Diffusion Approximations for Ridesharing Systems with Travel Times

Amir A. Alwan and Baris Ata

Booth School of Business, The University of Chicago

Abstract

This paper develops a closed queueing network model of a ride-hailing system by viewing cars as the jobs moving through various nodes of the network perpetually. The queueing network model crucially incorporates the travel times between different nodes of the network. We prove a novel heavy traffic limit theorem for this queueing network, providing an approximation for the original ride-hailing system.

Keywords: ridesharing, diffusion approximations, heavy traffic analysis, queueing theory

1 Introduction

This paper formulates and studies a closed queueing network motivated by ride-hailing systems such as Uber, Lyft, and Didi Chuxing. In these systems, customers request rides via a mobile application to travel from their current (pick-up) location to their final (drop-off) destination. Simultaneously, drivers wait for the mobile application to match them with the customer requests. Consequently, cars circulate through different regions of the city by picking up customers and delivering them to their desired destinations.

We consider a city partitioned into a finite set of geographical regions. Each such region constitutes a pick-up and a drop-off location. Crucially, our model incorporates the travel times between different regions of the city. We also allow customer heterogeneity by allowing customers in a region to have different destinations. Figure 1 displays an illustrative example. Namely, it partitions the map of New York City into nine geographical regions. For this example, our model tracks the movements of cars between the nine regions as they pick up arriving customers and

Figure 1: A map of New York City partitioned into nine geographical regions.[1]

deliver them to their destinations. However, before a customer is dropped off, there is a time delay due to the travel times between the pick-up and drop-off locations. This is an important feature of our model.

To be specific, the queueing network model has single-server and infinite-server nodes. The single-server nodes correspond to different regions of the city, whereas the infinite-server nodes model the travel times between different regions. Cars move between different server nodes according to a probabilistic routing structure; see Section 2 for details.

The ultimate goal of a ride-hailing system is to maximize its profit. As such, it is important for a ridesharing company to exert some control over the movement of cars on their platform. However, to effectively exert control requires first understanding how cars on their platform organically move throughout the city. This paper provides a first step in this important direction. In particular, we develop a diffusion approximation framework for ride-hailing systems with travel times. This approximation framework is justified via a weak convergence result for the queue length processes corresponding to the cars in the network; see Theorem 1.

[1]The figure was obtained from Ata et al. (2020).

1.1 Literature Review

Our paper is related to two streams of literature. The first stream is on the modeling and analysis of ride-hailing networks, while the second stream is on the heavy traffic analysis of queueing network models.

The literature on ride-hailing has greatly expanded in recent years. A majority of the operational work on ride-hailing is on how pricing, matching, and car repositioning can impact system performance. The effect of pricing in ridesharing has been studied in Banerjee et al. (2015), Banerjee et al. (2017), Cachon et al. (2017), Besbes et al. (2019), Bimpikis et al. (2019), and Ata et al. (2019), among others. Banerjee et al. (2015) studies dynamic pricing for a single-region system. They show that system performance under any dynamic pricing policy cannot exceed the performance under the optimal static pricing policy. Banerjee et al. (2017) designs pricing policies through a general approximation framework and show that the approximation ratio of the resulting pricing policy improves as the number of cars in each region grows. Despite the negative publicity of surge pricing, Cachon et al. (2017) demonstrate that surge pricing in a ridesharing platform can actually make both the platform and the customers better off. Besbes et al. (2019) also studies the problem of pricing in ridesharing systems with price sensitive customers and drivers.

Both Bimpikis et al. (2019) and Ata et al. (2019) use spatial search models to study spatial pricing in ride-hailing networks with strategic drivers. Predating these papers, Lagos (2000) is one of the first papers to study spatial search frictions in the taxi industry. This paper highlights how search frictions endogenously arise as a result of the strategic movement of taxi drivers. Building on this paper, Bimpikis et al. (2019) studies spacial pricing of ridesharing systems analytically. They show that when demand patterns in the network are not "balanced," spatial pricing can be very beneficial by increasing consumer surplus. On the other hand, Ata et al. (2019) take an empirical approach to study how spatial pricing and search frictions can impact the taxi market in New York City. To be specific, they use a mean field model to represent the taxis in the system and show how origin and origin-destination based pricing can reduce search frictions and increase consumer surplus. They do so empirically by performing a counterfactual analysis using data of taxi trips in New York City. Other related papers that study pricing in ridesharing include Chen and Sheldon

(2016), Castillo et al. (2018), Lu et al. (2018), Hu and Zhou (2019), Gokpinar and Selcuk (2019), Garg and Nazerzadeh (2019), Hu et al. (2019), and Afèche et al. (2020).

