EWGT 2012

15th meeting of the EURO Working Group on Transportation

An analytical approximation of the joint distribution of aggregate queue-lengths in an urban network

Carolina Osorio*, Carter Wang

Department of Civil and Environmental Engineering,
Massachusetts Institute of Technology, Cambridge, USA
*Email: osorioc@mit.edu

Abstract

Traditional queueing network models assume infinite queue capacities due to the complexity of capturing interactions between finite capacity queues. Accounting for this correlation can help explain how congestion propagates through a network. Joint queue-length distribution can be accurately estimated through simulation. Nonetheless, simulation is a computationally intensive technique, and its use for optimization purposes is challenging. By modeling the system analytically, we lose accuracy but gain efficiency and adaptability and can contribute novel information to a variety of congestion related problems, such as traffic signal optimization.

We formulate an analytical technique that combines queueing theory with aggregation-disaggregation techniques in order to approximate the joint network distribution, considering an aggregate description of the network. We propose a stationary formulation. We consider a tandem network with three queues.

The model is validated by comparing the aggregate joint distribution of the three queue system with the exact results determined by a simulation over several scenarios. It derives a good approximation of aggregate joint distributions.

Keywords: queueing theory, aggregation-disaggregation, congestion

1. Introduction

As congestion increases so does the dependency between links. Thus, it is important to model these dependencies in order to provide more accurate estimates of path, or network-wide, performance measures. This paper proposes a methodology to analytically approximate the joint distribution of queue-lengths over a network. The main challenges that arise in such an approach are the dimensionality and the complexity of modeling network-wide dependency analytically.

We focus on analytical approaches due to their computational efficiency and differentiability, which make them suitable to embed within traditional optimization frameworks, in order to address a variety of urban transportation problems. We consider a probabilistic setting and combine a queueing network model with an aggregation-disaggregation technique in order to approximate network-wide distributions.

Traditional queueing network models assume infinite queue (space) capacities due to the complexities of capturing interactions between finite capacity queues through blocking. Blocking occurs when a vehicle has completed its

service at a given queue, but cannot proceed downstream because the downstream queue is full. This is referred to as spillback in urban transportation. This phenomenon is not captured with infinite capacity queues, but is prevalent in congested networks. Finite capacity queues can allow us to describe spillbacks. Nonetheless, providing an analytical and tractable description of these is challenging [3].

The second main challenge is the dimensionality of the network-wide queue-length distribution. For a system of m queues, each with space capacity K (hereafter referred to as capacity), the state space of the joint system is $(K + 1)^m$. This is computationally intensive to solve exactly even for small values of m and K. In order to address this we resort to an aggregation-disaggregation mechanism. We simplify the state space of the full network by aggregating it, reducing the dimensionality of the joint distribution.

Section 3 presents the methodology. Section 4 validates this method versus simulated results, considering linear (i.e. tandem) topology networks and various demand/supply scenarios. Section 5 briefly presents the main conclusions and discusses ongoing and future work.

2. Literature review

A method for calculating steady state distributions of Markov chains was proposed by Takahashi [9]. The numerical method is appropriate when the state space can naturally be divided into disjoint sets, exploiting the structure of the chain for computational efficiency. This approach considers transitions between disaggregate states within an aggregate state as well as between aggregate states and individual disaggregate states.

Song and Takahashi [8] studied the application of the cross aggregation method, a nested family of approximate models that assumes different levels of dependence among queues, to tandem queueing systems of any size. By considering approximate models of subnetworks of various sizes, the method provides a series of stationary state probabilities for the subsystems, and relies on an assumption on the independence among nodes not in the same subsystem. The method requires computation that approximately scales linearly in the number of queues. Song and Takahashi examined tandem queueing systems with multiple servers and finite buffers, with exponentially distributed inter-arrival and service times, focusing on subsystems with one, two, or three queues. They conclude that the subsystem approach, which reduces the number of variables, can yield very accurate state probabilities for most situations. Specifically, in examining three queue subsystems, the relative error in both the marginal probabilities and the average number of customers were below 2% or 3%. Song and Takahashi do not aggregate within individual queues, an application explored by Dallery and Frein [2] and Boxma and Konheim [1].

