

Strong Approximations for Time Varying Infinite-Server Queues with Non-Renewal Arrival and Service Processes

Young Myoung Ko

Department of Industrial and Management Engineering
Pohang University of Science and Technology
77, Cheongam-ro, Nam-gu, Pohang, Gyeongbuk, 37673, Korea
youngko@postech.ac.kr

Jamol Pender

School of Operations Research and Information Engineering
Cornell University
228 Rhodes Hall, Ithaca, NY 14853
jjp274@cornell.edu

October 26, 2017

Abstract

In real stochastic systems, the arrival and service processes may not be renewal processes. For example, in many telecommunication systems such as internet traffic where data traffic is bursty, the sequence of inter-arrival times and service times are often correlated and dependent. One way to model this non-renewal behavior is to use Markovian Arrival Processes (MAP's) and Markovian Service Processes (MSP's). MAP's and MSP's allow for inter-arrival and service times to be dependent, while providing the analytical tractability of simple Markov processes. To this end, we prove fluid and diffusion limits for $MAP_t/MSP_t/\infty$ queues by constructing a new Poisson process representation for the queueing dynamics and leveraging strong approximations for Poisson processes. As a result, the fluid and diffusion limit theorems illuminate how the dependence structure of the arrival or service processes can affect the sample path behavior of the queueing process. Finally, our Poisson representation for MAP's and MSP's is useful for simulation purposes and may be of independent interest.

Keywords: Infinite-Server Queues, Markovian Arrival Processes, Heavy Traffic, Fluid and Diffusion Limits, Strong Approximations, Asymptotics, Time-Varying Rates, Uniform Acceleration.

Mathematics Subject Classification: Primary 60K25, Secondary 68M20; 33C45

1 Introduction

Counting processes are important stochastic processes that have importance in many applications and areas of study such as biology, finance, telecommunications, and queueing theory. In the queueing literature, counting processes are used to model the arrival process of the queue and count the number of jobs or customers that arrive to the system during a specific time interval. Perhaps the most important of these counting processes is the Poisson process, which is the canonical arrival process. There are various extensions of the Poisson process such as the compound Poisson process, renewal processes, Markov modulated Poisson processes, and semi-Markov processes just to name a few. These processes are generalizations of the Poisson process and serve to capture real world phenomena seen in arrival traffic.

Although the Poisson process is a good process to model customer arrivals who are independent, the Poisson process is known to not be a great stochastic arrival process to model the arrivals of internet data traffic in telecommunication networks. Despite the Palm-Khintine theorem, which asserts that the superposition of a large number of renewal processes will converge to a Poisson process, it is well known in the teletraffic literature that the arrival traffic is not Poisson. In fact, it is also well known that the arrival traffic is also not renewal. See for example, Kang et al. [15], Yang and Tsang [40], Ren and Ramamurthy [36] where ATM statistical multiplexers superimpose many different kinds of traffic sources and the resulting arrival process is no longer renewal. Moreover, in applications like ridesharing or bikesharing, arrivals might not be renewal and are often bursty and dependent on events such as concerts or shows that are transpiring in a particular city. For more work in the context of queueing theory or fitting arrival processes see for instance, Horváth et al. [14], Shroff and Schwartz [38], Heindl and Telek [13], Andersen and Nielsen [1], Gerhardt and Nelson [8], Vatamidou et al. [39], Mészáros and Telek [23].

This is not the first work to study queues where the primitive random variables have dependence. Much of the current literature is in the area of queues with dependence is where the arrival and service processes are assumed to be positively dependent, see for example Hadidi [9]. In [9], the author assumes that the arrival and service processes in a single server queue follow a bivariate negative exponential distribution. Using the bivariate gamma distribution, they derive an integral equation for the waiting time in terms of the parameters of the bivariate gamma distribution. In a follow-up paper, Hadidi [10] derives closed form expressions for the moments greater than two of the waiting time when the arrival and service processes are dependent. However, the current literature only considers the single server setting and does not explore the multi-server or infinite server setting.

Recently, the literature on queues with dependence has expanded to the infinite server setting. For example, the work of Pang and Whitt [26, 27, 28] explores the impact of dependence on the nonstationary infinite server queue, especially when the successive service times are correlated. This type of dependence is often observed in recalls or inquiries about consumer products since customers are calling about the same issues and questions. The authors provide an approximate analysis of the mean and variance of the queue length as a function of the dependence between service times or between arrival times. They show that the correlation significantly impacts the variance but not the mean behavior. Thus, their work is exploring dependence between arrivals or between services rather than that between arrivals and services. O’Cinneide and Purdue [25] and D’Auria [5] also address dependence

in an infinite server setting. The authors introduce a (semi-)Markov environment process that affects arrival and service processes and obtain factorial moments of the number of customers in steady state. Unlike our approach that focuses on the transient behavior (on a compact interval), they obtain performance measures in steady state.

In this paper, we propose a new approach for analyzing queues where the arrival process has dependence. The first step is to model the arrival process with a Markovian Arrival Process (MAP). MAP's unlike phase type distributions, allow one to consider non-renewal processes for the arrival process. This is because a phase type distribution is restarted independently of its past history. MAP's generalize this feature and allow for dependence on the past history of the Markov chain. In a MAP, unlike phase type distributions, the next interarrival time is dependent on the exit state of the Markov chain and this feature allows one to introduce a notion of *memory* into the process. The ultimate goal is for us to describe the sample path behavior of queues that have a MAP as its arrival process. One main obstacle that we have is that there exists no obvious way to model a MAP using Poisson processes. Since a MAP is constructed from the absorbing times of Markov chains, one major contribution of this work is to develop a novel way of modeling the MAP dynamics using unit rate Poisson processes. Given that we have a Poisson representation for the MAP, we can combine it with Markovian Service Processes (MSP's) for the service process to construct $MAP_t/MSP_t/\infty$ queueing models. Once we have a Poisson representation for the queue length process, we derive fluid and diffusion limits taking advantage of the theory of strong approximations.

1.1 Main Contributions of Paper

The contributions of this work can be summarized as follows. First, we consider the time varying dynamics of the $MAP_t/MSP_t/\infty$ queues as an approximation of more general queueing processes with non-renewal arrival and service processes. Even with its Markovian structure, $MAP_t/MSP_t/\infty$ queueing models are relatively intractable as we cannot analytically solve for the exact distribution of the queue length as a function of time. As a result, we derive fluid and diffusion limits for the $MAP_t/MSP_t/\infty$ queues using strong approximations of time changed Poisson processes by increasing the number of independent and identical input sources, a.k.a. the many-sources regime described in Shakkottai and Srikant [37], Eun and Shroff [7], Cecchi et al. [3]. These limit theorems allow us to gain insight into the sample path behavior of the $MAP_t/MSP_t/\infty$ queue and provide estimates of performance measures such as the mean and variance of the queue length process. We partially extend the result of this paper to a network setting in the follow-up paper ([18]).

1.2 Organization of Paper

The remainder of this paper is organized as follows. Section 2 describes the construction of a MAP and its time-varying extension. Section 3 constructs Poisson representations and derives fluid and diffusion limits. Section 3.1 constructs a Poisson representation of MAP's. Using it, Section 3.2 obtains a Poisson representation of $MAP_t/MSP_t/\infty$ queues. We also prove fluid and diffusion limits for this system. Lastly, Section 4 concludes and offers suggestions

for future research and the Appendix provides all of the proofs and derivations of our limit theorems.

2 Markovian Arrival Processes (MAP's)

In this section, we begin with describing MAP's. MAP's, unlike phase type distributions, allow one to model the dynamics of non-renewal processes to use a counting processes or inter-arrival process. Unlike the phase type distribution, which restarts independently of its past history, a MAP restarts dependent on the exit state of the Markov chain depends on the past history of the chain.

We follow the construction of a MAP in Chakravarthy [4]. We consider an m -state irreducible continuous time Markov chain (CTMC). After spending $\exp(\lambda_i)$ amount of time in state i , there are two possible transitions from state i to state j . The first possible transition to state j (including the case of $i = j$) happens with probability p_{ij} and incurs an event or arrival. The second possible transition to state j (not allowing the case of $i = j$) corresponds to no arrival and occurs with probability q_{ij} . Then, we define matrices $[\mathcal{D}_0]_{ij} = d_{ij}^0$ and $[\mathcal{D}_1]_{ij} = d_{ij}^1$ where $d_{ii}^0 = -\lambda_i$, $1 \leq i \leq m$; $d_{ij}^0 = \lambda_i q_{ij}$, $j \neq i$, $1 \leq i, j \leq m$; $d_{ij}^1 = \lambda_i p_{ij}$, $1 \leq i, j \leq m$, with $\left(\sum_{j=1}^m p_{ij} + \sum_{j \neq i} q_{ij}\right) = 1$, for $1 \leq i \leq m$. In our description of the MAP, we have suppressed its dependence on time. However, all of our results apply to the time varying setting and we explain it in Section 2.2.

With the above construction, a MAP is described by the two $m \times m$ matrices \mathcal{D}_0 and \mathcal{D}_1 . The matrix \mathcal{D}_0 where $[\mathcal{D}_0]_{ij} = d_{ij}^0$ corresponds to transitions where there is no arrival and the matrix \mathcal{D}_1 where $[\mathcal{D}_1]_{ij} = d_{ij}^1$ corresponds to the transitions that generate an actual arrival. With this construction, it is also obvious why this is more general than phase type distributions. Dependence is created by the fact that when an arrival is generated, then the Markov chain can re-enter the same state, however, when no arrival is generated, it cannot re-enter the same state. Now that we have defined a MAP, it is now important to understand how the MAP is a generalization of some well known stochastic arrival processes.

