks.

Keywords: Loss Networks, Blocking Probabilities, Multi-Server Queues, Abandonment, Dynamical Systems, Asymptotics, Time-Varying Rates, Time Inhomogeneous Markov Processes, Retrials, Multivariate Gaussian Distribution.

1 Introduction

In this paper, we study the time dependent behavior of non-stationary loss queues with the possibility of customer abandonment from the available waiting spaces. We denote the loss network with abandonment by the following Kendall notation, $(M_t/M_t/c_t/k_t + M_t)^N$. We assume that each station’s arrival process is a nonstationary Poisson process with a deterministic arrival rate function $\lambda_i(t), t \geq 0$ when there is space available for the customer to join the queue. Service times are independent random variables with service rate $\mu_i(t)$ and customers at station $i$ receive service from $c_i(t)$ parallel and homogenous servers. We also assume that there are $k_i(t)$ waiting spaces at each station. However, since customers are impatient, customers are allowed to abandon from the waiting spaces at rate $\beta_i(t)$ if they do not initiate service quickly enough. Lastly, once a customer either receives service or abandons, they are randomly routed to another station to receive service if there is enough space for them to join or they can leave the network permanently.
Loss queueing networks have been used for modeling a wide variety of stochastic systems from computer networks, internet networks, telecommunication networks, transportation networks, and more recently healthcare networks. Some example references are Hampshire et al [8, 9], Jagerman [10], Jennings and Massey [12], Alnowibet and Perros [1, 2]. A classic summary of stationary loss networks is given by Kelly [13]. Although stationary loss networks have been analyzed substantially by many authors, the literature on nonstationary loss networks is much less mature. This is partially because many of the techniques that are exploited in the stationary setting are not valid for in the nonstationary systems. Another major difficulty is that many of the properties that hold for stationary systems do not hold for nonstationary systems. One example of the difference is the insensitivity property of the Erlang loss queue. In fact, Davis et al [3] show that the performance of nonstationary loss queue is substantially influenced by the second moment and to some extent by the third moment of the service time distribution unlike its stationary counterpart.

Thus, many authors have developed approximations for the nonstationary loss queue and networks. One notable approximation method is the Pointwise Stationary Approximation (PSA), which was developed by Green and Kolesar [6] and shown to be asymptotically correct by Whitt [27]. To estimate the blocking probabilities, the PSA uses the Erlang blocking formula with $\frac{\lambda(t)}{\mu}$ as an approximation for the offered load. When the arrival rate changes very slowly, then the PSA works quite well. However, the PSA does not take into account any information of the service distribution beyond its mean value. Moreover, it is shown in Massey [17] that the PSA is the leading term in the asymptotic expansion using uniform acceleration and often needs further refinement when the arrival rate and service rates are not large or slowly varying.

Another approximation known as the modified offered load (MOL) was developed by Jagerman [10]. This approximation is constructed from the infinite server queue for which no blocking occurs. Hence, this approximation is only guaranteed to work well when the blocking probabilities are small and the arrival rate has small variation, see for example Massey and Whitt [21]. More recently authors Alnowibet and Perros [1, 2] have proposed the Fixed Point Approximation (FPA) point approximation for the loss queue; however, they do not estimate the variance and do not consider customer abandonment in their model, which are important aspects of real service systems that warrant study.

To analyze our loss network with abandonment we take advantage of the Markovian structure and appeal to the functional Kolmogorov forward equations for the queueing network. Similar to authors such as Taaffe et al [26] or Rothkopf et al [22], we encounter difficulties from the forward equations since they are not a closed system under general parameter settings. Thus, it is necessary to approximate the true distribution of the queueing network with a surrogate distribution that allows us to close the forward equations in a simple manner. More recently, authors Massey and Pender [18] and Pender [23, 24] have developed novel approximation methods for time varying queueing processes using orthogonal polynomial expansions for the queue length process. Furthermore, Massey and Pender [20] generalized the work of Ko et al [14] and [18] to Jackson networks with abandonment. We take a similar approach to the paper of [20] and use a multivariate Gaussian distribution to close the forward equations for our loss network. However, the loss network presents new difficulties.

The main difficulty of loss networks is when compared to their Jackson network counter-
parts is that if a customer is routed to another station, there must be enough capacity at the station to receive them, otherwise the customer cannot join the next station. This creates a dependence between stations that is not seen in traditional Jackson networks where customers can always join the next station since each station has an infinite capacity to hold them. Moreover, this dependency causes problems when trying to close the forward equations with a Gaussian distribution since each rate function includes an additional indicator function to maintain the loss nature of the network. To consider the full Gaussian distribution and to correctly characterize the true covariance under the Gaussian assumption, we must expand the rate functions in terms of an infinite series of Hermite polynomials. See Pender [25] for more details on this expansion. To circumvent the dependency issue and the infinite series of Hermite polynomials, we assume that the pairwise stations are independent. This assumption allows us to derive closed form expressions for the rate functions of the functional forward equations and we will show that our approximation works quite well. Our independence assumption is also supported by the recent work of Gurvich and Perry [7], which analyzes overflow networks in heavy traffic. One of their main results is to show that in heavy traffic, the overflow stations are asymptotically independent.