A survey of aggregation-disaggregation techniques was done by Schweitzer [7]. He first summarizes the stationary aggregation-disaggregation approach, explaining the system of equations representing the stationary equilibrium of a large Markov-chain, which assumes that the state space is finite and large, as well as aperiodic and communicative. When exact aggregation and disaggregation equations are known, this distribution can be calculated by decomposing the system of equations and iteratively solving each set. He also specifies alternate formulations to the exact aggregation and disaggregation equations as well as iterative procedure used in solving the exact equations.

3. Model formulation

3.1. Aggregation-disaggregation framework

In order to address the dimensionality issues mentioned in the previous section, we use the aggregation technique described by Schweitzer [6]. This technique is formulated for both stationary and transient models. The aggregation-disaggregation techniques assume that the state space of the Markov chain is finite and large and the chain is aperiodic and communicative, which is the case of the urban transportation network models that we are considering.

The exact solution of the stationary distribution a large Markov chain with state space Ω is given by the global balance equations:

$$\pi_i \sum_{j \in \Omega \setminus i} \lambda_{ij} = \sum_{j \in \Omega \setminus i} \pi_j \lambda_{ji}, \tag{1}$$

and the normalizing constraint:

$$\sum_{i \in \Omega} \pi_i = 1, \tag{2}$$

where π_i is the probability of being in the disaggregate state i and λ_{ij} is the transition rate from disaggregate state i to disaggregate state j.

Schweitzer proposes partitioning the N states of the Markov chain into $\bar{N}$ aggregate states, such that $\bar{N} \ll N$, and $\Omega = \{1, 2, ..., N\} = \bigcup_{a=1}^{\bar{N}} \Omega_a$, where Ω represents the complete state space, Ω_a an aggregate state. The probability of being in a given aggregate state, $\bar{\pi}_a$, is defined as:

$$\bar{\pi}_a = \sum_{i \in \Omega_a} \pi_i, \quad 1 \leq a \leq \bar{N}.$$

At equilibrium, transition rates out of an aggregate state, a, equal the transition rates into a from aggregate states $b \neq a$, i.e.:

$$\bar{\pi}_a \sum_{b \neq a} \bar{\lambda}_{ab} = \sum_{b \neq a} \bar{\pi}_b \bar{\lambda}_{ba}, \quad 1 \leq a \leq \bar{N} \tag{3}$$

with the normalizing constraint:

$$\sum_{a=1}^{\bar{N}} \bar{\pi}_a = 1. \tag{4}$$

where, $\bar{\lambda}_{ab}$ is the transition rate from aggregate state a to aggregate state b.

The aggregate transition rates are related to the disaggregate transitions made from all disaggregate states in aggregate state a to all disaggregate states in aggregate state b, through:

$$\bar{\lambda}_{ab} = \frac{\sum_{j \in \Omega_a} \sum_{i \in \Omega_b} \pi_j \lambda_{ji}}{\sum_{k \in \Omega_a} \pi_k}, \quad b \neq a. \tag{5}$$

The transition rate into and out of a disaggregate state i must also be equal, i.e.:

$$\pi_i \sum_{j \in \Omega \setminus i} \lambda_{ij} = \sum_{j \in \Omega_a \setminus i} \pi_j \lambda_{ji} + \sum_{b \neq a} \bar{\pi}_b \hat{\lambda}_{bi}, \quad i \in \Omega_a, \quad 1 \leq a \leq \bar{N}, \tag{6}$$

where disaggregate state i resides in the aggregate state a, and $\hat{\lambda}_{bi}$ is the transition rate from aggregate state b to disaggregate state i. This rate is given by:

$$\hat{\lambda}_{bi} = \frac{\sum_{j \in \Omega_b} \pi_j \lambda_{ji}}{\sum_{k \in \Omega_b} \pi_k}. \tag{7}$$

Equation (6) sets the transitions out of disaggregate state $i \in \Omega_a$ equal to the sum of the transitions made into i from $j \in \Omega_a$, $j \neq i$ and from aggregate states $b \neq a$ into i. Schweitzer suggests to iteratively solve the Equations (3) and (6) for the stationary distribution of the system.