Özkan and Ward (2019), Özkan (2018), and Banerjee et al. (2019) study matching (or dispatch control) in ridesharing platforms. In particular, Özkan and Ward (2019) models a ridesharing system as a non-stationary open queueing network where both drivers and customers exogenously arrive to the system. They propose a matching policy based on the solution to a continuous linear program and demonstrate that this policy is asymptotically optimal (in the fluid scale) in a large market regime where the arrival rates of drivers and customers are large. Özkan (2018) studies both the pricing and matching in ridesharing systems. The author demonstrates that joint pricing and matching provides significant performance increase over only optimizing over pricing or only over matching decisions. Finally, Banerjee et al. (2019) studies dispatch control in ridesharing systems. They model ridesharing systems as a closed queueing network model and proposes a family of state-dependent dispatch policies called Scaled MaxWeight policies. Under a complete resource pooling assumption, they show that the proportion of dropped demand under any Scaled MaxWeight policy decays exponentially. Other related papers include Agatz et al. (2017), Lam and Liu (2017), Korolko et al. (2019), Guda and Subramanian (2019), Kanoria and Qian (2019), and Bertsimas et al. (2019).

Other work considers repositioning control, including that of Braverman et al. (2019), Afèche et al. (2018), and Ata et al. (2020), among others. Braverman et al. (2019) models a ridesharing system as a closed queueing network and study it under fluid scaling. They focus on "empty-car routing" where a car without a customer can be repositioned to another region of the city. This paper proves that the solution to a suitable linear program serves as an upper bound for the utility obtained under any state-dependent routing policy in the finite-car system. From a modeling perspective, Braverman et al. (2019) is closely related to ours because both explicitly model the travel times between city regions. However, an important difference is that our work allows for stochastic variability, while Braverman et al. (2019) studies a fluid-based model. On the other hand, Afèche et al. (2018) studies both demand-side admission control of customers as well as supply-side repositioning control of cars. In particular, they develop several insights on the interplay between centralized and de-centralized admission and repositioning control on the effect they have on the system. Contrary to the previous two papers, Ata et al. (2020) considers a ride-hailing system with both repositioning and matching control. In particular, they model a ridesharing system as a stochastic processing network where the activities in the network correspond to repositioning cars from one area to another and dynamically matching customers with cars. By employing the general approach outlined in Harrison (2003), Ata et al. (2020) proposes a solution to the original stochastic control problem by studying a the related Brownian control problem, which arises as a heavy traffic approximation to the original system. Other related papers include He et al. (2019) and Hosseini et al. (2020).

On the other hand, from a methodological perspective our paper involves the analysis of queueing network models, but more specifically on the asymptotic analysis of closed queueing networks containing infinite-server nodes. Related work includes that of Krichagina and Puhalskii (1997), Kogan and Lipster (1993), Kogan et al. (1986), and Smorodinskii (1986). For example, Krichagina and Puhalskii (1997) studies a closed queueing model containing one single-server queue and one infinite-server queue (with a general service time distribution), and derive heavy traffic limits. On the other hand, Kogan and Lipster (1993) study a closed queueing network with many single-server queues and one infinite-server queue. In that paper, all the single-server queues except one are in a "light-usage regime," while the other single-server queue is studied in both a "moderate-usage regime" and a "heavy-usage regime." They prove a state-space collapse result for the single-server queues in the light-usage regime, but prove diffusion approximation results for the single-server queue in the moderate-usage and heavy-usage regimes. Limit results for the queue length processes in closed queueing results are also established in the latter two papers Kogan et al. (1986) and Smorodinskii (1986). Other papers that employ a similar type of analysis to ours include Ata and Kumar (2005), Ata and Lin (2008), Ata and Olsen (2009, 2013), and Reed and Ward (2008), among others.