In the same spirit, we develop three new approximations that exploit the functional Kolmogorov forward equations of the queueing network. The first approximation assumes that the queueing network is deterministic. Our second approximation assumes that the distribution at each station is independent and has a Gaussian distribution with the same mean and variance. This approximation is motivated by the infinite server queue and the fact that it has a Poisson distribution. Moreover, the Poisson is approximately Gaussian with the same mean and variance. Our last approximation assumes that the queue length can be approximated by a multivariate Gaussian distribution; this approximation is partially motivated by the many server heavy traffic limit theorems of Mandelbaum et al [16]. We will show that the Gaussian assumption allows us to estimate several performance measures such as the mean, covariance matrix, and blocking probabilities with good accuracy.

1.1 Contributions

In this work we make several key contributions to the literature of nonstationary loss networks. Our first contribution is the development of a queueing model that includes nonstationary arrivals with the possibility of customer abandonment. Our second fundamental contribution is to identify that the Gaussian distribution with the same mean and variance, which is inspired from the Poisson distribution of the infinite server queue, can be used as a surrogate distribution to close the functional Kolmogorov forward equations. To the best of our knowledge, the combination of this approximation with the functional forward equations has not been considered in the literature and is new. Moreover, we also show that a multivariate Gaussian distribution in addition to a simple independence assumption can be quite useful to close the functional forward equations and develop accurate approximations for the mean and variance of the queueing process. Lastly, our approximation methods are very simple to implement and only involve the numerical integration of at most $\frac{N^2+3N}{2}$ differential equations for a N-dimensional network, which is quite fast to solve numerically.
1.2 Organization of the Paper

The rest of the paper is organized as follows. We start with a description of the queueing network under consideration in Section 2. We also derive the functional forward equations for the model, which are important for the future analysis. We analyze three new approximations for the queueing model in Section 3. In Section 4, we provide approximate formulas for the probability of blocking for each station of the network. In Section 5, we provide more numerical examples to support the use of our approximation methods. In Section 6, we give concluding remarks.

1.3 Notation

The paper will use the following notation:

- $\lambda_i(t)$ is the external arrival rate to station $i$ at time $t$
- $\beta_i(t)$ is the abandonment rate for station $i$ at time $t$
- $\mu_i(t)$ is the service rate for station $i$ at time $t$
- $\tau_{ij}(t)$ is the abandonment routing probability from station $i$ to station $j$ at time $t$
- $\gamma_{ij}(t)$ is the service routing probability from station $i$ to station $j$ at time $t$
- $\tau_i(t)$ is the abandonment departure probability from station $i$ at time $t$
- $\gamma_i(t)$ is the service departure probability from station $i$ at time $t$
- $c_i(t)$ is the number of servers for station $i$ at time $t$
- $k_i(t)$ is the number of additional waiting spaces for station $i$ at time $t$
- $e_i$ is a N-dimensional unit vector with a 1 in the $i^{th}$ dimension.
- $x \wedge y = \min(x, y)$
- $(x - y)^+ = \max(0, x - y)$
- $\{x < y\}$ denotes an indicator function that equals one if the statement is true i.e. if $x < y$, and zero if the statement is false.

Moreover, we also require the following relations

$$\tau_i^i + \sum_{j \neq i}^{N} \tau_{ij} = 1 \quad \text{and} \quad \gamma_i^i + \sum_{j \neq i}^{N} \gamma_{ij} = 1 \quad (1.1)$$

and

$$\tau_{ii}(t) = \gamma_{ii}(t) = 0 \quad \forall t \geq 0 \quad (1.2)$$
2 The Loss Network Queueing Model

In this section, we give a overview of the queueing network that we will analyze in this paper. Using Kendall notation, the queueing model that we consider is the \((M_t/M_t/c_t/k_t + M_t)^N\) queue. We assume that each station’s arrival process is a nonstationary Poisson process with a deterministic arrival rate function \(\lambda_i(t), t \geq 0\) when there is space available for the customer to join the queue. Service times are independent random variables with service rate \(\mu_i(t)\) and customers at station \(i\) receive service from \(c_i(t)\) parallel and homogenous servers. We also assume that there are \(k_i(t)\) waiting spaces at each station. However, since customers are impatient, customers are allowed to abandon from the waiting spaces at rate \(\beta_i(t)\) if they do not initiate service quickly enough. Lastly, once a customer either receives service or abandons, they are randomly routed to another station to receive service if there is enough space for them to join or they can leave the network permanently.