We illustrate Schweitzer's approach for a single finite capacity M/M/1/K queue and consider stationary analysis. An M/M/1/K has service and inter-arrival times distributed as exponential variables and a finite capacity queue with a single server. In this paper, we will present the formulation for tandem topology networks. The generalization to finite capacity queueing networks with arbitrary topology is straightforward. We describe the state of the single queue by the number of users or vehicles in the queue. The state space is thus given by $\Omega = \{0, 1, .., K\}$. The corresponding state transition diagram is displayed in Figure 1. Each circle denotes a state. The arrows denote possible transitions between the states, with their corresponding rates. In this case, arrivals are determined by the arrival rate, λ, and departures are determined by the service rate, μ. Assume we would like to aggregate the $K + 1$ states into the following three states: the queue is empty, the queue is full, or neither. The choice of these three states is based on insights from vehicle traffic node models, where between-link interactions are mainly determined based on whether a vehicle is ready to be sent downstream (i.e. non-empty queue) and whether there is space downstream to receive this vehicle (i.e. non-full queue). The new aggregate states are depicted in Figure 2. There are now 3 aggregate states, state 0, state K, and the state defined by the dashed line in Figure 2. The aggregate system is now fully described by a set of four rates: λ, μ, $\bar{\lambda}$, and $\bar{\mu}$, where $\bar{\mu}$ and $\bar{\lambda}$ describe the transition rates from the new aggregate state to one of the other states. In Figure 3, the aggregate states 0, 1 and 2, denote, respectively, the disaggregate states 0, $\{1, ..., K-1\}$, and K.
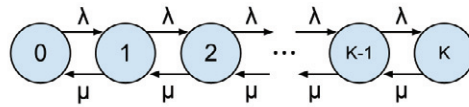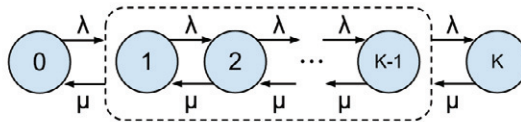
Figure 1: Single queue

Figure 2: Single aggregate queue

3.2. Stationary formulation for a single queue

In this section, we apply the aggregation-disaggregation technique to a single queue. This will give us insights into how to approximate the aggregate transition rates, so as to derive accurate aggregate distributions. For a single queue, as in Figure 1, given an external arrival rate, $\lambda \geq 0$, service rate, $\mu > 0$, and queue capacity, $K \in \mathbb{Z}^+$, the steady state queue-length probabilities are known exactly and are given by:

$$\pi_n = Pr\{N = n\} = \frac{(1 - \rho)\rho^n}{1 - \rho^{K+1}}, \quad \forall n \in [0, K] \tag{8}$$

where $\rho = \lambda/\mu$, and N is the random variable that denotes the number of users in the queue (i.e. total number of users or vehicles in the system).

As described in Section 2.1, the $K + 1$ states can be aggregated into three different states. The aggregate state is then described by the random variable N_A. Figure 3 represents this aggregated system. If the queue is empty ($\Omega_0 = \{N = 0\}$), then $N_A = 0$. If it is neither empty nor full ($\Omega_1 = \{N \in [1, K - 1]\}$), $N_A = 1$. If the queue is full ($\Omega_2 = \{N = K\}$), then $N_A = 2$. The probabilities of these aggregate states are denoted by $\bar{\pi}_i = Pr\{N_A = i\} = \sum_{j \in \Omega_i} \pi_j$. In steady state these probabilities satisfy the global balance equations, which are given by:

$$\begin{aligned} \lambda \bar{\pi}_0 &= \bar{\mu} \bar{\pi}_1 \\ \mu \bar{\pi}_2 &= \bar{\lambda} \bar{\pi}_1 \\ \sum_i \bar{\pi}_i &= 1 \end{aligned} \tag{9}$$

where $\bar{\mu}$ and $\bar{\lambda}$ are given by (5):

$$\bar{\lambda} = \frac{\sum_{j \in \Omega_1} \sum_{i \in \Omega_2} \pi_j \lambda_{ji}}{\sum_{k \in \Omega_1} \pi_k} = \frac{\sum_{j \in \Omega_1} \pi_j \lambda_{j,K}}{\sum_{k \in \Omega_1} \pi_k} \tag{10}$$

$$\bar{\mu} = \frac{\sum_{j \in \Omega_1} \sum_{i \in \Omega_0} \pi_j \mu_{ji}}{\sum_{k \in \Omega_1} \pi_k} = \frac{\sum_{j \in \Omega_1} \pi_j \mu_{j,0}}{\sum_{k \in \Omega_1} \pi_k}, \tag{11}$$

since Ω_0 and Ω_2 are associated with the disaggregate states $\{0\}$ and $\{K\}$. In the one-queue case, this implies:

$$\lambda_{i,K} = \begin{cases} 0 & \text{if } i \neq \{K - 1\} \\ \lambda & \text{if } i = \{K - 1\} \end{cases}$$
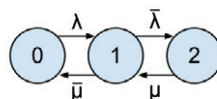

Figure 3: Simplified single aggregate queue

and:

$$\mu_{i,0} = \begin{cases} 0 & \text{if } i \neq \{1\} \\ \mu & \text{if } i = \{1\} \end{cases}$$

This implies, from (6) and (7) that:

$$\bar{\lambda} = \frac{\pi_{K-1}\lambda}{\sum_{k \in \Omega_1} \pi_k} = \lambda \frac{Pr\{N = K - 1\}}{Pr\{N_A = 1\}} = \lambda\, Pr\{N = K - 1 \mid N_A = 1\} \tag{12}$$

$$\bar{\mu} = \frac{\pi_1\mu}{\sum_{k \in \Omega_1} \pi_k} = \mu \frac{Pr\{N = 1\}}{Pr\{N_A = 1\}} = \mu\, Pr\{N = 1 \mid N_A = 1\}. \tag{13}$$

These equations state that the aggregate transition rates are the disaggregate transition rates, μ and λ, multiplied by the conditional probability that the queue is in the adjacent disaggregate state, given that it is in the aggregate state 1. In the case of a single queue, we can solve for $\bar{\lambda}$ and $\bar{\mu}$ using (8). Thus, the exact joint aggregate distribution, represented by the linear system of equations in (9), can be solved.

In general, we denote:

$$\alpha^e = Pr\{N = 1 \mid N_A = 1\} \tag{14}$$

$$\alpha^f = Pr\{N = K - 1 \mid N_A = 1\}. \tag{15}$$

3.3. Stationary formulation for a tandem network

A network of M/M/1/K queues presents additional problems not faced in the single queue scenario. If the system has two queues in tandem, then service completions by the upstream queue may be blocked by a full downstream queue (i.e. spillbacks may occur). This implies that the service rates of the upstream queue will change if it is blocked by the downstream queue, affecting both the aggregate and disaggregate transition rates. Thus, we need to approximate the conditional probabilities described in the previous section (Equations (14) and (15)) as well as the corresponding service rates (that may vary depending on whether there is blocking, and if so which queue is at the source of the blocking).

We consider a system of three linear (i.e. tandem) queues, with external arrivals to the first queue and external departures solely from the third queue. This is the simplest topology where a queue is affected by both upstream and downstream traffic conditions. We denote the external arrival rate to the first queue, $\gamma \geq 0$, service rates, $\mu_i > 0$, and queue capacities, $K_i \in \mathbb{Z}^+$, for each queue $i \in \{1, 2, 3\}$, where 1 (resp. 3) denotes the most upstream (resp. downstream) queue.