To the best of our knowledge, this is the first paper to develop a diffusion approximation for a ride-hailing system while incorporating travel times between regions. Ata et al. (2020) also proposes a diffusion approximation for ride-sharing systems,[2] but does not incorporate travel times. Travel times are important from a practical perspective. However, with the exception of Braverman

[2] We are not aware of any other work proving diffusion approximations for ride-hailing systems.

et al. (2019), most of the ride-hailing literature appears to ignore them effectively by assuming instantaneous pick-up and drop-off of customers. Our work builds on but differs from Braverman et al. (2019). Namely, we model the uncertainty through the second moments of stochastic primitives, whereas Braverman et al. (2019) work with a deterministic model.

The rest of the paper is organized as follows. Section 2 formalizes our ridesharing model by introducing the model primitives and making a sample-path construction of the queue-length processes describing our ridesharing network. Section 3 puts forth our heavy traffic assumption and diffusion scaling regime, and states the main result of the paper (Theorem 1). Section 4 outlines the main tools needed to prove Theorem 1. Section 5 is devoted to a proof of Theorem 1, which involves proving convergence of the fluid scaled processes. Section 6 concludes and discusses future research. Relevant notation and technical preliminaries are given in Appendix A. Proofs of the results in Section 4, as well as all auxiliary results and derivations, are given in Appendices B, C, and D.

2 The Ridesharing Model

The model has three key components: (i) J city regions where customers are picked up from and dropped off to, (ii) K travel times that correspond to the time that rides from one city region to another take, and (iii) a probabilistic routing structure that governs movement of cars. To be more specific, our model contains two levels of probabilistic routing. The first level is from the city regions to travel time nodes, while the second level is from travel time nodes to city regions. In the first level, a customer arriving to region $j \in [J]$ requests a ride requiring travel time $k \in [K]$ with probability p_{jk}. In the second level, a car taking travel node k will, with probability q_{kj}, deliver the customer to region j. It should be noted that the arbitrary nature of the routing structure and the arbitrary number of travel nodes makes our model almost completely general. In particular, our model subsumes the $K = J^2$ case, where an arriving customer to region j will get routed to region $j' \in [J]$ with probability $p_{jj'}$. This occurs when the travel times between any two regions of the city are different. The two-level probabilistic routing with K travel times not only allows for arbitrarily general routing structures but also provides modeling flexibility. It also can help lower the dimension of the state space.

Each city region is modeled by a single-server queue, where a service completion corresponds to a customer arrival to the region who is subsequently picked up by a car. On the other hand, the K travel time nodes are modeled by infinite-server queues. A service completion at an infinite-server queue corresponds to a car finishing its travel with a customer from the pick-up location to the drop-off location.

In what follows, we consider a sequence of systems indexed by the total number of cars n. Each system is a closed queueing network with J single-server queues, K infinite-server queues, and the probabilistic routing structure mentioned above. We study this sequence of systems in the heavy traffic asymptotic regime as $n \to \infty$. A superscript of n will be attached to various quantities of interest to indicate they correspond to the nth system.

2.1 Model Primitives

The service rate at a single-server queue reflects the demand rate at the corresponding city region. We consider a regime in which both the number of cars and the demand gets large. Thus, the service rate μ_j^n at the single-server queue $j \in [J]$ in the nth system is given by

$$\mu_j^n = n\mu_j, \tag{1}$$

where $\mu_j > 0$ is a fixed parameter for $j \in [J]$. The service rates at the infinite-server queues do not vary with n. In particular, the service rate of infinite server $k \in [K]$ is denoted by $\eta_k > 0$.

To facilitate the description of the system dynamics in Section 2.2, we define the following Poisson processes (one for each queue): For $j \in [J]$ and $k \in [K]$ we let

$$N_j = \{N_j(t) : t \geq 0\} \qquad \text{and} \qquad M_k = \{M_k(t) : t \geq 0\} \tag{2}$$

be independent rate-one Poisson processes.