If we did not consider the possibility of customers being lost when there is no capacity to hold them, then our model reduces to a Jackson network with abandonment or the \((M_t/M_t/c_t/k_t + M_t)^N\) queue. Jackson networks with abandonment fall into the class of queueing models known as Markovian service networks, which were analyzed extensively Mandelbaum et al [16]. It is also shown in [16] that Jackson networks with abandonment have fluid and diffusion limits and be approximated by a Gaussian diffusion under mild technical conditions. However, the indicator function for the loss network, which prevents customers from joining a queue if there are not enough waiting spaces, precludes the same techniques to be applied to the loss systems, since they rely on the rate functions to be Lipschitz continuous functions of the queue length. Thus, we develop a new approach to estimate the performance of these types of models using the functional Kolmogorov forward equations of the queue length process. The functional Kolmogorov forward equations for the \((M_t/M_t/c_t/k_t + M_t)^N\) queue have the following form

\[
\dot{E}[f(Q)] = E[\alpha_i(Q) \cdot (f(Q + e_i) - f(Q))] + E[\delta_i(Q) \cdot (f(Q - e_i) - f(Q))] + \sum_{j=1}^{N} E[D_{ij}(Q) \cdot (f(Q - e_i + e_j) - f(Q))] + \sum_{j=1}^{N} E[D_{ij}(Q) \cdot (f(Q + e_i - e_j) - f(Q))]
\]

(2.3)

for all appropriate real-valued functions \(f: \mathbb{R}^n \rightarrow \mathbb{R}\) and where we have the following expressions for the rate functions

\[
\alpha_i(Q) = \lambda_i \cdot \{Q_i \leq c_i + k_i - 1\},
\]

(2.4)

\[
\delta_i(Q) = \mu_i \cdot (Q_i \land c_i) + \beta_i \cdot (Q_i - c_i)^+,
\]

(2.5)

\[
\tilde{D}_{ij}(Q) = \mu_i \cdot \tau_{ij} \cdot (Q_i \land c_i) \cdot \{Q_j \leq c_j + k_j - 1\} + \beta_i \cdot \gamma_{ij} \cdot (Q_i - c_i)^+ \cdot \{Q_j \leq c_j + k_j - 1\},
\]

(2.6)

and

\[
D_{ij}(Q) = \mu_j \cdot \tau_{ji} \cdot (Q_j \land c_j) \cdot \{Q_i \leq c_i + k_i - 1\} + \beta_j \cdot \gamma_{ji} \cdot (Q_j - c_j)^+ \cdot \{Q_i \leq c_i + k_i - 1\}.
\]

(2.7)
This implies that we have the following expressions for the mean, variance, and covariance matrix of the loss network

\[ E[Q_i] = E[\alpha_i(Q)] - E[\delta_i(Q)] - \sum_{j=1}^{N} E[D_{ij}(Q)] + \sum_{j=1}^{N} E[\tilde{D}_{ij}(Q)], \quad (2.8) \]

\[ \text{Var}[Q_i] = E[\alpha_i(Q)] + E[\delta_i(Q)] + \sum_{j=1}^{N} E[D_{ij}(Q)] + \sum_{j=1}^{N} E[\tilde{D}_{ij}(Q)] + 2 \cdot \text{Cov}[Q_i, \alpha_i(Q)] - 2 \cdot \text{Cov}[Q_i, \delta_i(Q)] \]

\[ - 2 \cdot \sum_{j=1}^{N} \text{Cov}[Q_i, \tilde{D}_{ij}(Q)] + 2 \cdot \sum_{j=1}^{N} \text{Cov}[Q_i, D_{ij}(Q)], \quad (2.9) \]

and

\[ \text{Cov}[Q_i, Q_l] = - \sum_{j=1}^{N} E[D_{ij}(Q)] - \sum_{j=1}^{N} E[\tilde{D}_{ij}(Q)] \]

\[ + \text{Cov}[Q_i, \alpha_i(Q)] + \text{Cov}[Q_i, \alpha_l(Q)] - \text{Cov}[Q_i, \delta_i(Q)] - \text{Cov}[Q_l, \delta_i(Q)] \]

\[ + \sum_{j=1}^{N} \text{Cov}[Q_i, D_{ij}(Q) - \tilde{D}_{ij}(Q)] - \sum_{j=1}^{N} \text{Cov}[Q_l, D_{ij}(Q) - \tilde{D}_{ij}(Q)]. \quad (2.10) \]

3 New Approximations

3.1 Deterministic Mean Approximation

Our first approximation will assume that the queueing network is deterministic. We call this the Deterministic Mean Approximation (DMA) since we assume that \( q(t) \equiv \{q_1(t)\ldots q_N(t)\} | t \geq 0 \) is a deterministic process that approximates the queueing process. This approximation gives us our first theorem.