2.1 Versatility of the MAP

The MAP is a very versatile process for modeling arrival processes that are ubiquitous in service systems or queueing theory. There are various special cases of MAP's that are important for many queueing processes and we outline some of them below in the one and two dimensional cases.

$$\text{Poisson Process : } \mathcal{D}_0 = [-\lambda], \quad \mathcal{D}_1 = [\lambda]$$

$$\text{Erlang Renewal Process : } \mathcal{D}_0 = \begin{pmatrix} -\lambda & \lambda \\ 0 & -\lambda \end{pmatrix}, \quad \mathcal{D}_1 = \begin{pmatrix} 0 & 0 \\ \lambda & 0 \end{pmatrix}$$

$$\text{Hyperexponential Renewal Process : } \mathcal{D}_0 = \begin{pmatrix} -\lambda_1 & 0 \\ 0 & -\lambda_2 \end{pmatrix}, \quad \mathcal{D}_1 = \begin{pmatrix} p_1 \lambda_1 & p_2 \lambda_1 \\ p_1 \lambda_2 & p_2 \lambda_2 \end{pmatrix},$$

$$p_1 + p_2 = 1$$

$$\text{Markov Modulated Poisson Process : } \mathcal{D}_0 = \begin{pmatrix} -\lambda_1 & \lambda_{1,2} \\ \lambda_{2,1} & -\lambda_2 \end{pmatrix}, \quad \mathcal{D}_1 = \begin{pmatrix} \lambda_{1,1}^* & 0 \\ 0 & \lambda_{2,2}^* \end{pmatrix}$$

$$\text{Interrupted Poisson Process : } \mathcal{D}_0 = \begin{pmatrix} -\lambda_1 & \lambda_{1,2} \\ \lambda_{2,1} & -\lambda_2 \end{pmatrix}, \quad \mathcal{D}_1 = \begin{pmatrix} \lambda_{1,1}^* & \lambda_{1,2}^* \\ 0 & 0 \end{pmatrix}$$

$$\text{Acyclic MAP(2) : } \mathcal{D}_0 = \begin{pmatrix} -\lambda_1 & \lambda_{1,2} \\ 0 & -\lambda_2 \end{pmatrix}, \quad \mathcal{D}_1 = \begin{pmatrix} \lambda_{1,1}^* & 0 \\ \lambda_{2,1}^* & \lambda_{2,2}^* \end{pmatrix}$$

$$\text{MAP(2) : } \mathcal{D}_0 = \begin{pmatrix} -\lambda_1 & \lambda_{1,2} \\ \lambda_{2,1} & -\lambda_2 \end{pmatrix}, \quad \mathcal{D}_1 = \begin{pmatrix} \lambda_{1,1}^* & \lambda_{1,2}^* \\ \lambda_{2,1}^* & \lambda_{2,2}^* \end{pmatrix}$$

Beyond the fact that the MAP generalizes many well known arrival processes, it is also important to note that the MAP is still a Markov process. Therefore, much of the analysis of Markov processes is still applicable in the MAP setting. The only drawback is that fitting a MAP is slightly more involved than a Poisson process and the MAP is high dimensional when there are a large number of transient states. However, with the added dimensionality, it allows for more flexibility in fitting and more realistic performance approximations for stochastic models in practice. Now that we have defined a MAP, it is important to understand how to derive a Poisson process representation for the MAP. This Poisson representation of the MAP is integral to proving the fluid and diffusion limits of the $MAP_t/MSP_t/\infty$ queues since we will leverage strong approximations developed for time changed Poisson processes Kurtz [19], Mandelbaum et al. [20], Hampshire et al. [11]. The Poisson representation naturally allows time varying rate functions. We, therefore, explain the time varying extension of the MAP in the next section.

2.2 Time varying extension of the MAP

The MAP described in the previous section does not have a time varying structure. The extension to time varying parameters can be made by changing constant parameters into the functions of time, i.e., $d_{ij}^0 \rightarrow d_{ij}^0(t)$ and $d_{ij}^1 \rightarrow d_{ij}^1(t)$. However, in order to derive the limit theorems in Section 3.2, the time varying rates should be locally bounded as follows:

$$\int_0^T d_{ij}^0(t)dt < \infty, \quad \int_0^T d_{ij}^1(t)dt < \infty \quad \text{for } i, j \in \{1, \dots, m\} \text{ and } T < \infty.$$

In the rest of the paper, we will assume that all time varying rates are locally bounded with respect to time t .

3 Poisson Construction and limit theorems

In this section, we construct the Poisson representation for the $MAP_t/MSP_t/\infty$ queues and derive fluid and diffusion limits of the queue length process. We first explain the Poisson representation for MAP's in Section 3.1. Extending the representation, Section 3.2 constructs the Poisson representation of $MAP_t/MSP_t/\infty$ queues and derives fluid/diffusion limits for them.

3.1 Poisson Construction of MAP's

In this section, we describe the Poisson process construction of the MAP. Our construction uses Poisson processes since they are well studied and strong approximations for Poisson processes yield Brownian motion approximations. To this end, assuming m_A number of phases in the MAP, we let $U_j(t)$ be the number of customers in phase j of the MAP at time t . Since $U_j(t)$'s keep track of the current phase of the MAP, $\sum_{j=1}^{m_A} U_j(t) = 1$. Then, we construct a Poisson process representation for the MAP as follows:

$$\begin{aligned}
 U_j(t) &= U_j(0) + \sum_{k \neq j}^{m_A} \Pi_{kj}^{A0} \left(\int_0^t d_{kj}^{A0}(s) U_k(s) ds \right) + \sum_{k=1}^{m_A} \Pi_{kj}^{A1} \left(\int_0^t d_{kj}^{A1}(s) U_j(s) ds \right) \\
 &\quad - \sum_{k \neq j}^{m_A} \Pi_{jk}^{A0} \left(\int_0^t d_{jk}^{A0}(s) U_j(s) ds \right) - \sum_{k=1}^{m_A} \Pi_{jk}^{A1} \left(\int_0^t d_{jk}^{A1}(s) U_j(s) ds \right) \text{ for } 1 \leq j \leq m_A,
 \end{aligned} \tag{3.1}$$

Poisson processes, $\Pi_{kj}^{A0}(\cdot)$'s and $\Pi_{kj}^{A1}(\cdot)$'s count the phase transition from k to j without and with arrivals respectively. Note that we assume that the Poisson processes explained above have rate 1 (with random time changes) and are mutually independent. One way to view the Poisson construction of the MAP is to view it as a token moving across the phases since we have that $\sum_{j=1}^{m_A} U_j(t) = 1$. Sometimes, it will not generate an arrival and when an arrival is not generated, the continuous time Markov chain must move to different phase than it is currently in. However, when an arrival is generated, the Markov chain can move to another state or stay in the same state according to the transition probabilities.

3.2 Poisson Construction of $MAP_t/MSP_t/\infty$ Queue

In this section, we construct a Poisson representation and derive the fluid and diffusion limits for the case where both inter-arrival and service times are not necessarily independent and identically distributed random variables. Dependent service times are very practical and often arise in telephone call centers when customers call about recalled or defective products. For example, the work of Pang and Whitt [26] explores the impact of dependence on the nonstationary infinite server queue, especially when the successive service times are correlated with one another. However, the combination of the non-renewal arrival and service times has not been explored in the literature in the infinite server context. Moreover, our approach lends itself to computational methods that have been developed for MAP's and MSP's. The $MAP_t/MSP_t/\infty$ has the following representation in terms of unit rate Poisson processes:

$$U_j(t) = \underbrace{U_j(0)}_{\text{Initial Value of Token}} + \underbrace{\sum_{k \neq j}^{m_A} \Pi_{kj}^{A0} \left(\int_0^t d_{kj}^{A0}(s) U_k(s) ds \right)}_{\text{MAP moves from state } k \text{ to } j \text{ (no arrival generated)}} \quad (3.2)$$

$$+ \underbrace{\sum_{k=1}^{m_A} \sum_{i=1}^{m_S} \Pi_{kji}^{A1} \left(\int_0^t d_{kj}^{A1}(s) \beta_{ji} U_k(s) ds \right)}_{\text{MAP moves from state } k \text{ to } j \text{ (arrival is placed in phase } i \text{ of service)}}$$

$$- \underbrace{\sum_{k \neq j} \Pi_{jk}^{A0} \left(\int_0^t d_{jk}^{A0}(s) U_j(s) ds \right)}_{\text{MAP moves from state } j \text{ to } k \text{ (no arrival generated)}}$$

$$- \underbrace{\sum_{k=1}^{m_A} \sum_{i=1}^{m_S} \Pi_{jki}^{A1} \left(\int_0^t d_{jk}^{A1}(s) \beta_{ki} U_j(s) ds \right)}_{\text{MAP moves from state } j \text{ to } k \text{ (arrival is placed in phase } i \text{ of service)}} \quad \text{for } 1 \leq j \leq m_A,$$

$$X_i(t) = \underbrace{\sum_{j=1}^{m_A} \sum_{k=1}^{m_A} \Pi_{jki}^{A1} \left(\int_0^t d_{jk}^{A1}(s) \beta_{ki} U_j(s) ds \right)}_{\text{MAP moves from state } j \text{ to } k \text{ (arrival is placed in phase } i \text{ of service)}} \quad (3.3)$$

$$+ \underbrace{\sum_{l \neq i}^{m_S} \Pi_{li}^{S0} \left(\int_0^t d_{li}^{S0}(s) X_l(s) ds \right)}_{\text{MSP moves from state } l \text{ to } i \text{ (no service completion)}}$$

$$- \underbrace{\sum_{l \neq i}^{m_S} \Pi_{il}^{S0} \left(\int_0^t d_{il}^{S0}(s) X_i(s) ds \right)}_{\text{MSP moves from state } i \text{ to } l \text{ (no service completion)}}$$

$$- \underbrace{\Pi_i^{S1} \left(\int_0^t d_i^{S1}(s) X_i(s) ds \right)}_{\text{MSP leaves the system from state } i \text{ (service completion)}} \quad \text{for } 1 \leq i \leq m_S.$$

In this construction, $X_i(t)$ represents the number of customers that are in phase i of the MSP at time t . This can be also interpreted as the number of customers that are in phase i of service. The probability vector, $\beta_j = (\beta_{j1}, \dots, \beta_{jm_S})$, is the initial distribution to the service process when the arrival is generated in phase j of the MAP. Moreover, Poisson processes, $\Pi_{il}^{S0}(\cdot)$'s count the number of transitions from phase i to phase l without service completions and Poisson processes, $\Pi_i^{S1}(\cdot)$'s count the number of service completions. For the remainder of the paper, we will use the following notation for the stochastic queue length process.