To model the probabilistic routing of cars, we take as given the stochastic matrices

$$P = (p_{jk}) \in \mathbb{R}^{J \times K} \qquad \text{and} \qquad Q = (q_{kj}) \in \mathbb{R}^{K \times J} \tag{3}$$

representing routing from single-server queues to infinite-server queues and from infinite-server queues to single-server queues, respectively. To be more specific, for $j \in [J]$ and $k \in [K]$ we let

$$\phi_j = \{\phi_j(l) : l \geq 1\} \qquad \text{and} \qquad \psi_k = \{\psi_k(l) : l \geq 1\} \tag{4}$$

denote independent sequences of i.i.d. random (routing) vectors. Their probability distributions are given by

$$P\left(\phi_j(1) = e_k\right) = p_{jk} \qquad \text{and} \qquad P\left(\psi_k(1) = e_j\right) = q_{kj}, \tag{5}$$

where e_k and e_j are the kth and jth standard unit basis vector in $\mathbb{R}^K$ and $\mathbb{R}^J$, respectively. We assume that these random routing vectors are independent of all other stochastic model primitives. The associated cumulative routing processes $\Phi_{jk} = \left\{\Phi_{jk}(m) : m \geq 1\right\}$ and $\Psi_{kj} = \left\{\Psi_{kj}(m) : m \geq 1\right\}$ are defined by the following:

$$\Phi_{jk}(m) := \sum_{l=1}^{m} \phi_{jk}(l) \qquad \text{and} \qquad \Psi_{kj}(m) := \sum_{l=1}^{m} \psi_{kj}(l), \tag{6}$$

where $\phi_{jk}(l)$ and $\psi_{kj}(l)$ are the kth and jth components of $\phi_j(l)$ and $\psi_k(l)$, respectively. In particular, $\Phi_{jk}(m)$ represents the total number of customers that are routed from single-server queue j to infinite-server queue k among the first m customers arriving to region j. The expression $\Psi_{kj}(m)$ can be interpreted similarly.

2.2 The System State and Its Evolution

For $j \in [J]$ we denote by $Q_j^n(t)$ the number of jobs in the jth single-server queue at time t in the nth system. Similarly, for $k \in [K]$ we denote by $V_k^n(t)$ the number of jobs in the kth infinite-server queue at time t in the nth system. Equations (8) – (9) below define these processes Q_j^n and V_k^n in the natural way. Also, we take as given random variables $Q_j^n(0)$ and $V_k^n(0)$ representing the initial number of cars in the network. We assume that the collection of these random variables is independent of all other stochastic primitives and that

$$\sum_{j=1}^{J} Q_j^n(0) + \sum_{k=1}^{K} V_k^n(0) = n, \quad \text{a.s.}, \tag{7}$$

as required for any closed system.

By simple conservation of flow, we write

$$Q_j^n(t) = Q_j^n(0) + A_j^n(t) - D_j^n(t), \qquad j \in [J], \tag{8}$$

$$V_k^n(t) = V_k^n(0) + E_k^n(t) - F_k^n(t), \qquad k \in [K], \tag{9}$$

where $A_j^n = \left\{A_j^n(t) : t \geq 0\right\}$ and $D_j^n = \left\{D_j^n(t) : t \geq 0\right\}$ denote the arrival and departure processes for single-server queue j. Similarly, $E_k^n = \left\{E_k^n(t) : t \geq 0\right\}$ and $F_k^n = \left\{F_k^n(t) : t \geq 0\right\}$ denote the arrival and departure processes for infinite-server queue k. That is, A_j^n tracks the total number of cars that have arrived to region j, while D_j^n tracks the total number of cars that have departed from region j with a customer. (The processes E_k^n and F_k^n are interpreted similarly.) The following equations define these arrival and departure processes:

$$A_j^n(t) := \sum_{k=1}^{K} \Psi_{kj}\left(F_k^n(t)\right), \tag{10}$$

$$E_k^n(t) := \sum_{j=1}^{J} \Phi_{jk}\left(D_j^n(t)\right), \tag{11}$$

$$D_j^n(t) := N_j\left(n\mu_j T_j^n(t)\right), \tag{12}$$

$$F_k^n(t) := M_k\left(\eta_k \int_0^t V_k^n(s)\, ds\right), \tag{13}$$

where Φ_{jk} and Ψ_{kj} are the cumulative routing processes defined in Equation (6) and where

$$T_j^n = \left\{T_j^n(t) : t \geq 0\right\} \tag{14}$$

is the busy time process for the jth single-server queue. To be specific, $T_j^n(t)$ is the cumulative amount of time the server is busy over $[0,t]$ at single-server queue $j \in [J]$. The corresponding idleness process $I_j^n = \left\{I_j^n(t) : t \geq 0\right\}$ at single-server queue j is defined by

$$I_j^n(t) := t - T_j^n(t). \tag{15}$$

Note that by Equations (7) and (10) – (13), it is straightforward to verify that

$$\sum_{j=1}^{J} Q_j^n(t) + \sum_{k=1}^{K} V_k^n(t) = n, \quad \text{a.s. for all } t \geq 0. \tag{16}$$