**Theorem 3.1.** If we substitute the approximate distribution \( Q = q \) for the queue length process of the nonstationary loss network with abandonment, then we have the following differential equation for the mean queue length at the \( i^{th} \) station

\[ \dot{q}_i = \lambda_i \cdot \{q_i < c_i + k_i\} - \mu_i \cdot \gamma_i \cdot (q_i \wedge c_i) - \beta_i \cdot \tau_i \cdot (q_i - c_i) \]

\[ - \sum_{j=1}^{N} (q_i - c_i)^+ \cdot \beta_i \cdot \tau_{ij} \cdot \{q_j < c_j + k_j\} - \sum_{j=1}^{N} (q_i \wedge c_i) \cdot \mu_i \cdot \gamma_{ij} \cdot \{q_j < c_j + k_j\} \]

\[ + \sum_{j=1}^{N} (q_j - c_j)^+ \cdot \beta_j \cdot \tau_{ji} \cdot \{q_i < c_i + k_i\} + \sum_{j=1}^{N} (q_j \wedge c_j) \cdot \mu_j \cdot \gamma_{ji} \cdot \{q_i < c_i + k_i\}. \quad (3.11) \]
Table 1: Two Station Loss Network Model Parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\lambda_1$</td>
<td>$20 + 10 \sin(t)$</td>
<td>$\lambda_2$</td>
<td>$30 + 15 \sin(t)$</td>
</tr>
<tr>
<td>$\mu_1$</td>
<td>1</td>
<td>$\mu_2$</td>
<td>1</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>2</td>
<td>$\beta_2$</td>
<td>2</td>
</tr>
<tr>
<td>$c_1$</td>
<td>20</td>
<td>$c_2$</td>
<td>30</td>
</tr>
<tr>
<td>$k_1$</td>
<td>5</td>
<td>$k_2$</td>
<td>5</td>
</tr>
<tr>
<td>$\tau_{11}$</td>
<td>0</td>
<td>$\tau_{22}$</td>
<td>0</td>
</tr>
<tr>
<td>$\gamma_{11}$</td>
<td>0</td>
<td>$\gamma_{22}$</td>
<td>0</td>
</tr>
<tr>
<td>$\tau_{12}$</td>
<td>.25</td>
<td>$\tau_{21}$</td>
<td>0</td>
</tr>
<tr>
<td>$\gamma_{12}$</td>
<td>.25</td>
<td>$\gamma_{21}$</td>
<td>0</td>
</tr>
</tbody>
</table>

Figure 1: Simulation vs. DMA of $Q_1$ Mean (Left). Simulation vs. DMA of $Q_2$ Mean (Right).

The DMA in the network setting can be viewed as a N-dimensional projection onto the deterministic function $q(t)$. For the covariance, the DMA approximation yields a value of zero since the DMA implicitly assumes that the covariance and other cumulant moments are equal to zero. If we want to appropriately model other moments such as the covariance, we must add an additional term to the approximation in order to add some randomness into the approximation.

In Figure 1 we see that the DMA does well at approximating the mean dynamics of the loss network, especially in the underloaded region where the mean number of customers is far away from the total capacity of the station. However, we see that the DMA can be improved when the queueing network is near the capacity level of each station. In fact the DMA overestimates the value of the queue length since it does not take into account the stochastic fluctuations of the queue length process, which spends some time under the capacity threshold. This inspires us to add some randomness to our next approximation method.
3.2 Gaussian Infinite Server Approximation

Our second method for estimating the dynamics of the mean is inspired by the dynamics of the infinite server queueing process. In Eick et al [4], it is shown that the infinite server queue has a Poisson distribution when initialized at zero or with a Poisson distribution. Furthermore, the Poisson distribution is characterized by all of its cumulant moments having the same value as the mean, which implies that the mean and variance are equal. Moreover, when the mean is large, the Poisson distribution can be approximated very well by a Gaussian distribution with the same mean and variance. This approximation of the infinite server queue motivates our next approximation for the loss network. This new approximation is called the Gaussian Infinite Server Approximation (GISA). To construct the GISA we first let

\[ Q_i \overset{D}{=} q_i + \sqrt{q_i} \cdot X_i. \]  

(3.12)

This construction implies that

\[ \mathbb{E}[Q_i] = q_i \]  

(3.13)

\[ \text{Cov}[Q_i, Q_i] = q_i \]  

(3.14)

\[ \text{Cov}[Q_i, Q_j \neq i] \equiv 0. \]  

(3.15)

**Theorem 3.2.** If we substitute the above approximate distribution in Equation 3.12 for each of the queue length processes of the nonstationary loss network with abandonment, we get the following differential equations for the mean

\[ \dot{\mathbb{E}}[Q_i] = \mathbb{E}[\alpha_i(Q)] - \mathbb{E}[\delta_i(Q)] - \sum_{j=1}^{N} \mathbb{E}[D_{ij}(Q)] + \sum_{j=1}^{N} \mathbb{E}[\tilde{D}_{ij}(Q)], \]  

(3.16)

where we have the following closed form expressions for the rate functions

\[ \mathbb{E}[\alpha_i(Q)] = \lambda_i \cdot \Phi(\psi_i), \]  