The possible aggregate and disaggregate states of a queue are given in Table 1. A queue is in state 0 if it is empty, 2 if it is full and not blocked, and 1 if it is not blocked and neither full nor empty. The subscript represents a blocked queue, which occurs when a queue has a service completion and the downstream queue is full. Accounting for blocking captures additional detail about the effective service rates, (the realized rate of service, rather than the rate the queue is capable of serving at) that would otherwise be lost.

Table 1: Mapping from disaggregate to aggregate states

Aggregate (N)	Disaggregate (N_A)	
0	$\{0\}$	empty
1	$[1, K_i - 1]$	not blocked, neither empty nor full
2	$\{K_i\}$	not blocked and full
1_B	$[1, K_i - 1]_B$	blocked, neither empty nor full
2_B	$\{K_i\}_B$	blocked and full

Let N_i denote the random variable that represents the number of users in each queue, distinguishing between whether the queue is blocked or not. The disaggregate joint distribution of the three queue system is denoted by $Pr\{N_1 = i, N_2 = j, N_3 = k\}$, $\forall\, i \in \{0, 1, 1_B, ..., K_1, K_{1,B}\}$, $j \in \{0, 1, 1_B, ..., K_2, K_{2,B}\}$, $k \in \{0, 1, ..., K_3\}$.

The aggregate three queue system is represented by a set of states $s = (i, j, k)$, where $i, j \in \{0, 1, 1_B, 2, 2_B\}$, $k \in \{0, 1, 2\}$. The random variable $N_{i,A}$ represents the aggregate number of users in queue i, as well as whether or note the queue is blocked, such that the aggregate joint distribution of the system is given by:

$$\bar{\pi}_s = Pr\{N_{1,A} = i, N_{2,A} = j, N_{3,A} = k\}, \quad i, j \in \{0, 1, 1_B, 2, 2_B\}, k \in \{0, 1, 2\}.$$

There are a total of 41 possible joint aggregate states. Thus, the dimension of the state space is now independent of the space capacity of the individual queues.

If an individual queue, i, is in the aggregate state 0, 2, or 2_B, then the disaggregate state of the queue is known (see Table 1). If it is in aggregate state 1 or 1_B, then the disaggregate state is in the interval $[1, K_i - 1]$. Transitions from state 1 or 1_B only occur when the disaggregate state is either 1 or $K_i - 1$. Given that the queue is in the aggregate state 1 or 1_B, we need to approximate the conditional probability that it is in the disaggregate state 1 or $K_i - 1$, (see Equations (12) and (13)). Thus we need to approximate the parameters α^e and α^f, defined in Equations (14) and (15). Assume that the marginal disaggregate queue length distributions have the functional form of Equation (8). Then α^e and α^f can be approximated as follows, note that the queue index i has been dropped for clarity:

$$\alpha^e = Pr\{N = 1 \mid N_A = 1\} = \frac{Pr\{N = 1, N_A = 1\}}{Pr\{N_A = 1\}} = \frac{Pr\{N = 1\}}{Pr\{N_A = 1\}}$$

$$\alpha^f = Pr\{N = K - 1 \mid N_A = 1\} = \frac{Pr\{N = K - 1, N_A = 1\}}{Pr\{N_A = 1\}} = \frac{Pr\{N = K - 1\}}{Pr\{N_A = 1\}}$$

where, for $\rho = \frac{\lambda}{\mu}$:

$$Pr\{N = 1\} = \frac{(1 - \rho)\rho}{1 - \rho^{K+1}}, \quad Pr\{N = K - 1\} = \frac{(1 - \rho)\rho^{K-1}}{1 - \rho^{K+1}}$$

$$Pr\{N_A = 1\} = \sum_{j=1}^{K-1} \frac{(1 - \rho)\rho^j}{1 - \rho^{K+1}}$$

Then:

$$\alpha^e = \frac{\frac{(1-\rho)\rho}{1-\rho^{K+1}}}{\sum_{j=1}^{K-1} \frac{(1-\rho)\rho^j}{1-\rho^{K+1}}} = \sum_{j=1}^{K-1} \frac{\rho}{\rho^j} = \sum_{j=0}^{K-2} \frac{1}{\rho^j} = \frac{1 - \rho}{1 - \rho^{K-1}} \tag{16}$$

$$\alpha^f = \frac{\frac{(1-\rho)\rho^{K-1}}{1-\rho^{K+1}}}{\sum_{j=1}^{K-1} \frac{(1-\rho)\rho^j}{1-\rho^{K+1}}} = \sum_{j=1}^{K-1} \frac{\rho^{K-1}}{\rho^j} = \rho^{K-2} \sum_{j=0}^{K-2} \frac{1}{\rho^j} = \frac{(1 - \rho)\rho^{K-2}}{1 - \rho^{K-1}} \tag{17}$$

Thus, α^e and α^f depend on K and on ρ. This functional form is clearly a good approximation in light traffic conditions, where the queues in a network behave as if they are independent. In order to also use this to accurately approximate the conditional probabilities under congested traffic, we will propose approximations for ρ. For a given aggregate state, we define ρ as the ratio between the prevailing arrival rate and the prevailing service rate. These prevailing rates vary depending on whether a queue is blocked and which queue it is being blocked by. Our model distinguishes between six different cases to determine these rates, as listed in Table 2 and detailed in the next paragraph.

Recall that α^e (resp. α^f) represents the conditional probability of a queue being in disaggregate state 1 (resp. $K - 1$) given that it is in aggregate state 1 or 1_B. To approximate this conditional probability, we enumerate the possible states of the downstream queues.

Lines 1 through 3 of Table 2 consider the cases where the first queue is not blocked, is blocked by the second queue, or is blocked by the third queue, respectively. These three cases have separate prevailing service rates, μ_1, μ_2, and μ_3, respectively. For example, line 3 indicates that if the first queue is blocked and the second and third queues are full, then we assume that the third queue is blocking both queues 1 and 2 and thus the prevailing service rate is that of the third queue (i.e. the third queue is the active bottleneck). In all three cases (lines $1 - 3$), the prevailing arrival rate is the external arrival rate, γ.

Table 2: Values of λ and μ, conditioned on the state of the downstream queues

	Conditions	Prevailing arrival rate	Prevailing service rate
Queue 1			
α_1	$N_{2,A} \neq 2$	γ	μ_1
α_2	$N_{2,A} = 2, N_{3,A} \neq 2$	γ	μ_2
α_3	$N_{2,A} = 2, N_{3,A} = 2$	γ	μ_3
Queue 2			
α_4	$N_{3,A} \neq 2$	$Pr\{N_{1,A} \neq 0\}\mu_1$	μ_2
α_5	$N_{3,A} = 2$	$Pr\{N_{1,A} \neq 0\}\mu_1$	μ_3
Queue 3			
α_6	No conditions	$Pr\{N_{2,A} \neq 0\}\mu_2$	μ_3

Lines 4 and 5 of Table 2 consider the cases where the second queue is not blocked and blocked by the third queue. The prevailing arrival rate to the second queue for an arbitrarily small time interval is approximated by the service rate of the first queue multiplied by the probability that the first queue is nonempty. The prevailing service rate is conditioned on the state of the third queue. If it is not full, then we say that the second queue serves at rate μ_2, otherwise it serves at rate μ_3.

Line 6 of Table 2 states that the third queue cannot be blocked. The prevailing arrival rate to the third queue is approximated for an arbitrarily small time interval as the service rate of the second queue multiplied by the probability that the second queue is nonempty.