$$\mathbf{Q}(t) = (U_1(t), \dots, U_{m_A}(t), X_1(t), \dots, X_{m_S}(t))'. \\ \mathbf{q} = (u_1, \dots, u_{m_A}, x_1, \dots, x_{m_S})'.$$

- $f_{jk}^{A0}(t, \mathbf{q})$: rate function of the (integrand) in $\mathbf{\Pi}_{jk}^{A0}(\cdot)$,
 $f_{jki}^{A1}(t, \mathbf{q})$: rate function of the (integrand) in $\mathbf{\Pi}_{jki}^{A1}(\cdot)$,
 $f_{il}^{S0}(t, \mathbf{q})$: rate function of the (integrand) in $\mathbf{\Pi}_{il}^{S0}(\cdot)$,
 $f_i^{S1}(t, \mathbf{q})$: rate function of the (integrand) in $\mathbf{\Pi}_i^{S1}(\cdot)$
 \mathbf{l}_{jk}^{A0} : $(m_A + m_A) \times 1$ vector, j^{th} element is -1, k^{th} element is 1, and other elements are 0.
 \mathbf{l}_{jki}^{A1} : $(m_A + m_S) \times 1$ vector, j^{th} element is -1, k^{th} element is 1,
 $(m_A + i)^{\text{th}}$ element is 1, and other elements are 0.
 \mathbf{l}_{il}^{S0} : $(m_A + m_S) \times 1$ vector, $(m_A + i)^{\text{th}}$ element is -1, $(m_A + l)^{\text{th}}$ element is 1,
and other elements are 0.
 \mathbf{l}_i^{S1} : $(m_A + m_S) \times 1$ vector, $(m_A + i)^{\text{th}}$ element is -1, and other elements are 0.

Then, we can express equations (3.2) and (3.3) as follows:

$$\begin{aligned}
\mathbf{Q}(t) = & \mathbf{Q}(0) + \sum_{j=1}^{m_A} \sum_{k \neq j}^{m_A} \mathbf{l}_{jk}^{A0} \mathbf{\Pi}_{jk}^{A0} \left(\int_0^t f_{jk}^{A0}(s, \mathbf{Q}(s)) ds \right) \\
& + \sum_{j=1}^{m_A} \sum_{k=1}^{m_A} \sum_{i=1}^{m_S} \mathbf{l}_{jki}^{A1} \mathbf{\Pi}_{jki}^{A1} \left(\int_0^t f_{jki}^{A1}(s, \mathbf{Q}(s)) ds \right) + \sum_{i=1}^{m_S} \sum_{l \neq i}^{m_S} \mathbf{l}_{il}^{S0} \mathbf{\Pi}_{il}^{S0} \left(\int_0^t f_{il}^{S0}(s, \mathbf{Q}(s)) ds \right) \\
& + \sum_{i=1}^{m_S} \mathbf{l}_i^{S1} \mathbf{\Pi}_i^{S1} \left(\int_0^t f_i^{S1}(s, \mathbf{Q}(s)) ds \right).
\end{aligned}$$

Remark. If the β_j 's are the same for all $j \in \{1, \dots, m_A\}$, the Poisson representation in equations (3.2)-(3.3) is for the $MAP_t/Ph_t/\infty$ queue, a special case of the $MAP_t/MSP_t/\infty$ queue.

3.2.1 Fluid Limits

We prove fluid limits for the queue length process using the Poisson representation and strong approximations. We first define a sequence of processes $\{\mathbf{Q}^\eta(t), \eta \in \mathcal{N}, t \in \mathcal{R}+\}$ as

follows:

$$\begin{aligned}
\mathbf{Q}^\eta(t) &= \mathbf{Q}^\eta(0) + \sum_{j=1}^{m_A} \sum_{k \neq j}^{m_A} \mathbf{I}_{jk}^{A0} \mathbf{\Pi}_{jk}^{A0} \left(\int_0^t f_{jk}^{A0}(s, \mathbf{Q}^\eta(s)) ds \right) \\
&\quad + \sum_{j=1}^{m_A} \sum_{k=1}^{m_A} \sum_{i=1}^{m_S} \mathbf{I}_{jki}^{A1} \mathbf{\Pi}_{jki}^{A1} \left(\int_0^t f_{jki}^{A1}(s, \mathbf{Q}^\eta(s)) ds \right) + \sum_{i=1}^{m_S} \sum_{l \neq i}^{m_S} \mathbf{I}_{il}^{S0} \mathbf{\Pi}_{il}^{S0} \left(\int_0^t f_{il}^{S0}(s, \mathbf{Q}^\eta(s)) ds \right) \\
&\quad + \sum_{i=1}^{m_S} \mathbf{I}_i^{S1} \mathbf{\Pi}_i^{S1} \left(\int_0^t f_i^{S1}(s, \mathbf{Q}^\eta(s)) ds \right) \\
&= \mathbf{Q}^\eta(0) + \sum_{j=1}^{m_A} \sum_{k \neq j}^{m_A} \mathbf{I}_{jk}^{A0} \mathbf{\Pi}_{jk}^{A0} \left(\int_0^t \eta f_{jk}^{A0}(s, \bar{\mathbf{Q}}^\eta(s)) ds \right) \\
&\quad + \sum_{j=1}^{m_A} \sum_{k=1}^{m_A} \sum_{i=1}^{m_S} \mathbf{I}_{jki}^{A1} \mathbf{\Pi}_{jki}^{A1} \left(\int_0^t \eta f_{jki}^{A1}(s, \bar{\mathbf{Q}}^\eta(s)) ds \right) + \sum_{i=1}^{m_S} \sum_{l \neq i}^{m_S} \mathbf{I}_{il}^{S0} \mathbf{\Pi}_{il}^{S0} \left(\int_0^t \eta f_{il}^{S0}(s, \bar{\mathbf{Q}}^\eta(s)) ds \right) \\
&\quad + \sum_{i=1}^{m_S} \mathbf{I}_i^{S1} \mathbf{\Pi}_i^{S1} \left(\int_0^t \eta f_i^{S1}(s, \bar{\mathbf{Q}}^\eta(s)) ds \right) \text{ by the linearity of the functions } f(\cdot, \cdot)\text{'s,}
\end{aligned}$$

where we define $\bar{\mathbf{Q}}^\eta(t) = \mathbf{Q}^\eta(t)/\eta$.

Note that we increase the number of independent arrival processes (sources) from the MAP by setting $\sum_{j=1}^{m_A} U_j^\eta(t) = \eta$ for $t \geq 0$ instead of accelerating the transition rates (d_{kj}^{A0} 's and d_{kj}^{A1} 's). By doing that, we can directly find a sequence satisfying the initial condition for deriving the fluid limit. One should also note that this scaling is equivalent to scaling the rates where $\sum_{j=1}^{m_A} \bar{U}_j^\eta(t) = \sum_{j=1}^{m_A} U_j^\eta(t)/\eta = 1$ and the transitions rates are $\eta \cdot d_{ij}(t)$. The following theorem explains the fluid limits for the $MAP_t/MSP_t/\infty$ queue.

Theorem 3.1. *Suppose $\mathbf{Q}^\eta(0)/\eta \rightarrow \mathbf{q}(0)$ almost surely as $\eta \rightarrow \infty$, then*

$$\lim_{\eta \rightarrow \infty} \frac{1}{\eta} \mathbf{Q}^\eta(t) = \mathbf{q}(t) \text{ almost surely,}$$

where $\mathbf{q}(t) = (u_1(t), \dots, u_{m_A}(t), x_1(t), \dots, x_{m_S}(t))'$ is the solution to the following system of ordinary differential equations:

$$\begin{aligned}
\frac{d}{dt} \mathbf{q}(t) &= \sum_{j=1}^{m_A} \sum_{k \neq j}^{m_A} \mathbf{I}_{jk}^{A0} f_{jk}^{A0}(t, \mathbf{q}(t)) + \sum_{j=1}^{m_A} \sum_{k=1}^{m_A} \sum_{i=1}^{m_S} \mathbf{I}_{jki}^{A1} f_{jki}^{A1}(t, \mathbf{q}(t)) \\
&\quad + \sum_{i=1}^{m_S} \sum_{l \neq i}^{m_S} \mathbf{I}_{il}^{S0} f_{il}^{S0}(t, \mathbf{q}(t)) + \sum_{i=1}^{m_S} \mathbf{I}_i^{S1} f_i^{S1}(t, \mathbf{q}(t))
\end{aligned} \tag{3.4}$$

Proof. See Appendix. □

3.2.2 Diffusion Limits

With the fluid limit, $\mathbf{q}(t)$, derived in Section 3.2.1, we can derive the diffusion limit as follows:

Theorem 3.2. Let $\mathbf{D}^\eta(t) = \sqrt{\eta} \left(\frac{1}{\eta} \mathbf{Q}^\eta(t) - \mathbf{q}(t) \right)$ and suppose that $\sqrt{\eta} \left(\frac{1}{\eta} \mathbf{Q}^\eta(0) - \mathbf{q}(0) \right) \Rightarrow \mathbf{D}(0)$ in distribution as $\eta \rightarrow \infty$, then we have that

$$\lim_{\eta \rightarrow \infty} \mathbf{D}^\eta(t) = \mathbf{D}(t) \text{ in distribution,}$$

where $\mathbf{D}(t)$ is the solution to the following stochastic differential equation

$$d\mathbf{D}(t) = d\mathbf{H}(t, \mathbf{q}(t)) + \partial\mathbf{F}(t, \mathbf{q}(t))\mathbf{D}(t)dt, \quad (3.5)$$

and $\partial\mathbf{F}(t, \mathbf{q}(t))$ is the gradient matrix of $\mathbf{F}(t, \mathbf{q}(t))$ with respect to $\mathbf{q}(t)$. Moreover,

$$\begin{aligned} \mathbf{F}(t, \mathbf{q}(t)) &= \sum_{j=1}^{m_A} \sum_{k \neq j}^{m_A} \mathbf{l}_{jk}^{A0} f_{jk}^{A0}(t, \mathbf{q}(t)) + \sum_{j=1}^{m_A} \sum_{k=1}^{m_A} \sum_{i=1}^{m_S} \mathbf{l}_{jki}^{A1} f_{jki}^{A1}(t, \mathbf{q}(t)) \\ &\quad + \sum_{i=1}^{m_S} \sum_{l \neq i}^{m_S} \mathbf{l}_{il}^{S0} f_{il}^{S0}(t, \mathbf{q}(t)) + \sum_{i=1}^{m_S} \mathbf{l}_i^{S1} f_i^{S1}(t, \mathbf{q}(t)) \\ d\mathbf{H}(t, \mathbf{q}(t)) &= \sum_{j=1}^{m_A} \sum_{k \neq j}^{m_A} \mathbf{l}_{jk}^{A0} \sqrt{f_{jk}^{A0}(t, \mathbf{q}(t))} dW_{jk}^{A0}(t) + \sum_{j=1}^{m_A} \sum_{k=1}^{m_A} \sum_{i=1}^{m_S} \mathbf{l}_{jki}^{A1} \sqrt{f_{jki}^{A1}(t, \mathbf{q}(t))} dW_{jki}^{A1}(t) \\ &\quad + \sum_{i=1}^{m_S} \sum_{l \neq i}^{m_S} \mathbf{l}_{il}^{S0} \sqrt{f_{il}^{S0}(t, \mathbf{q}(t))} dW_{il}^{S0}(t) + \sum_{i=1}^{m_S} \mathbf{l}_i^{S1} \sqrt{f_i^{S1}(t, \mathbf{q}(t))} dW_i^{S1}(t). \end{aligned}$$

where $W_{jk}^{A0}(t), W_{jki}^{A1}(t), W_{il}^{S0}(t), W_i^{S1}(t)$ are mutually independent standard Brownian motions.

Proof. See Appendix. □

3.2.3 Performance Measures

Proposition 3.3. Let $\mathbf{M}(t) = E[\mathbf{D}(t)]$ and $\Sigma(t) = \text{Cov}[\mathbf{D}(t), \mathbf{D}(t)]$. Then, $\mathbf{M}(t)$ and $\Sigma(t)$ are the unique solution to the following ordinary equations:

$$\frac{d}{dt} \mathbf{M}(t) = \partial\mathbf{F}(t, \mathbf{q}(t))\mathbf{M}(t), \quad (3.6)$$

$$\frac{d}{dt} \Sigma(t) = \partial\mathbf{F}(t, \mathbf{q}(t))\Sigma(t) + \Sigma(t)\partial\mathbf{F}(t, \mathbf{q}(t))' + \mathbf{G}(t, \mathbf{q}(t)). \quad (3.7)$$

where

$$\mathbf{G}(t, \mathbf{q}(t)) = \sum_{j=1}^{m_A} \sum_{k \neq j}^{m_A} \mathbf{l}_{jk}^{A0} \mathbf{l}_{jk}^{A0'} f_{jk}^{A0}(t, \mathbf{q}(t)) + \sum_{j=1}^{m_A} \sum_{k=1}^{m_A} \sum_{i=1}^{m_S} \mathbf{l}_{jki}^{A1} \mathbf{l}_{jki}^{A1'} f_{jki}^{A1}(t, \mathbf{q}(t)) \quad (3.8)$$

$$+ \sum_{i=1}^{m_S} \sum_{l \neq i}^{m_S} \mathbf{l}_{il}^{S0} \mathbf{l}_{il}^{S0'} f_{il}^{S0}(t, \mathbf{q}(t)) + \sum_{i=1}^{m_S} \mathbf{l}_i^{S1} \mathbf{l}_i^{S1'} f_i^{S1}(t, \mathbf{q}(t)). \quad (3.9)$$

If $\mathbf{M}(0) = \mathbf{0}$, $\mathbf{M}(t) = \mathbf{0}$ for all $t \geq 0$.

Proof. See Arnold [2] Theorem 8.2.6 on page 131. □

Recall that we start with an empty queue, which implies that we do not have to solve Equation (3.6), i.e., $\mathbf{M}(t) = \mathbf{0}$ for all $t \geq 0$. By solving differential equations (3.4) and (3.7), we can approximate $E[\mathbf{Q}(t)]$ and $\text{Cov}[\mathbf{Q}(t), \mathbf{Q}(t)]$ as follows:

$$\begin{aligned} E[\mathbf{Q}(t)] &\approx \mathbf{q}(t), \\ \text{Cov}[\mathbf{Q}(t), \mathbf{Q}(t)] &\approx \mathbf{\Sigma}(t). \end{aligned}$$

Let $Z^*(t)$ be the number of customers in the queueing system at time t . Then,

$$Z^*(t) = \sum_{i=1}^{m_S} X_i(t).$$

Note that $\{Z^*(t), t \geq 0\}$ is a Gaussian process and therefore, we can obtain the mean and variance of the number of customers in the queue at time t , $Z^*(t)$, as follows:

$$\begin{aligned} m^*(t) &\equiv E[Z^*(t)] = \sum_{i=1}^{m_S} E[X_i(t)], \\ \sigma^*(t) &\equiv \sqrt{\text{Var}[Z^*(t)]} = \sqrt{\sum_{i=1}^{m_S} \text{Var}[X_i(t)] + \sum_{i=1}^{m_S} \sum_{l \neq i}^{m_S} \text{Cov}[X_i(t), X_l(t)]}. \end{aligned}$$

4 Conclusion and Final Remarks

In this paper, we analyze the $MAP_t/MSP_t/\infty$ queues and prove fluid and diffusion limits via strong approximation techniques. It is our hope that our analysis of the infinite server queue will guide us in the analysis for the $MAP_t/MSP_t/n_t$ and $MAP_t/MSP_t/n_t + MAP_t$ queues. Note the second MAP term in the $MAP_t/MSP_t/n_t + MAP_t$ queue can be regarded as the Markovian Abandonment Process (MAP). By extending our analysis to these types queues will allow us to model queueing systems with non-renewal arrival, service, and abandonment random variables, which will advance the state of the art with queueing models with dependence and correlation structures. The finite server setting is especially interesting because it is well known in Ko and Gautam [17], Massey and Pender [21, 22], Pender and Ko [34], Pender and Massey [35] that the fluid and diffusion limits need refining when the number of servers is not large. Moreover, in the finite server setting, we also know that lingering can have a substantial impact on the accuracy of the fluid and diffusion approximations. Thus, methods by Engblom and Pender [6], Pender [33, 29, 30, 31, 32] could be very relevant to improving the fluid and diffusion limits in the finite server setting.

Another extension that is even more interesting is the control of these networks with general distributions. Methods such as the fluid control of Hampshire et al. [12], Niyirora and Pender [24] seem promising as ways to optimally control the stochastic network on the fluid and diffusion scale respectively. We plan to pursue these extensions in later work.

Acknowledgment

This research was supported in part by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT & Future Planning (NRF-2014R1A1A1005019) and in part by Basic Science Research Program through the NRF funded by the Ministry of Education (NRF-2016R1D1A1B04933453).

A Appendix

In this section, we now provide the proof for the fluid and diffusion limit theorems of the queue length process for the $MAP_t/MSP_t/\infty$ queue. Before we begin the proof, we present two lemmas that are vital to understanding and constructing the proof via strong approximation theory.

Lemma A.1 (Kurtz 1978). *A standard Poisson process $\{\mathbf{\Pi}(t)\}_{t \geq 0}$ can be realized on the same probability space as a standard Brownian motion $\{W(t)\}_{t \geq 0}$ in such a way that the almost surely finite random variable*

$$Z \equiv \sup_{t \geq 0} \frac{|\mathbf{\Pi}(t) - t - W(t)|}{\log(2 \vee t)}$$

has finite moment generating function in the neighborhood of the origin and in particular finite mean.

Lemma A.2 (Kurtz 1978). *For any standard Brownian motion $\{W(t)\}_{t \geq 0}$ and any $\epsilon > 0$, $n \in \mathbb{N}$, and $T > 0$*

$$\tilde{M} \equiv \sup_{u, v, \leq n\epsilon T} \frac{|W(u) - W(v)|}{\sqrt{|u - v| (1 + \log(n\epsilon T / |u - v|))}} < \infty \quad a.s.$$

A.1 Proof of Fluid Limit

In this section, we will provide the proof of the fluid limit. From equations (3.2) and (3.3), we know the scaled queue length satisfies the following equation

$$\begin{aligned} \mathbf{Q}^\eta(t) &= \mathbf{Q}^\eta(0) + \sum_{j=1}^{m_A} \sum_{k \neq j}^{m_A} \mathbf{1}_{jk}^{A0} \mathbf{\Pi}_{jk}^{A0} \left(\eta \int_0^t f_{jk}^{A0}(s, \bar{\mathbf{Q}}^\eta(s)) ds \right) \\ &+ \sum_{j=1}^{m_A} \sum_{k=1}^{m_A} \sum_{i=1}^{m_S} \mathbf{1}_{jki}^{A1} \mathbf{\Pi}_{jki}^{A1} \left(\eta \int_0^t f_{jki}^{A1}(s, \bar{\mathbf{Q}}^\eta(s)) ds \right) + \sum_{i=1}^{m_S} \sum_{l \neq i}^{m_S} \mathbf{1}_{il}^{S0} \mathbf{\Pi}_{il}^{S0} \left(\eta \int_0^t f_{il}^{S0}(s, \bar{\mathbf{Q}}^\eta(s)) ds \right) \\ &+ \sum_{i=1}^{m_S} \mathbf{1}_i^{S1} \mathbf{\Pi}_i^{S1} \left(\eta \int_0^t f_i^{S1}(s, \bar{\mathbf{Q}}^\eta(s)) ds \right). \end{aligned}$$