We assume that the following hold for each $j \in [J]$:

$$I_j^n \text{ is continuous and nondecreasing with } I_j^n(0) = 0, \tag{17}$$

$$\int_0^\infty \mathbb{1}_{\left\{Q_j^n(s)>0\right\}} \, dI_j^n(s) = 0, \tag{18}$$

$$Q_j^n(t) \geq 0 \text{ for all } t \geq 0, \tag{19}$$

$$I_j^n(t) - I_j^n(s) \leq t - s \text{ for all } 0 \leq s \leq t. \tag{20}$$

Clearly, we must have $Q_j(t) \geq 0$ for $j \in [J]$. Also, we restrict attention to work-conserving policies. That is, the server idleness does not increase as long as the queue is not empty. These are reflected in Equations (18) – (19). Lastly, Equations (17) and (20) are natural consequences of the interpretation of I_j^n as server idleness.

3 The Main Result: A Heavy Traffic Limit Theorem

As a preliminary to stating our main result, we first introduce the heavy traffic assumption.

Assumption 1 (Heavy Traffic Assumption)**.** The following conditions hold. First,

$$\sum_{k=1}^{K} \sum_{i=1}^{J} \mu_i p_{ik} q_{kj} = \mu_j \quad \text{for} \quad j \in [J]. \tag{21}$$

Second, letting $m_k = \eta_k^{-1} \sum_{j=1}^{J} \mu_j p_{jk}$ for $k \in [K]$,

$$\sum_{k=1}^{K} m_k = 1. \tag{22}$$

Roughly speaking, the first condition assumes that every single-server is fully utilized, whereas the second condition corresponds to assuming almost all cars are in transit.[3] To elaborate on the first condition, note that $n\mu_i p_{ik}$ is the rate of jobs from single-server queue i to infinite-server queue k, i.e., the rate of customer arrivals into region i that require travel time k. The sum $\sum_{i=1}^{J} n\mu_i p_{ik}$ therefore represents the total rate of jobs leaving the city regions that require travel time k. Then, the expression $q_{kj} \sum_{i=1}^{J} n\mu_i p_{ik}$ represents the rate of cars entering region j from travel node k, and thus the entire sum $\sum_{k=1}^{K} q_{kj} \sum_{i=1}^{J} n\mu_i p_{ik}$ represents the total rate of cars entering region j from travel nodes. Because $n\mu_j$ is the customer arrival rate to region j, the first condition stats the

[3] These heavy traffic conditions can easily be generalized to limit conditions. To state it, suppose that μ_j depends on n, for which we write μ_j^n, and that $\mu_j^n \to \mu_j \in \mathbb{R}$ as $n \to \infty$ for each $j \in [J]$. Then Equation (21) can be replaced with the following condition: $\sqrt{n}\left[\sum_{k=1}^{K} \sum_{i=1}^{J} \mu_i^n p_{ik} q_{kj} - \mu_j^n\right] \to c_j \in \mathbb{R}$ for all $j \in [J]$. The main convergence result (Theorem 1) remains unchanged, except that we have an additional drift term c_j. A similar limit condition can be used in place of Equation (22).

arrival rate of cars is balanced out by the customer demand in each region of the city.

We expect the first condition to hold in large cities under optimal spatial pricing provided the demand is elastic. To be more specific, if it was the case that for some region j that $\sum_{k=1}^{K}\sum_{i=1}^{J}\mu_i p_{ik} q_{kj} \gg \mu_j$, then the supply of cars in region j greatly exceeds customer demand. By decreasing the price for rides requested from region j, more people in region j will switch from their current form of transportation to ridesharing as their means for travel. This will increase the customer arrival rate μ_j and would eventually balance out supply. It would also increase the profit because demand is elastic. On the other hand, if it was the case that for some region j that $\sum_{k=1}^{K}\sum_{i=1}^{J}\mu_i p_{ik} q_{kj} \ll \mu_j$, then customer demand greatly exceeds the supply of cars in region j. In this case, the ridesharing platform could increase prices for rides requested in region j, while maintaining the current prices in all other regions $i \in [J] \setminus \{j\}$.