The stationary aggregate joint distribution, $\bar{\pi}$ is obtained by solving the nonlinear system of equations given by the global balance equations:

$$\bar{\pi}Q = 0, \tag{18}$$

where Q is the transition matrix, which contains the rates at which there are transitions between aggregate states. Element $Q_{i,j}$ is the transition rate from state i to state j. The transition rate matrix is a described by the external arrival rate (γ), the service rates (μ), and the parameters α^e and α^f, which themselves depend on γ, μ, and K.

4. Experiments

Table 3: Experiments

Scenario	1	2	3	4	5	6	7	8	9
γ	1.8	1.8	1.8	1.8	1.8	1.8	1.8	1.8	1.8
μ	{2,2,2}	{2,2,2}	{2,2,2}	{6,4,2}	{6,4,2}	{6,4,2}	{2,3,6}	{2,3,6}	{2,3,6}
K	{2,2,2}	{5,5,5}	{10,10,10}	{2,2,2}	{5,5,5}	{10,10,10}	{2,2,2}	{5,5,5}	{10,10,10}

We compare the results of the stationary model with the results given by a discrete event simulator over nine scenarios with differing service rates and queue capacities, shown in Table 3. For the stationary case, we consider 100,000 separate replications, each with a run time of 10,000. The model results are simplified into 27 states, such that the aggregate states {1,2} count as both blocked and unblocked. The error is calculated by determining the Euclidean norm of the difference between the simulation and analytical model predictions.

For scenarios 1, 4, and 7, with queue capacities of {2,2,2}, the aggregate state 1 maps directly onto one state, and thus the approximations of α^e and α^f are exact and equal to 1. In these cases, the errors between the model and simulation results are very small, see Figure 4, and fall within the error bars. Thus errors in other scenarios may be attributed to approximations of α^e and α^f.

Outside of scenarios with queue capacities of {2,2,2}, the model also approximates scenarios 5 and 6 very well. Scenarios 5 and 6 have increasing service rates of {2,4,6} and queue capacities of {5,5,5} and {10,10,10}, respectively. In these scenarios, any arrival to a full first queue is lost, and the service rates increase as a user moves through the system, reducing the chances that the second or third queue becomes full and causes spillback. As a result, the system
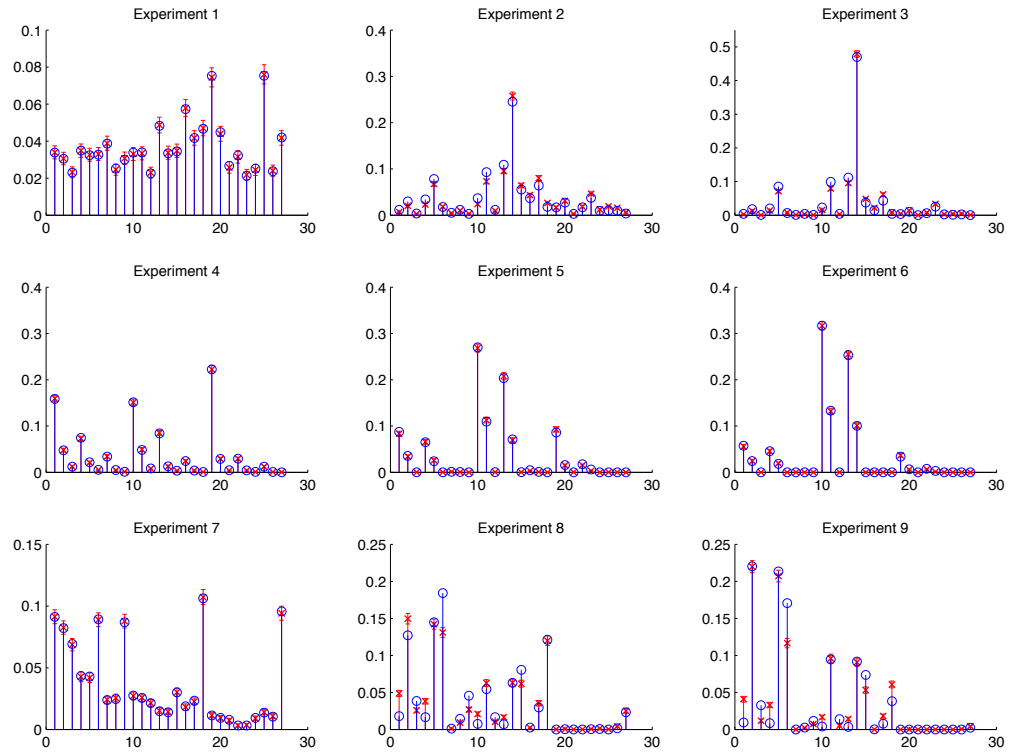
Figure 4: Stationary distributions for all scenarios. The blue circles represent the model predictions, and the red crosses represent the simulation results, with error bars for the simplified 27 states.