Thus, by adding and subtracting the integrand of each Poisson process, we now have the following bound of the scaled queue length and the fluid limit,

$$\begin{aligned}
& \left| \frac{1}{\eta} \mathbf{Q}^\eta(t) - \mathbf{q}(t) \right| \\
& \leq \left| \frac{1}{\eta} \mathbf{Q}^\eta(0) - \mathbf{q}(0) \right| + \left| \sum_{j=1}^{m_A} \sum_{k \neq j}^{m_A} \mathbf{I}_{jk}^{A0} \int_0^t (f_{jk}^{A0}(s, \bar{\mathbf{Q}}^\eta(s)) - f_{jk}^{A0}(s, \mathbf{q}(s))) ds \right. \\
& + \sum_{j=1}^{m_A} \sum_{k=1}^{m_A} \sum_{i=1}^{m_S} \mathbf{I}_{jki}^{A1} \int_0^t (f_{jki}^{A1}(s, \bar{\mathbf{Q}}^\eta(s)) - f_{jki}^{A1}(s, \mathbf{q}(s))) ds \\
& + \sum_{i=1}^{m_S} \sum_{l \neq i}^{m_S} \mathbf{I}_{il}^{S0} \int_0^t (f_{il}^{S0}(s, \bar{\mathbf{Q}}^\eta(s)) - f_{il}^{S0}(s, \mathbf{q}(s))) ds \\
& + \sum_{i=1}^{m_S} \mathbf{I}_i^{S1} \int_0^t (f_i^{S1}(s, \bar{\mathbf{Q}}^\eta(s)) - f_i^{S1}(s, \mathbf{q}(s))) ds \left. \right| \\
& + \sum_{j=1}^{m_A} \sum_{k \neq j}^{m_A} \left| \mathbf{I}_{jk}^{A0} \frac{1}{\eta} \left(\bar{\mathbf{\Pi}}_{jk}^{A0} \left(\eta \int_0^t f_{jk}^{A0}(s, \bar{\mathbf{Q}}^\eta(s)) ds \right) \right) \right| \\
& + \sum_{j=1}^{m_A} \sum_{k=1}^{m_A} \sum_{i=1}^{m_S} \left| \mathbf{I}_{jki}^{A1} \frac{1}{\eta} \bar{\mathbf{\Pi}}_{jki}^{A1} \left(\eta \int_0^t f_{jki}^{A1}(s, \bar{\mathbf{Q}}^\eta(s)) ds \right) \right| \\
& + \sum_{i=1}^{m_S} \sum_{l \neq i}^{m_S} \left| \mathbf{I}_{il}^{S0} \frac{1}{\eta} \bar{\mathbf{\Pi}}_{il}^{S0} \left(\eta \int_0^t f_{il}^{S0}(s, \bar{\mathbf{Q}}^\eta(s)) ds \right) \right| \\
& + \sum_{i=1}^{m_S} \left| \mathbf{I}_i^{S1} \frac{1}{\eta} \bar{\mathbf{\Pi}}_i^{S1} \left(\eta \int_0^t f_i^{S1}(s, \bar{\mathbf{Q}}^\eta(s)) ds \right) \right|
\end{aligned}$$

where we define $\bar{\mathbf{\Pi}}(\cdot)$ as

$$\bar{\mathbf{\Pi}} \left(\eta \int_0^t f(s, \bar{\mathbf{Q}}^\eta(s)) ds \right) = \mathbf{\Pi} \left(\eta \int_0^t f(s, \bar{\mathbf{Q}}^\eta(s)) ds \right) - \eta \int_0^t f(s, \bar{\mathbf{Q}}^\eta(s)) ds.$$

Now we use the Lipschitz continuity of the rate functions in the integrand of each Poisson

process to show that

$$\begin{aligned}
& \left| \frac{1}{\eta} \mathbf{Q}^\eta(t) - \mathbf{q}(t) \right| \\
& \leq \left| \frac{1}{\eta} \mathbf{Q}^\eta(0) - \mathbf{q}(0) \right| + C \int_0^t \left| \frac{1}{\eta} \mathbf{Q}^\eta(s) - \mathbf{q}(s) \right| ds \\
& + \sum_{j=1}^{m_A} \sum_{k \neq j}^{m_A} \left| \mathbf{I}_{jk}^{A0} \frac{1}{\eta} \overline{\Pi}_{jk}^{A0} \left(\eta \int_0^t f_{jk}^{A0}(s, \bar{\mathbf{Q}}^\eta(s)) ds \right) \right| \\
& + \sum_{j=1}^{m_A} \sum_{k=1}^{m_A} \sum_{i=1}^{m_S} \left| \mathbf{I}_{jki}^{A1} \frac{1}{\eta} \overline{\Pi}_{jki}^{A1} \left(\eta \int_0^t f_{jki}^{A1}(s, \bar{\mathbf{Q}}^\eta(s)) ds \right) \right| \\
& + \sum_{i=1}^{m_S} \sum_{l \neq i}^{m_S} \left| \mathbf{I}_{il}^{S0} \frac{1}{\eta} \overline{\Pi}_{il}^{S0} \left(\eta \int_0^t f_{il}^{S0}(s, \bar{\mathbf{Q}}^\eta(s)) ds \right) \right| + \sum_{i=1}^{m_S} \left| \mathbf{I}_i^{S1} \frac{1}{\eta} \overline{\Pi}_i^{S1} \left(\eta \int_0^t f_i^{S1}(s, \bar{\mathbf{Q}}^\eta(s)) ds \right) \right|
\end{aligned}$$

In view of the strong approximation results given in Lemma A.1,

$$\begin{aligned}
& \sum_{j=1}^{m_A} \sum_{k \neq j}^{m_A} \left| \mathbf{I}_{jk}^{A0} \left[\overline{\Pi}_{jk}^{A0} \left(\eta \int_0^t f_{jk}^{A0}(s, \bar{\mathbf{Q}}^\eta(s)) ds \right) - W_{jk}^{A0} \left(\eta \int_0^t f_{jk}^{A0}(s, \bar{\mathbf{Q}}^\eta(s)) ds \right) \right] \right| \\
& + \sum_{j=1}^{m_A} \sum_{k=1}^{m_A} \sum_{i=1}^{m_S} \left| \mathbf{I}_{jki}^{A1} \left[\overline{\Pi}_{jki}^{A1} \left(\eta \int_0^t f_{jki}^{A1}(s, \bar{\mathbf{Q}}^\eta(s)) ds \right) - W_{jki}^{A1} \left(\eta \int_0^t f_{jki}^{A1}(s, \bar{\mathbf{Q}}^\eta(s)) ds \right) \right] \right| \\
& + \sum_{i=1}^{m_S} \sum_{l \neq i}^{m_S} \left| \mathbf{I}_{il}^{S0} \left[\overline{\Pi}_{il}^{S0} \left(\eta \int_0^t f_{il}^{S0}(s, \bar{\mathbf{Q}}^\eta(s)) ds \right) - W_{il}^{S0} \left(\eta \int_0^t f_{il}^{S0}(s, \bar{\mathbf{Q}}^\eta(s)) ds \right) \right] \right| \\
& + \sum_{i=1}^{m_S} \left| \mathbf{I}_i^{S1} \left[\overline{\Pi}_i^{S1} \left(\eta \int_0^t f_i^{S1}(s, \bar{\mathbf{Q}}^\eta(s)) ds \right) - W_i^{S1} \left(\eta \int_0^t f_i^{S1}(s, \bar{\mathbf{Q}}^\eta(s)) ds \right) \right] \right|
\end{aligned}$$

is of the order $\Theta(\log(\eta))$ almost surely. Since the $W(\cdot)$ terms are standard Brownian motions and the rate functions or integrands of the Brownian motion processes are Lipschitz continuous, the law of the iterated logarithm for Brownian motion asserts that

$$\limsup_{\eta \rightarrow \infty} \sup_{t \leq T} \frac{1}{\eta} W_{jk}^{A0} \left(\eta \int_0^t f_{jk}^{A0}(s, \bar{\mathbf{Q}}^\eta(s)) ds \right) = 0 \quad \text{almost surely,}$$

$$\limsup_{\eta \rightarrow \infty} \sup_{t \leq T} \frac{1}{\eta} W_{jki}^{A1} \left(\eta \int_0^t f_{jki}^{A1}(s, \bar{\mathbf{Q}}^\eta(s)) ds \right) = 0 \quad \text{almost surely,}$$

$$\limsup_{\eta \rightarrow \infty} \sup_{t \leq T} \frac{1}{\eta} W_{il}^{S0} \left(\eta \int_0^t f_{il}^{S0}(s, \bar{\mathbf{Q}}^\eta(s)) ds \right) = 0 \quad \text{almost surely,}$$

$$\limsup_{\eta \rightarrow \infty} \sup_{t \leq T} \frac{1}{\eta} W_i^{S1} \left(\eta \int_0^t f_i^{S1}(s, \bar{\mathbf{Q}}^\eta(s)) ds \right) = 0 \quad \text{almost surely.}$$