To shed light on the second condition, note that the arrival rate to the kth infinite-server queue is $n\sum_{j=1}^{J}\mu_j p_{jk}$ whereas its service rate is η_k. Based on intuition from the classical $M/M/\infty$ queue, we expect the steady-state average queue length at the kth infinite-server to be $n\sum_{j=1}^{J}\mu_j p_{jk}/\eta_k$. Thus, the expected fraction of jobs at the kth infinite-server is m_k. So the condition $\sum_{k=1}^{K} m_k = 1$ means (almost) all jobs are at the infinite-server queues, i.e., (almost) all cars are in transit. This is consistent with the first condition because of the following: For each single-sever queue, the arrival and service rates are equal and of order n. Thus, by the intuition from the central limit theorem, we expect the queue lengths to be of order $\sqrt{n}$ for each single-server queue. Consequently, the total number of jobs in the single-server queues is of order $\sqrt{n}$, meaning (almost) all jobs are at the infinite-server queues because this is a closed network.

To facilitate the analysis to follow, we next define the following diffusion and fluid scaled processes:

Diffusion Scaled Processes: For $j \in [J]$, $k \in [K]$, and $t \geq 0$, let

$$\hat{Q}_j^n(t) := \frac{Q_j^n(t)}{\sqrt{n}}, \tag{23}$$

$$\hat{V}_k^n(t) := \frac{V_k^n(t) - nm_k}{\sqrt{n}}, \tag{24}$$

$$\hat{I}_j^n(t) := \sqrt{n} I_j^n(t), \tag{25}$$

$$\hat{T}_j^n(t) := \sqrt{n} T_j^n(t), \tag{26}$$

$$\hat{\Phi}_{jk}^n(t) := \frac{\Phi_{jk}(\lfloor nt \rfloor) - p_{jk} nt}{\sqrt{n}}, \tag{27}$$

$$\hat{\Psi}_{kj}^n(t) := \frac{\Psi_{kj}(\lfloor nt \rfloor) - q_{kj} nt}{\sqrt{n}}, \tag{28}$$

$$\hat{N}_j^n(t) := \frac{N_j(nt) - nt}{\sqrt{n}}, \tag{29}$$

$$\hat{M}_k^n(t) := \frac{M_k(nt) - nt}{\sqrt{n}}. \tag{30}$$

Fluid Scaled Processes: For $j \in [J]$, $k \in [K]$, and $t \geq 0$, let

$$\bar{Q}_j^n(t) := \frac{\hat{Q}_j^n(t)}{\sqrt{n}}, \tag{31}$$

$$\bar{V}_k^n(t) := \frac{\hat{V}_k^n(t)}{\sqrt{n}}, \tag{32}$$

$$\bar{\bar{V}}_k^n(t) := \frac{V_k^n(t)}{n}, \tag{33}$$

$$\bar{N}_j^n(t) := \frac{N_j(nt)}{n}, \tag{34}$$

$$\bar{M}_k^n(t) := \frac{M_k(nt)}{n}. \tag{35}$$

By applying the scaling in Equations (23) – (35) and using the heavy traffic conditions, it is straightforward to show that the following diffusion scaled system equations hold:

$$\begin{aligned} \hat{Q}_j^n(t) = \hat{Q}_j^n(0) &+ \sum_{k=1}^K \hat{\Psi}_{kj}^n\left(\bar{M}_k^n\left(\eta_k \int_0^t \bar{\bar{V}}_k^n(s)\,ds\right)\right) - \hat{N}_j^n\left(\mu_j T_j^n(t)\right) \\ &+ \sum_{k=1}^K q_{kj} \hat{M}_k^n\left(\eta_k \int_0^t \bar{\bar{V}}_k^n(s)\,ds\right) + \sum_{k=1}^K q_{kj}\eta_k \int_0^t \hat{V}_k^n(s)\,ds + \mu_j \hat{I}_j^n(t), \end{aligned} \tag{36}$$

$$\begin{aligned} \hat{V}_k^n(t) = \hat{V}_k^n(0) &+ \sum_{j=1}^J \hat{\Phi}_{jk}^n\left(\bar{N}_j^n\left(\mu_j T_j^n(t)\right)\right) - \hat{M}_k^n\left(\eta_k \int_0^t \bar{\bar{V}}_k^n(s)\,ds\right) \\ &+ \sum_{j=1}^J p_{jk} \hat{N}_j^n\left(\mu_j T_j^n(t)\right) - \eta_k \int_0^t \hat{V}_k^n(s)\,ds - \sum_{j=1}^J p_{jk}\mu_j \hat{I}_j^n(t). \end{aligned} \tag{37}$$