(3.17)

\[ \mathbb{E}[\delta_i(Q)] = \mu_i \cdot \tau_i \cdot (q_i - \sqrt{q_i} \cdot (\varphi(\chi_i) - \chi_i \cdot \Phi(\chi_i))) + \beta_i \cdot \gamma_i \cdot \sqrt{q_i} \cdot (\varphi(\chi_i) - \chi_i \cdot \Phi(\chi_i)), \]  

(3.18)

\[ \mathbb{E}[D_{ij}(Q)] = \mu_i \cdot \tau_{ij} \cdot (q_i - \sqrt{q_i} \cdot (\varphi(\chi_i) - \chi_i \cdot \Phi(\chi_i))) \cdot \Phi(\psi_j) \]  

(3.19)

\[ + \beta_i \cdot \gamma_{ij} \cdot \sqrt{q_i} \cdot (\varphi(\chi_i) - \chi_i \cdot \Phi(\chi_i)) \cdot \Phi(\psi_j), \]

\[ \mathbb{E}[\tilde{D}_{ij}(Q)] = \mu_j \cdot \tau_{ji} \cdot (q_j - \sqrt{q_j} \cdot (\varphi(\chi_j) - \chi_j \cdot \Phi(\chi_j))) \cdot \Phi(\psi_i) \]  

(3.20)

\[ + \beta_j \cdot \gamma_{ji} \cdot \sqrt{q_j} \cdot (\varphi(\chi_j) - \chi_j \cdot \Phi(\chi_j)) \cdot \Phi(\psi_i), \]

and where we define

\[ \chi_i = \frac{c_i - q_i}{\sqrt{q_i}} \text{ and } \psi_i = \frac{c_i + k_i - q_i - 1}{\sqrt{q_i}} \]  

(3.21)
Proof. The proof of this theorem is similar to the one we will give for GVA, thus we omit it.

In Figure 2 we simulate the queueing network and compare the mean of the DMA and GISA methods. On the left and the right of Figure 2 we see that the GISA method outperforms the DMA method in approximating the mean dynamics of the network. This can be explained because the GISA approximation takes into account some of the stochastic fluctuations of the queueing network that DMA cannot.

Figure 2: Simulated, DMA, and GISA Means of $Q_1$ (Left). Simulated, DMA, and GISA Means of $Q_2$ (Right).

In Figure 3 we also compare the simulated variance to the mean of the GISA method since it assumes that the mean and variance are equal. We see on the left and right of Figure 3 that the GISA method is not accurate at estimating the variance dynamics of the queueing network since it assumes that the mean and variance are equal. We see that the variance is vastly overestimated by the GISA method and this can be explained by the loss feature of the queueing network. When the queueing network has a large number of customers arriving and being lost since the queue is at capacity, then the mean of the sample path of the queueing network does not change much and sits at the capacity level until the arrival rate starts to dip below the service and abandonment rates. Thus, when the queueing process constantly remains at the capacity level, the variance becomes very small and approaches zero. The GISA method is not able to estimate this type of queueing dynamics since this is exactly where the mean queue length is at the capacity level and non-zero if the number of server and waiting spaces is non-zero. However, if one is only interested in the mean dynamics of the loss network, then the GISA method will work quite well. To be able to estimate more than just the mean dynamics, we must develop another method that can capture the dynamics of the mean and variance separately. This is the subject of the next section.

3.3 Gaussian Variance Approximation

Our third approximation for the dynamics involves using the Gaussian distribution with separate mean and variance terms. To construct the GVA we follow the same procedure of
Figure 3: Simulated and GISA Variances of $Q_1$ (Left). Simulated and GISA Variances of $Q_2$ (Right).

[20]. We first let $X_1, X_2, ..., X_N$ be $N$ independent Gaussian(0,1) random variables and then we define recursively

$$Z_i = Z_{i-1} \cdot \cos \theta_{i-1} + X_i \cdot \sin \theta_{i-1} \quad (3.22)$$

where $Z_1 = X_1$. Using this recursive definition for the $Z_i$ terms, it is easily seen that the $Z_i$ random variables are distributed as Gaussian(0,1) random variables themselves. Thus, in order to construct the GVA for each individual station of the network we define

$$Q_i \overset{D}{=} q_i + \sqrt{v_i} \cdot Z_i. \quad (3.23)$$

This construction of the GVA gives us more flexibility and allows for different approximations for the mean and variance terms unlike the GISA. With the construction of the GVA, we are led to our next theorem.