This implies that

$$\begin{aligned}
& \sum_{j=1}^{m_A} \sum_{k \neq j}^{m_A} \left| \mathbf{I}_{jk}^{A0} \frac{1}{\eta} W_{jk}^{A0} \left(\eta \int_0^t f_{jk}^{A0}(s, \bar{\mathbf{Q}}^\eta(s)) ds \right) \right| \\
& + \sum_{j=1}^{m_A} \sum_{k=1}^{m_A} \sum_{i=1}^{m_S} \left| \mathbf{I}_{jki}^{A1} \frac{1}{\eta} W_{jki}^{A1} \left(\eta \int_0^t f_{jki}^{A1}(s, \bar{\mathbf{Q}}^\eta(s)) ds \right) \right| \\
& + \sum_{i=1}^{m_S} \sum_{l \neq i}^{m_S} \left| \mathbf{I}_{il}^{S0} \frac{1}{\eta} W_{il}^{S0} \left(\eta \int_0^t f_{il}^{S0}(s, \bar{\mathbf{Q}}^\eta(s)) ds \right) \right| + \sum_{i=1}^{m_S} \left| \mathbf{I}_i^{S1} \frac{1}{\eta} W_i^{S1} \left(\eta \int_0^t f_i^{S1}(s, \bar{\mathbf{Q}}^\eta(s)) ds \right) \right|
\end{aligned}$$

converges to zero uniformly over compact sets of time as η goes to ∞ . As a result, for sufficiently large $\eta^* \in \mathbb{N}$ and $\varepsilon > 0$, we have that for all $\eta \geq \eta^*$, that

$$\begin{aligned}
\varepsilon/2 & \geq \sum_{j=1}^{m_A} \sum_{k \neq j}^{m_A} \left| \mathbf{I}_{jk}^{A0} \frac{1}{\eta} W_{jk}^{A0} \left(\eta \int_0^t f_{jk}^{A0}(s, \bar{\mathbf{Q}}^\eta(s)) ds \right) \right| \\
& + \sum_{j=1}^{m_A} \sum_{k=1}^{m_A} \sum_{i=1}^{m_S} \left| \mathbf{I}_{jki}^{A1} \frac{1}{\eta} W_{jki}^{A1} \left(\eta \int_0^t f_{jki}^{A1}(s, \bar{\mathbf{Q}}^\eta(s)) ds \right) \right| \\
& + \sum_{i=1}^{m_S} \sum_{l \neq i}^{m_S} \left| \mathbf{I}_{il}^{S0} \frac{1}{\eta} W_{il}^{S0} \left(\eta \int_0^t f_{il}^{S0}(s, \bar{\mathbf{Q}}^\eta(s)) ds \right) \right| \\
& + \sum_{i=1}^{m_S} \mathbf{I}_i^{S1} \frac{1}{\eta} \left| \bar{\mathbf{\Pi}}_i^{S1} \left(\eta \int_0^t f_i^{S1}(s, \bar{\mathbf{Q}}^\eta(s)) ds \right) \right| + \frac{\Theta(\log(\eta))}{\eta}.
\end{aligned}$$

Thus, for some large enough constant \tilde{C} and sufficiently large enough η^* , we have that

$$\left| \frac{1}{\eta} \mathbf{Q}^\eta(t) - \mathbf{q}(t) \right| \leq \left| \frac{1}{\eta} \mathbf{Q}^\eta(0) - \mathbf{q}(0) \right| + C \int_0^t \left| \frac{1}{\eta} \mathbf{Q}^\eta(s) - \mathbf{q}(s) \right| ds + \varepsilon/2$$

Now by assuming that η is large enough that

$$\varepsilon/2 \geq \left| \frac{1}{\eta} \mathbf{Q}^\eta(0) - \mathbf{q}(0) \right|,$$

then we have that

$$\left| \frac{1}{\eta} \mathbf{Q}^\eta(t) - \mathbf{q}(t) \right| \leq \varepsilon + C \int_0^t \left| \frac{1}{\eta} \mathbf{Q}^\eta(s) - \mathbf{q}(s) \right| ds.$$

Finally, our fluid limit result follows by Gronwall's lemma given in Problem 2.7 of Karatzas and Shreve [16].

A.2 Proof of Diffusion Limit

In order to construct the diffusion limit, we need to subtract the fluid limit and multiply by $\sqrt{\eta}$. This yields the following expression for $\mathbf{D}^\eta(t)$

$$\begin{aligned}\mathbf{D}^\eta(t) &= \sqrt{\eta} \left(\frac{1}{\eta} \mathbf{Q}^\eta(t) - \mathbf{q}(t) \right) \\ &= \sqrt{\eta} \int_0^t (\mathbf{F}(s, \bar{\mathbf{Q}}^\eta(s)) - \mathbf{F}(s, \mathbf{q}(s))) ds + \mathbf{V}^\eta(t).\end{aligned}$$

where

$$\begin{aligned}\sqrt{\eta} \mathbf{V}^\eta(t) &= \sum_{j=1}^{m_A} \sum_{k \neq j}^{m_A} \mathbf{I}_{jk}^{A0} \bar{\mathbf{\Pi}}_{jk}^{A0} \left(\int_0^t \eta f_{jk}^{A0}(s, \bar{\mathbf{Q}}^\eta(s)) ds \right) \\ &+ \sum_{j=1}^{m_A} \sum_{k=1}^{m_A} \sum_{i=1}^{m_S} \mathbf{I}_{jki}^{A1} \bar{\mathbf{\Pi}}_{jki}^{A1} \left(\int_0^t \eta f_{jki}^{A1}(s, \bar{\mathbf{Q}}^\eta(s)) ds \right) \\ &+ \sum_{i=1}^{m_S} \sum_{l \neq i}^{m_S} \mathbf{I}_{il}^{S0} \bar{\mathbf{\Pi}}_{il}^{S0} \left(\int_0^t \eta f_{il}^{S0}(s, \bar{\mathbf{Q}}^\eta(s)) ds \right) + \sum_{i=1}^{m_S} \mathbf{I}_i^{S1} \bar{\mathbf{\Pi}}_i^{S1} \left(\int_0^t \eta f_i^{S1}(s, \bar{\mathbf{Q}}^\eta(s)) ds \right).\end{aligned}$$

Now we need two propositions that will be helpful in proving our main result.

Proposition A.3. *Let $\mathbf{M}^\eta(t)$ be defined by the following equation*

$$\begin{aligned}\mathbf{M}^\eta(t) &= \sum_{j=1}^{m_A} \sum_{k \neq j}^{m_A} \mathbf{I}_{jk}^{A0} W_{jk}^{A0} \left(\int_0^t f_{jk}^{A0}(s, \bar{\mathbf{Q}}^\eta(s)) ds \right) \\ &+ \sum_{j=1}^{m_A} \sum_{k=1}^{m_A} \sum_{i=1}^{m_S} \mathbf{I}_{jki}^{A1} W_{jki}^{A1} \left(\int_0^t f_{jki}^{A1}(s, \bar{\mathbf{Q}}^\eta(s)) ds \right) \\ &+ \sum_{i=1}^{m_S} \sum_{l \neq i}^{m_S} \mathbf{I}_{il}^{S0} W_{il}^{S0} \left(\int_0^t f_{il}^{S0}(s, \bar{\mathbf{Q}}^\eta(s)) ds \right) + \sum_{i=1}^{m_S} \mathbf{I}_i^{S1} W_i^{S1} \left(\int_0^t f_i^{S1}(s, \bar{\mathbf{Q}}^\eta(s)) ds \right),\end{aligned}$$

then

$$\lim_{\eta \rightarrow \infty} \sup_{0 \leq t \leq T} |\mathbf{M}^\eta(t) - \mathbf{V}^\eta(t)| = 0 \quad \text{in distribution.} \quad (\text{A.1})$$

We will show the result for one of the Brownian motion terms and one of the centered Poisson processes. The proof for the remaining terms will follow in a similar manner and are therefore omitted.

Proof. Using the strong approximation result of Lemma A.1, we obtain

$$\sup_{t \geq 0} \frac{1}{\sqrt{\eta}} \frac{\left| \bar{\mathbf{\Pi}}_{jk}^{A0} \left(\eta \int_0^t f_{jk}^{A0}(s, \bar{\mathbf{Q}}^\eta(s)) ds \right) - W_{jk}^{A0} \left(\eta \int_0^t f_{jk}^{A0}(s, \bar{\mathbf{Q}}^\eta(s)) ds \right) \right|}{\log \left(2 \vee \eta \int_0^t f_{jk}^{A0}(s, \bar{\mathbf{Q}}^\eta(s)) ds \right)} \leq \frac{B_{jk}^{A0}}{\sqrt{\eta}} \quad (\text{A.2})$$

where the distribution of B_{jk}^A is independent η . Using the above strong approximation result, which the assumption that the rate functions are locally bounded by a constant K , then we have that

$$\sup_{0 \leq t \leq T} |\mathbf{M}^\eta(t) - \mathbf{V}^\eta(t)| \leq \log(2 \vee \eta KT) \sup_{0 \leq t \leq T} \frac{|\mathbf{M}^\eta(t) - \mathbf{V}^\eta(t)|}{\log(2 \vee \eta Kt)} \quad (\text{A.3})$$

$$\leq \log(2 \vee \eta KT) \frac{B_{jk}^{A0}}{\sqrt{\eta}} \quad (\text{A.4})$$

Since the distribution of B_{jk}^{A0} is independent η and we have that

$$\lim_{\eta \rightarrow \infty} \frac{\log(2 \vee \eta KT)}{\sqrt{\eta}} = 0,$$

it implies that as $\eta \rightarrow \infty$ we have that

$$\sup_{0 \leq t \leq T} |\mathbf{M}^\eta(t) - \mathbf{V}^\eta(t)| \Rightarrow 0 \quad \text{in distribution as } \eta \rightarrow \infty. \quad (\text{A.5})$$

All the other terms can be proved similarly with the same technique. \square

Now that we have related the centered Poisson processes with time changed Brownian motions, it remains for us to show that the fluid scaled randomly time changed Brownian motion terms converge to Brownian motions time changed with the deterministic fluid equations. The following Proposition A.4 proves this result.

Proposition A.4. *The sequence of stochastic processes $\mathbf{M}^\eta(t)$ converges in distribution to the process $\mathbf{M}(t)$ where*

$$\begin{aligned} \mathbf{M}(t) &= \sum_{j=1}^{m_A} \sum_{k \neq j}^{m_A} \mathbf{I}_{jk}^{A0} W_{jk}^{A0} \left(\int_0^t f_{jk}^{A0}(s, \mathbf{q}(s)) ds \right) \\ &+ \sum_{j=1}^{m_A} \sum_{k=1}^{m_A} \sum_{i=1}^{m_S} \mathbf{I}_{jki}^{A1} W_{jki}^{A1} \left(\int_0^t f_{jki}^{A1}(s, \mathbf{q}(s)) ds \right) \\ &+ \sum_{i=1}^{m_S} \sum_{l \neq i}^{m_S} \mathbf{I}_{il}^{S0} W_{il}^{S0} \left(\int_0^t f_{il}^{S0}(s, \mathbf{q}(s)) ds \right) + \sum_{i=1}^{m_S} \mathbf{I}_i^{S1} W_i^{S1} \left(\int_0^t f_i^{S1}(s, \mathbf{q}(s)) ds \right) \end{aligned} \quad (\text{A.6})$$

In order to prove the convergence of the scaled Brownian motions, we will use Lemma A.2. Moreover, we will provide the full proof for one term and the proofs for the remaining terms follow analogously. We now define a new function $\gamma^\eta(t)$ as follows.