See Appendix D for a straightforward, albeit tedious, derivation of the diffusion scaled system equations (36) and (37). Defining

$$\begin{aligned} \hat{\xi}_j^n(t) := \hat{Q}_j^n(0) &+ \sum_{k=1}^K \hat{\Psi}_{kj}^n\left(\bar{M}_k^n\left(\eta_k \int_0^t \bar{\bar{V}}_k^n(s)\,ds\right)\right) - \hat{N}_j^n\left(\mu_j T_j^n(t)\right) \\ &+ \sum_{k=1}^K q_{kj} \hat{M}_k^n\left(\eta_k \int_0^t \bar{\bar{V}}_k^n(s)\,ds\right), \end{aligned} \tag{38}$$

$$\hat{\zeta}_k^n(t) := \hat{V}_k^n(0) + \sum_{j=1}^{J} \hat{\Phi}_{jk}^n \left(\bar{N}_j^n \left(\mu_j T_j^n(t)\right)\right) - \hat{M}_k^n \left(\eta_k \int_0^t \bar{\bar{V}}_k^n(s)\, ds\right) + \sum_{j=1}^{J} p_{jk} \hat{N}_j^n \left(\mu_j T_j^n(t)\right), \tag{39}$$

we see that Equations (36) – (37) can be rewritten as follows:

$$\hat{Q}_j^n(t) = \hat{\xi}_j^n(t) + \sum_{k=1}^{K} q_{kj}\eta_k \int_0^t \hat{V}_k^n(s)\, ds + \mu_j \hat{I}_j^n(t), \tag{40}$$

$$\hat{V}_k^n(t) = \hat{\zeta}_k^n(t) - \eta_k \int_0^t \hat{V}_k^n(s)\, ds - \sum_{j=1}^{J} p_{jk}\mu_j \hat{I}_j^n(t). \tag{41}$$

Furthermore, by applying the fluid scaling in Equations (31) – (32) to Equations (40) – (41), we obtain the following fluid scaled system equations:

$$\bar{Q}_j^n(t) = \bar{\xi}_j^n(t) + \sum_{k=1}^{K} q_{kj}\eta_k \int_0^t \bar{V}_k^n(s)\, ds + \mu_j I_j^n(t), \tag{42}$$

$$\bar{V}_k^n(t) = \bar{\zeta}_k^n(t) - \eta_k \int_0^t \bar{V}_k^n(s)\, ds - \sum_{j=1}^{J} p_{jk}\mu_j I_j^n(t), \tag{43}$$

where

$$\bar{\xi}_j^n(t) := n^{-1/2}\hat{\xi}_j^n(t), \tag{44}$$

$$\bar{\zeta}_k^n(t) := n^{-1/2}\hat{\zeta}_k^n(t). \tag{45}$$

We make the following assumptions on the initial conditions:

Assumption 2 (Joint Convergence of the Initial Conditions)**.** As $n \to \infty$,

$$\left(\hat{Q}_1^n(0), \ldots, \hat{Q}_J^n(0), \hat{V}_1^n(0), \ldots, \hat{V}_K^n(0)\right) \Rightarrow (Q_1(0), \ldots, Q_J(0), V_1(0), \ldots, V_K(0)) \in D^{J+K}. \tag{46}$$