is closer in behavior to three separate finite capacity queues. Since the method utilizes the known distribution of a single finite capacity queue, it accurately models these scenarios.

In scenarios 2 and 3, which have constant service rates of {2,2,2} and queue capacities of {5,5,5} and {10,10,10}, our model provides a very good approximation. The model overestimates the probability of the first queue being empty, states {1 − 9}, and underestimates the first queue being full, states {19 − 27}.

Scenarios 8 and 9 are the least accurate estimations. These scenarios have decreasing service rates of {6,4,2} and queue capacities of {5,5,5} and {10,10,10}, so that blocking is mostly likely to occur as a result of the third queue. The resulting error is approximately 0.08, and likely due to the increased probability of blocking states when compared to the other seven scenarios.

5. Conclusion

We proposed an analytical technique that approximates the stationary joint aggregate distribution of three tandem queues. The method combines a queueing network model with an aggregation/disaggregation technique. The model was validated versus simulated results. These methods are currently being extended to tandem networks of any length by considering subsystems of three queues and imposing consistency between overlapping subsystems using the cross aggregation method, similar to that proposed by Song and Takahashi [8]. We can also incorporate external arrivals to multiple queues, providing a more similar comparison to actual road networks. The method currently under development also allows for external arrivals and departures to occur at any queue of the network. We are currently using this model along with its urban traffic formulation, defined by Osorio and Bierlaire [3], to address a traffic signal control problem, and to evaluate the added value of accounting for between-queue dependencies.

References

[1] Boxma, O., and A. Konheim, "Approximate analysis of exponential queueing systems with blocking", *Acta Informatica*, Vol. 15, 19-66, 1981.

[2] Dallery, Y., and Y. Frein. "On decomposition methods for tandem queueing networks with blocking", *Operations Research*, Vol. 41, No. 2, 386-399, 1993.

[3] Osorio, C. and M. Bierlaire. "An analytic finite capacity queueing network model capturing the propagation of congestion and blocking." *European Journal of Operational Research*, 196, pp. 996-1007, 2009.

[4] Osorio, C., et al. "Dynamic network loading: a stochastic differentiable model that derives link state distributions." *Transportation Research Part B (2011)*

[5] Rogers, D., R. Plante, R. Wong, and J. Evans. "Aggregation and disaggregation techniques and methodology in optimization", *Operations Research*, Vol. 39, No. 4, 553-582, 1991.

[6] Schweitzer, P., "Aggregation methods for large Markov chains", *Computer Performance and Reliability*, 275-286, 1984.

[7] Schweitzer, P., "A survey of aggregation-disaggregation in large Markov chains." *Numerical Solution of Markov Chains*, 63-88, 1991.

[8] Song, Y., and Y. Takahashi "Aggregate approximation for tandem queueing systems with production blocking." *Journal of the Operations Research*, 329-353, 1991.

[9] Takahashi, Y., "A lumping method for numerical calculations of stationary distributions of Markov chains." *Research Reports on Information Sciences, Series B: Operations Research*, 1975.

[10] Takahashi, Y., "A new type aggregation method for large Markov chains and its application to queueing networks." *ITC 11*, 1985.