**Theorem 3.3.** If we substitute the above approximate distribution in Equation 3.23 for each of the queue length processes of the nonstationary loss network with abandonment, we get the following differential equations for the mean, variance, and covariance matrix

$$\begin{align*}
\dot{E}[Q_i] &= E[\alpha_i(Q)] - E[\delta_i(Q)] - \sum_{j=1}^{N} E[D_{ij}(Q)] + \sum_{j=1}^{N} E[\tilde{D}_{ij}(Q)], \\
\text{Var}[Q_i] &= E[\alpha_i(Q)] + E[\delta_i(Q)] + \sum_{j=1}^{N} E[D_{ij}(Q)] + \sum_{j=1}^{N} E[\tilde{D}_{ij}(Q)] \\
&\quad + 2 \cdot \text{Cov}[Q_i, \alpha_i(Q)] - 2 \cdot \text{Cov}[Q_i, \delta_i(Q)] \\
&\quad - 2 \cdot \sum_{j=1}^{N} \text{Cov}[Q_i, \tilde{D}_{ij}(Q)] + 2 \cdot \sum_{j=1}^{N} \text{Cov}[Q_i, D_{ij}(Q)],
\end{align*}$$

(3.24) (3.25)
\[ 
\begin{align*}
\mathbf{Cov}[Q_i, Q_i] & = - \sum_{j=1}^{N} E[D_{ij}(Q)] - \sum_{j=1}^{N} E\left[ \tilde{D}_{ij}(Q) \right] \\
& + \mathbf{Cov}[Q_t, \alpha_i(Q)] + \mathbf{Cov}[Q_i, \alpha_i(Q)] - \mathbf{Cov}[Q_i, \delta_i(Q)] - \mathbf{Cov}[Q_t, \delta_i(Q)] \\
& + \sum_{j=1}^{N} \mathbf{Cov}[Q_i, D_{ij}(Q) - \tilde{D}_{ij}(Q)] - \sum_{j=1}^{N} \mathbf{Cov}[Q_t, D_{ij}(Q) - \tilde{D}_{ij}(Q)].
\end{align*} 
\tag{3.26} 
\]

where we have the following closed form expressions for the rate functions:

\[ 
E[\alpha_i(Q)] = \lambda_i \cdot \Phi(\psi_i), 
\]
\[ 
E[\delta_i(Q)] = \mu_i \cdot \tau_i \cdot (q_i - \sqrt{v_i} \cdot (\varphi(\chi_i) - \chi_i \cdot \Phi(\chi_i))) + \beta_i \cdot \gamma_i \cdot \sqrt{v_i} \cdot (\varphi(\chi_i) - \chi_i \cdot \Phi(\chi_i)), 
\]
\[ 
E[D_{ij}(Q)] \approx \mu_j \cdot \tau_{ji} \cdot (q_j - \sqrt{v_j} \cdot (\varphi(\chi_j) - \chi_j \cdot \Phi(\chi_j))) \cdot \Phi(\psi_j) + \beta_j \cdot \gamma_{ji} \cdot \sqrt{v_j} \cdot (\varphi(\chi_j) - \chi_j \cdot \Phi(\chi_j)) \cdot \Phi(\psi_j), 
\]
\[ 
E[\tilde{D}_{ij}(Q)] \approx \mu_j \cdot \tau_{ji} \cdot (q_j - \sqrt{v_j} \cdot (\varphi(\chi_j) - \chi_j \cdot \Phi(\chi_j))) \cdot \Phi(\psi_i) + \beta_j \cdot \gamma_{ji} \cdot \sqrt{v_j} \cdot (\varphi(\chi_j) - \chi_j \cdot \Phi(\chi_j)) \cdot \Phi(\psi_i), 
\]
\[ 
\mathbf{Cov}[Q_t, \alpha_i(Q)] = - \frac{v_i}{\sqrt{v_i}} \cdot \varphi(\psi_i), 
\]
\[ 
\mathbf{Cov}[Q_t, \delta_i(Q)] = \mu_i \cdot \tau_i \cdot \nu_i \cdot \Phi(\chi_i) + \beta_i \cdot \gamma_i \cdot \nu_i \cdot \Phi(\chi_i), 
\]
\[ 
\mathbf{Cov}[Q_t, D_{ij}(Q)] \approx \mu_i \cdot \tau_{ij} \cdot \nu_i \cdot \Phi(\chi_i) \cdot \Phi(\psi_j) + \beta_i \cdot \gamma_{ij} \cdot \nu_i \cdot \Phi(\chi_i) \cdot \Phi(\psi_j), 
\]
\[ 
\mathbf{Cov}[Q_t, \tilde{D}_{ij}(Q)] \approx \mu_j \cdot \tau_{ji} \cdot \nu_j \cdot \Phi(\chi_j) \cdot \Phi(\psi_i) + \beta_j \cdot \gamma_{ji} \cdot \nu_j \cdot \Phi(\chi_j) \cdot \Phi(\psi_i), 
\]

and where we now define:

\[ 
\chi_i = \frac{c_i - q_i}{\sqrt{v_i}} \quad \text{and} \quad \psi_i = \frac{c_i + k_i - 1 - q_i}{\sqrt{v_i}}, 
\]
\[ 
\tag{3.35} 
\]