Proof.

$$\gamma^\eta(t) \equiv \left| \int_0^t f_{jk}^{A0}(s, \bar{\mathbf{Q}}^\eta(s)) ds - \int_0^t f_{jk}^{A0}(s, \mathbf{q}(s)) ds \right|$$

and

$$\bar{\gamma}^\eta \equiv \sup_{0 \leq t \leq T} \gamma^\eta(t).$$

This implies that

$$\begin{aligned} & \left| W_{jk}^{A0} \left(\int_0^t f_{jk}^{A0}(s, \bar{\mathbf{Q}}^\eta(s)) ds \right) - W_{jk}^{A0} \left(\int_0^t f_{jk}^{A0}(s, \mathbf{q}(s)) ds \right) \right| \\ &= \frac{\left| W_{jk}^{A0} \left(\int_0^t f_{jk}^{A0}(s, \bar{\mathbf{Q}}^\eta(s)) ds \right) - W_{jk}^{A0} \left(\int_0^t f_{jk}^{A0}(s, \mathbf{q}(s)) ds \right) \right|}{\sqrt{\gamma^\eta(t) \cdot (1 + \log(KT/\gamma^\eta(t)))}} \cdot \sqrt{\gamma^\eta(t) \cdot (1 + \log(KT/\gamma^\eta(t)))}. \end{aligned}$$

However, from Lemma A.2, we obtain

$$\begin{aligned} & \left| W_{jk}^{A0} \left(\int_0^t f_{jk}^{A0}(s, \bar{\mathbf{Q}}^\eta(s)) ds \right) - W_{jk}^{A0} \left(\int_0^t f_{jk}^{A0}(s, \mathbf{q}(s)) ds \right) \right| \\ & \leq \tilde{M} \cdot \sqrt{\gamma^\eta(t) \cdot (1 + \log(KT/\gamma^\eta(t)))}. \end{aligned}$$

From the Lipschitz continuity of the rate functions, we have that

$$\bar{\gamma}^\eta \leq KT \cdot \sup_{0 \leq t \leq T} \left| \frac{1}{\eta} \mathbf{Q}^\eta(t) - \mathbf{q}(t) \right|.$$

Therefore, by convergence of the fluid limit, we have that

$$\bar{\gamma}^\eta \Rightarrow 0.$$

By observing that the distribution of \tilde{M} is independent of η and that the following limit

$$\lim_{\delta \rightarrow 0} \sqrt{\delta \cdot (1 + \log(KT/\delta))} = 0,$$

we conclude that

$$\tilde{M} \cdot \sqrt{\bar{\gamma}^\eta \cdot (1 + \log(KT/\bar{\gamma}^\eta))} \Rightarrow 0$$

and therefore,

$$\lim_{\eta \rightarrow \infty} \sup_{0 \leq t \leq T} \left| W_{jk}^{A0} \left(\int_0^t f_{jk}^{A0}(s, \bar{\mathbf{Q}}^\eta(s)) ds \right) - W_{jk}^{A0} \left(\int_0^t f_{jk}^{A0}(s, \mathbf{q}(s)) ds \right) \right| \Rightarrow 0.$$

The remaining terms can be shown to converge by identical arguments and therefore, we do not provide their proofs. \square

The following lemma shows that the sequence $\mathbf{D}^\eta(t)$ is stochastically bounded.

Lemma A.5. *For any $\epsilon > 0$, there exists $\eta^* \in \mathbb{N}$ and $K < \infty$ such that*

$$\mathbb{P} \left(\sup_{0 \leq t \leq T} |\mathbf{D}^\eta(t)| > K \right) < \epsilon \quad \text{for all } \eta \geq \eta^*. \quad (\text{A.7})$$

Proof. The strong approximation for the Brownian motion yields the following representation

$$\mathbf{D}^\eta(t) = \sqrt{\eta} \int_0^t (\mathbf{F}(s, \bar{\mathbf{Q}}^\eta(s)) - \mathbf{F}(s, \mathbf{q}(s))) ds + \mathbf{V}^\eta(t).$$

We know that $\mathbf{V}^\eta(t)$ is tight since it converges to a time changed Brownian motion, which is a continuous stochastic processes. Therefore, the tightness of $\mathbf{V}^\eta(t)$ implies that it is bounded in probability. Moreover, by using the Lipschitz continuity of the rate functions we have that

$$\sup_{0 \leq t \leq T} |\mathbf{D}^\eta(t)| \leq L \int_0^T \sup_{0 \leq t \leq s} |\mathbf{D}^\eta(s)| ds + \sup_{0 \leq t \leq T} |\mathbf{V}^\eta(t)|$$

for some Lipschitz constant L . Thus, by Gronwall's inequality in Problem 2.7 of Karatzas and Shreve [16] we have almost surely that

$$\sup_{0 \leq t \leq T} \mathbf{D}^\eta(t) \leq e^{LT} \sup_{0 \leq t \leq T} \mathbf{V}^\eta(t)$$

and this concludes the proof. \square

Lemma A.6. *If $\{f^\eta(t), \eta \in \mathcal{N}, t \in \mathcal{R}_+\}$ be a sequence of non-negative random processes such that*

$$\lim_{\eta \rightarrow \infty} \int_0^T f^\eta(s) ds = 0 \quad \text{in probability,} \quad (\text{A.8})$$

then, for all $\delta > 0$,

$$\lim_{\eta \rightarrow \infty} \mathbb{P} \left(\sup_{0 \leq t \leq T} \left| \int_0^t f^\eta(s) \mathbf{D}^\eta(s) ds \right| > \delta \right) = 0. \quad (\text{A.9})$$

Proof. If we fix $\epsilon > 0$, then we know that there exists a constant $\eta^* \in \mathbb{N}$ such that for all $\eta > \eta^*$, there exists sets $\Omega_{\eta,1}$ and $\Omega_{\eta,2}$ such that

$$\int_0^T f^\eta(s) ds < \epsilon/2 \quad \text{on } \Omega_{\eta,1} \text{ and such that } \mathbb{P}(\Omega_{\eta,1}) \geq 1 - \epsilon/2, \quad (\text{A.10})$$

and

$$\sup_{0 \leq t \leq T} |\mathbf{D}^\eta(t)| < K \quad \text{on } \Omega_{\eta,2} \text{ and such that } \mathbb{P}(\Omega_{\eta,2}) \geq 1 - \epsilon/2, \quad (\text{A.11})$$

Therefore, we have that

$$\sup_{0 \leq t \leq T} \left| \int_0^t f^\eta(s) \mathbf{D}^\eta(s) ds \right| \leq \sup_{0 \leq t \leq T} |\mathbf{D}^\eta(t)| \int_0^T f^\eta(s) ds < K\epsilon \quad \text{on } \Omega_{\eta,1} \cap \Omega_{\eta,2}. \quad (\text{A.12})$$

The result follows since ϵ was chosen arbitrarily. \square

Theorem A.7. *Suppose that we define $\tilde{\mathbf{D}}^\eta(t)$ as*

$$\tilde{\mathbf{D}}^\eta(t) \equiv \int_0^t \partial \mathbf{F}(s, \mathbf{q}(s)) \tilde{\mathbf{D}}^\eta(s) ds + \mathbf{V}^\eta(t), \quad (\text{A.13})$$

then

$$\lim_{\eta \rightarrow \infty} \sup_{0 \leq t \leq T} |\mathbf{D}^\eta(t) - \tilde{\mathbf{D}}^\eta(t)| = 0 \quad \text{in probability.} \quad (\text{A.14})$$

We know by the continuous mapping theorem and Proposition A.4, which shows that $\mathbf{V}^\eta(t)$ converges to $\mathbf{M}(t)$ in Equation A.6, then we know that that $\tilde{\mathbf{D}}^\eta(t)$ converges to $\tilde{\mathbf{D}}(t)$ given

in Equation 3.5. Thus, in order to show our diffusion limit results in Theorem 3.2, it now suffices to show the following convergence

$$\lim_{\eta \rightarrow \infty} \sup_{0 \leq t \leq T} |\mathbf{D}^\eta(t) - \tilde{\mathbf{D}}^\eta(t)| = 0 \quad \text{in probability.} \quad (\text{A.15})$$

To prove this, we define the difference between the two processes as

$$\begin{aligned} \mathbf{E}^\eta(t) &= \mathbf{D}^\eta(t) - \tilde{\mathbf{D}}^\eta(t) \\ &= \sqrt{\eta} \int_0^t (\mathbf{F}(s, \bar{\mathbf{Q}}^\eta(s)) - \mathbf{F}(s, \mathbf{q}(s))) ds + \mathbf{V}^\eta(t) - \left(\int_0^t \partial \mathbf{F}(s, \mathbf{q}(s)) \tilde{\mathbf{D}}^\eta(s) ds + \mathbf{V}^\eta(t) \right) \\ &= \sqrt{\eta} \int_0^t (\mathbf{F}(s, \bar{\mathbf{Q}}^\eta(s)) - \mathbf{F}(s, \mathbf{q}(s))) ds - \int_0^t \partial \mathbf{F}(s, \mathbf{q}(s)) \tilde{\mathbf{D}}^\eta(s) ds \\ &= \int_0^t \partial \mathbf{F}(s, \mathbf{q}(s)) \mathbf{E}^\eta(s) ds + \sqrt{\eta} \int_0^t (\mathbf{F}(s, \bar{\mathbf{Q}}^\eta(s)) - \mathbf{F}(s, \mathbf{q}(s))) ds \\ &\quad - \int_0^t \partial \mathbf{F}(s, \mathbf{q}(s)) \mathbf{D}^\eta(s) ds \end{aligned}$$