**Proof.** Since many of the terms are similar to the paper of [20], we only give the terms that are not contained in the paper [20] and hence involve the loss feature of the network to show where the independence assumption is needed. Moreover, we do not give the proof of the terms that contain \( Q_i \wedge c_i \). This is because \( Q_i \wedge c_i = Q_i - (Q_i - c_i)^+ \) and therefore it suffices to provide the calculations for the \((Q_i - c_i)^+\) terms. The first rate function that we calculate is for the arrival process.
\[
E[\alpha_i(Q)] = \lambda_i \cdot E[\{Q_i \leq c_i + k_i - 1\}] \\
= \lambda \cdot E[\{Z_i \leq \psi_i\}] \\
= \lambda \cdot \mathbb{P}(\{Z_i \leq \psi_i\}) \\
= \lambda_i \cdot \Phi(\psi_i).
\]

The next rate function that we calculate is for the mean number of abandonments that leave station \(i\) and move to station \(j\). We use the independence assumption on the first line of the calculation. The rest simply exploits the fact that each \(Z_i\) is a standard Gaussian random variable.

\[
E[(Q_i - c_i)^+ \cdot \{Q_j \leq c_j + k_j - 1\}] \approx E[(Q_i - c_i)^+] \cdot E[\{Q_j \leq c_j + k_j - 1\}] \\
= E[\sqrt{v_i} \cdot (Z_i - \chi_i)^+] \cdot E[\{Z_j \leq \psi_j\}] \\
= \sqrt{v_i} \cdot (\varphi(\chi_i) - \chi_i \cdot \Phi(\chi_i)) \cdot \Phi(\psi_j).
\]

The next term that we calculate is for the covariance of the arrival rate at station \(i\) and the queue length of station \(j\).

\[
\text{Cov}[\alpha_i(Q), Q_j] = \lambda_i \cdot \text{Cov}[Q_i, \{Q_j \leq c_j + k_j - 1\}] \\
= \lambda_i \cdot \text{Cov}[\sqrt{v_i} \cdot Z_i, \{Z_j \leq \psi_j\}] \\
= \lambda_i \cdot \text{Cov}[\sqrt{v_i} \cdot Z_i, \{Z_j \leq \psi_j\}] \\
= -\frac{v_{ij}}{v_j} \cdot \varphi(\psi_j).
\]

Finally the last term that we calculate is the covariance of the queue length of station \(i\) with the abandonment process of station \(i\) and the arrival process of station \(j\). This has the following expression

\[
\text{Cov}[Q_i, (Q_i - c_i)^+ \cdot \{Q_j \leq c_j + k_j - 1\}] \\
= \text{Cov}[\sqrt{v_i} \cdot Z_i, \sqrt{v_i} \cdot (Z_i - \chi_i)^+ \cdot \{Z_j \leq \psi_j\}] \\
\approx \text{Cov}[\sqrt{v_i} \cdot Z_i, \sqrt{v_i} \cdot (Z_i - \chi_i)^+] \cdot E[\{Z_j \leq \psi_j\}] \\
= v_i \cdot \text{Cov}[Z_i, (Z_i - \chi_i)^+] \cdot \Phi(\psi_j) \\
= v_i \cdot \Phi(\chi_i) \cdot \Phi(\psi_j).
\]

In Figure 4 we see that the GVA method also approximates the mean of our network quite well. So we do not lose any performance of the mean. In fact the GVA approximation is slightly better than the GISA method as well. Furthermore in Figure 5 we see the true value of the GVA method since it is able to estimate the variance of the queueing network as well. In Figure 5, the estimates of the variance are quite accurate when compared to the Monte Carlo simulation of the queueing network. Although we can approximate the mean and variance quite well with the GVA method, it comes at an additional cost. Instead of
numerically integrating only $N$ equations for a $N$-dimensional network, we must integrate $\frac{N^2+3N}{2}$ differential equations. However, this is still very fast given the current speed of computers. Thus, we have seen that the GVA method can approximate the mean and variance of the queueing network with good accuracy, but what about other performance measures such as the blocking probability.

4 Approximating the Blocking Probability

Perhaps the most important performance measure of loss networks is the blocking probability at each station. In addition to estimating the mean and covariance of the loss network, we
will show that we can use the GISA and GVA methods to approximate the probability of blocking for each station in the network.

4.1 Probability of Blocking GISA

To construct the probability of blocking approximation, we must derive a formula for the probability of blocking for each station of the network. This can be written as the following expression

\[ P(Q_i > c_i + k_i - 1) = P(X > \psi_i) = \Phi(\psi_i) \]  (4.36)

where

\[ \psi_i = \frac{c_i + k_i - q_i - 1}{\sqrt{q_i}}. \]  (4.37)

4.2 Probability of Delay GVA

Similar to the GISA method, we can also derive an approximation for the blocking probabiltiy using the GVA. The probability of blocking for each station of the network is

\[ P(Q_i > c_i + k_i - 1) = P(X > \psi_i) = \Phi(\psi_i) \]  (4.38)

where

\[ \psi_i = \frac{c_i + k_i - q_i - 1}{\sqrt{v_i}}. \]  (4.39)

Thus, by using the Gaussian tail cdf function, we can approximate the probability of blocking for each station of the network using the GISA and GVA methods. In Figure 6 we see that we can provide good estimates for the probability of blocking for our loss network example. However, we see on both sides of Figure 6 that the GVA method is slightly more accurate at approximating the blocking probability. Thus, we see that both GISA and GVA are good at estimating other performance measures beyond the mean and variance of the network.

5 Additional Numerical Examples

In this section, we provide more numerical evidence to support our approximation methods. For Figures 7-9, we use the parameters given in Table 2 for a two dimensional loss network with abandonment. For Figures 10-12, we use Table 3.

In Figure 7 we see once again that the GVA method is very accurate at approximating the mean dynamics of the network and better than the DMA and GISA methods. This is also true of the variance in Figure 8 and the probability of blocking in Figure 9.

In Figure 10 we also see that the GVA method is quite accurate at approximating the mean behavior of the network and better than the DMA and GISA methods. This is also true of the variance in Figure 11 and the probability of blocking in Figure 12.
Figure 6: Simulated, GISA, GVA Blocking Probabilities of $Q_1$ (Left). Simulated, GISA, GVA Blocking Probabilities of $Q_2$ (Right).

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\lambda_1$</td>
<td>$20 + 10 \sin(t)$</td>
<td>$\lambda_2$</td>
<td>$30 + 15 \sin(t)$</td>
</tr>
<tr>
<td>$\mu_1$</td>
<td>1</td>
<td>$\mu_2$</td>
<td>1</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>.5</td>
<td>$\beta_2$</td>
<td>.5</td>
</tr>
<tr>
<td>$c_1$</td>
<td>20</td>
<td>$c_2$</td>
<td>30</td>
</tr>
<tr>
<td>$k_1$</td>
<td>5</td>
<td>$k_2$</td>
<td>5</td>
</tr>
<tr>
<td>$\tau_{11}$</td>
<td>0</td>
<td>$\tau_{22}$</td>
<td>0</td>
</tr>
<tr>
<td>$\gamma_{11}$</td>
<td>0</td>
<td>$\gamma_{22}$</td>
<td>0</td>
</tr>
<tr>
<td>$\tau_{12}$</td>
<td>.25</td>
<td>$\tau_{21}$</td>
<td>0</td>
</tr>
<tr>
<td>$\gamma_{12}$</td>
<td>.25</td>
<td>$\gamma_{21}$</td>
<td>0</td>
</tr>
</tbody>
</table>

6 Conclusion and Final Remarks

In this paper, we analyze the nonstationary loss network with abandonment. We present an approximation algorithm for the calculation of the mean, covariance matrix, and blocking probabilities of each station in the network. Simulation experiments are conducted to verify that our approximations are indeed accurate. To the best of our knowledge, such a study of the nonstationary loss network with abandonment has not been analyzed in the literature before.

We provide three different approximations that can be used to the discretion of the user. We find that the GVA method is most accurate, however, is grows quadratically with the network size. The GISA on the other hand only grows linearly and also provides accurate estimate of performance. The use of the GISA or GVA will depend on the time constraints of the user and their accuracy tolerance. However, we believe that GISA strikes a natural balance between accuracy, simplicity, and computational expense since it incorporates
stochastic elements into the approximation while still only grows linearly with the size of the network. What is also good about the GISA method, is that the functions are smooth as a function of the queue length and can be used for optimal control problems. We plan to use these differential equations for optimal control analysis in a future paper.

One potential idea for future work is to use the forward equations and multivariate Gaussian surrogate distribution to estimate the performance of nonstationary overflow networks studied in [7]. We believe that our method should also yield accurate results when applied in the overflow network setting. Moreover, we plan to extend our results to non-exponential distributions for service and abandonment distributions. This would enable us to model real networks that have non-Markovian behavior.
Figure 8: Simulated, GISA, and GVA Variances of $Q_1$ (Left). Simulated, GISA, and GVA Variances of $Q_2$ (Right).

References


Figure 9: Simulated, GISA, GVA Blocking Probabilities of $Q_1$ (Left). Simulated, GISA, GVA Blocking Probabilities of $Q_2$ (Right).


Figure 10: Simulated, DMA, GISA, and GVA Means of $Q_1$ (Left). Simulated, DMA, GISA, and GVA Means of $Q_2$ (Right).


Figure 11: Simulated, GISA, and GVA Variances of $Q_1$ (Left). Simulated, GISA, and GVA Variances of $Q_2$ (Right).

Figure 12: Simulated, GISA, GVA Blocking Probabilities of $Q_1$ (Left). Simulated, GISA, GVA Blocking Probabilities of $Q_2$ (Right).