Thus, by the mean value theorem and the fact that the rate functions in the Poisson representations are continuously differentiable, there exists a vector $\zeta^\eta(s)$ that is in between $\mathbf{q}(s)$ and $\bar{\mathbf{Q}}^\eta(s)$ such that

$$\begin{aligned} \mathbf{F}(s, \bar{\mathbf{Q}}^\eta(s)) - \mathbf{F}(s, \mathbf{q}(s)) &= \partial \mathbf{F}(s, \zeta^\eta(s)) \cdot (\bar{\mathbf{Q}}^\eta(s) - \mathbf{q}(s)) \\ &= \partial \mathbf{F}(s, \zeta^\eta(s)) \cdot \frac{1}{\sqrt{\eta}} \cdot \sqrt{\eta} (\bar{\mathbf{Q}}^\eta(s) - \mathbf{q}(s)) \\ &= \frac{1}{\sqrt{\eta}} \partial \mathbf{F}(s, \zeta^\eta(s)) \mathbf{D}^\eta(s). \end{aligned}$$

From this equivalence provided by the mean value theorem, it now implies that

$$\mathbf{E}^\eta(t) = \int_0^t (\partial \mathbf{F}(s, \zeta^\eta(s)) - \partial \mathbf{F}(s, \mathbf{q}(s))) \mathbf{D}^\eta(s) ds + \int_0^t \partial \mathbf{F}(s, \mathbf{q}(s)) \mathbf{E}^\eta(s) ds.$$

We also know that

$$\lim_{\eta \rightarrow \infty} \sup_{0 \leq t \leq T} \|\partial \mathbf{F}(t, \zeta^\eta(t)) - \partial \mathbf{F}(t, \mathbf{q}(t))\| = 0 \quad \text{a.s.} \quad (\text{A.16})$$

in lieu of the fluid limit convergence and the continuity of the function $\partial \mathbf{F}(\cdot, \zeta^\eta(\cdot))$. Moreover, since $\mathbf{D}^\eta(u)$ is bounded in probability and Lemma A.6 is true, we have that the process

$$\lim_{\eta \rightarrow \infty} \sup_{0 \leq t \leq T} \int_0^t (\partial \mathbf{F}(s, \zeta^\eta(s)) - \partial \mathbf{F}(s, \mathbf{q}(s))) \mathbf{D}^\eta(s) ds = 0 \quad \text{in probability.}$$

Finally by the application of Gronwall's inequality in Problem 2.7 of Karatzas and Shreve [16] and Lemma A.6, we obtain our diffusion limit result.

References

- [1] Allan T Andersen and Bo Friis Nielsen. A markovian approach for modeling packet traffic with long-range dependence. *Selected Areas in Communications, IEEE Journal on*, 16(5):719–732, 1998.
- [2] Ludwig Arnold. *Stochastic Differential Equations: Theory and Applications*. Krieger Publishing Company, 1992.
- [3] F. Cecchi, S. C. Borst, J. S. H. van Leeuwaarden, and P. A. Whiting. CSMA networks in a many-sources regime: A mean-field approach. In *Proceedings of IEEE INFOCOM 2016*, pages 1–9. IEEE, apr 2016.
- [4] Srinivas R Chakravathy. Markovian arrival processes. *Wiley Encyclopedia of Operations Research and Management Science*, 2010.
- [5] B. D’Auria. $M/M/\infty$ queues in semi-Markovian random environmen. *Queueing Systems*, 58(3):221–237, 2008.
- [6] Stefan Engblom and Jamol Pender. Approximations for the moments of nonstationary and state dependent birth-death queues. *arXiv preprint arXiv:1406.6164*, 2014.
- [7] Do Young Eun and Ness B. Shroff. Network Decomposition in the Many-Sources Regime. *Advances in Applied Probability*, 36(3):898–918, 2004.
- [8] Ira Gerhardt and Barry L Nelson. On capturing dependence in point processes: Matching moments and other techniques. Technical report, Technical report, Northwestern Univ, 2009.
- [9] Nasser Hadidi. Queues with partial correlation. *SIAM Journal on Applied Mathematics*, 40(3):467–475, 1981.
- [10] Nasser Hadidi. Further results on queues with partial correlation. *Operations research*, 33(1):203–209, 1985.
- [11] Robert C. Hampshire, Mor Harchol-Balter, and William A. Massey. Fluid and diffusion limits for transient sojourn times of processor sharing queues with time varying rates. *Queueing Systems*, 53(1-2):19–30, June 2006.
- [12] Robert C Hampshire, Otis B Jennings, and William A Massey. A time-varying call center design via lagrangian mechanics. *Probability in the Engineering and Informational Sciences*, 23(02):231–259, 2009.
- [13] Armin Heindl and Miklós Telek. Output models of $MAP/PH/1(/K)$ queues for an efficient network decomposition. *Performance Evaluation*, 49(1):321–339, 2002.
- [14] Gábor Horváth, Miklós Telek, and Benny Van Houdt. Commuting matrices in the sojourn time analysis of map/map/1 queues. 2014.

- [15] Sang H Kang, Yong Han Kim, Dan K Sung, and Bong D Choi. An application of markovian arrival process (map) to modeling superposed atm cell streams. *Communications, IEEE Transactions on*, 50(4):633–642, 2002.
- [16] Ioannis Karatzas and Steven Shreve. *Brownian motion and stochastic calculus*, volume 113. Springer Science & Business Media, 2012.
- [17] Young Myoung Ko and Natarajan Gautam. Critically Loaded Time-Varying Multi-server Queues: Computational Challenges and Approximations. *INFORMS Journal on Computing*, 25(2):285–301, May 2013.
- [18] Young Myoung Ko and Jamol Pender. Diffusion limits for the $(MAP_t/Ph_t/\infty)^N$ queueing network. *Operations Research Letters*, 45(3):248–253, 2017.
- [19] T Kurtz. Strong approximation theorems for density dependent Markov chains. *Stochastic Processes and their Applications*, 6(3):223–240, February 1978.
- [20] Avi Mandelbaum, William Massey, and Martin Reiman. Strong approximations for Markovian service networks. *Queueing Systems*, 30(1):149–201, 1998.
- [21] William A Massey and Jamol Pender. Poster: skewness variance approximation for dynamic rate multiserver queues with abandonment. *ACM SIGMETRICS Performance Evaluation Review*, 39(2):74–74, 2011.
- [22] William A. Massey and Jamol Pender. Gaussian skewness approximation for dynamic rate multi-server queues with abandonment. *Queueing Systems*, 75(2-4):243–277, February 2013.
- [23] András Mészáros and Miklós Telek. A two-phase map fitting method with aph interarrival time distribution. In *Proceedings of the Winter Simulation Conference*, page 425. Winter Simulation Conference, 2012.
- [24] Jerome Niyirora and Jamol Pender. Optimal staffing in nonstationary service centers with constraints. *Naval Research Logistics (NRL)*, 2016.
- [25] C. A. O’Cinneide and P. Purdue. The $M/M/\infty$ Queue in a Random Environment. *Journal of Applied Probability*, 23(1):175, mar 1986.
- [26] Guodong Pang and Ward Whitt. Heavy-traffic limits for many-server queues with service interruptions. *Queueing Systems*, 61(2-3):167–202, Mar 2009.
- [27] Guodong Pang and Ward Whitt. The impact of dependent service times on large-scale service systems. *Manufacturing & Service Operations Management*, 14(2):262–278, 2012.
- [28] Guodong Pang and Ward Whitt. Infinite-server queues with batch arrivals and dependent service times. *Probability in the Engineering and Informational Sciences*, 26(02):197–220, 2012.

- [29] Jamol Pender. Gram charlier expansion for time varying multiserver queues with abandonment. *SIAM Journal on Applied Mathematics*, 74(4):1238–1265, 2014.
- [30] Jamol Pender. Nonstationary loss queues via cumulant moment approximations. *Probability in the Engineering and Informational Sciences*, 29(01):27–49, 2015.
- [31] Jamol Pender. An analysis of nonstationary coupled queues. *Telecommunication Systems*, pages 1–16, 2015.
- [32] Jamol Pender. The truncated normal distribution: Applications to queues with impatient customers. *Operations Research Letters*, 43(1):40–45, 2015.
- [33] Jamol Pender. Sampling the functional kolmogorov forward equations for nonstationary queueing networks. *INFORMS Journal on Computing*, 29(1):1–17, 2016.
- [34] Jamol Pender and Young Myoung Ko. Approximations for the Queue Length Distributions of Time-Varying Many-Server Queues. *INFORMS Journal on Computing*, to appear, 2017.
- [35] Jamol Pender and William A Massey. Approximating and stabilizing dynamic rate jackson networks with abandonment. *Probability in the Engineering and Informational Sciences*, 31(1):1–42, 2017.
- [36] Qiang Ren and G Ramamurthy. A hybrid model and measurement based connection admission control and bandwidth allocation scheme for atm networks. In *Global Telecommunications Conference, 1998. GLOBECOM 1998. The Bridge to Global Integration. IEEE*, volume 4, pages 2030–2038. IEEE, 1998.
- [37] Sanjay Shakkottai and R. Srikant. Many-Sources Delay Asymptotics with Applications to Priority Queues. *Queueing Systems*, 39(2):183–200, 2001.
- [38] Ness B Shroff and Mischa Schwartz. Improved loss calculations at an atm multiplexer. *IEEE/ACM Transactions on Networking (TON)*, 6(4):411–421, 1998.
- [39] Eleni Vatamidou, Ivo Adan, Maria Vlasidou, and Bert Zwart. Corrected phase-type approximations for the workload of the map/g/1 queue with heavy-tailed service times. *ACM SIGMETRICS Performance Evaluation Review*, 41(2):53–55, 2013.
- [40] Tao Yang and Danny HK Tsang. A novel approach to estimating the cell loss probability in an atm multiplexer loaded with homogeneous on-off sources. *Communications, IEEE Transactions on*, 43(1):117–126, 1995.