The result below establishes the joint convergence of the diffusion scaled queue length processes and the idleness processes at the single-server stations to a multidimensional diffusion process. To facilitate the statement of the result, let $(\xi^*, \zeta^*) \equiv (\xi_1^*, \ldots, \xi_J^*, \zeta_1^*, \ldots, \zeta_K^*)$ be a $(0, \Sigma)$ Brownian motion with initial state $(\xi^*(0), \zeta^*(0)) = (Q(0), V(0))$, where $(Q(0), V(0)) \equiv (Q_1(0), \ldots, Q_J(0), V_1(0), \ldots, V_K(0))$. Note by construction and Equation (46) that $\xi^*(0) \geq 0$ and $\sum_{j=1}^{J} \xi_j^*(0) + \sum_{k=1}^{K} \zeta_k^*(0) = 0$. The covariance matrix $\Sigma \in \mathbb{R}^{(J+K)\times(J+K)}$ is given by

$$\Sigma_{j,j} = \sum_{k=1}^{K} q_{kj}(1 - q_{kj})\eta_k m_k + \mu_j + \sum_{k=1}^{K} q_{kj}^2 \eta_k m_k, \qquad j \in [J], \tag{47}$$

$$\Sigma_{J+k,J+k} = \sum_{j=1}^{J} p_{jk}(1 - p_{jk})\mu_j + \eta_k m_k + \sum_{j=1}^{J} p_{jk}^2 \mu_j, \qquad k \in [K], \tag{48}$$

$$\Sigma_{i,j} = \sum_{k=1}^{K} q_{ki} q_{kj} \eta_k m_k, \qquad i, j \in [J],\ i \neq j, \tag{49}$$

$$\Sigma_{j,J+k} = -p_{jk}\mu_j - q_{kj}\eta_k m_k, \qquad j \in [J],\ k \in [K], \tag{50}$$

$$\Sigma_{J+l,J+k} = \sum_{j=1}^{J} p_{jl} p_{jk} \mu_j, \qquad l, k \in [K],\ l \neq k. \tag{51}$$

Theorem 1 (Joint Convergence of Diffusion Scaled Processes)**.** *We have that as $n \to \infty$,*

$$\left(\hat{Q}^n, \hat{I}^n, \hat{V}^n\right) \Rightarrow (Q^*, I^*, V^*) \quad in \quad D^{2J+K}, \tag{52}$$

where Q^*, I^*, and V^* are multidimensional diffusion processes satisfying the following for all $j \in [J]$, $k \in [K]$, and $t \geq 0$:

$$Q_j^*(t) = \xi_j^*(t) + \sum_{k=1}^{K} q_{kj}\eta_k \int_0^t V_k^*(s)\,ds + \mu_j I_j^*(t) \geq 0, \tag{53}$$

$$V_k^*(t) = \zeta_k^*(t) - \eta_k \int_0^t V_k^*(s)\,ds - \sum_{j=1}^{J} p_{jk}\mu_j I_j^*(t), \tag{54}$$

$$\int_0^\infty \mathbb{1}_{\left\{Q_j^*(t)>0\right\}}\,dI_j^*(t) = 0. \tag{55}$$

4 Auxiliary Results

This section establishes the existence of suitably defined continuous functions that will aid in the proof of Theorem 1 via a continuous mapping argument; see Appendix B for the proofs of the results in this section. To that end, fix $\xi \equiv (\xi_1, \ldots, \xi_J) \in D^J$ and $\zeta \equiv (\zeta_1, \ldots, \zeta_K) \in D^K$ such that

$$\sum_{j=1}^{J} \xi_j(t) + \sum_{k=1}^{K} \zeta_k(t) = 0, \qquad \text{for all } t \geq 0, \tag{56}$$

$$\xi_j(0) \geq 0, \qquad \text{for all } j \in [J]. \tag{57}$$

The following collection of equations corresponds to our closed ridesharing network with J single-server stations and K infinite-server stations. That is, for $j \in [J]$, $k \in [K]$, and $t \geq 0$ we consider the following set of equations:

$$x_j(t) = \xi_j(t) + \sum_{k=1}^{K} q_{kj}\eta_k \int_0^t y_k(s)\,ds + \mu_j u_j(t) \in [0, \infty), \tag{58}$$

$$y_k(t) = \zeta_k(t) - \eta_k \int_0^t y_k(s)\,ds - \sum_{j=1}^{J} p_{jk}\mu_j u_j(t), \tag{59}$$

$$\sum_{j=1}^{J} x_j(t) + \sum_{k=1}^{K} y_k(t) = 0, \tag{60}$$

$$u_j \text{ is nondecreasing with } u_j(0) = 0, \tag{61}$$

$$\int_0^\infty \mathbb{1}_{\{x_j(t)>0\}}\, du_j(t) = 0. \tag{62}$$

We first establish the existence and uniqueness of solutions to these equations (for fixed ξ and ζ):

Proposition 1. *For every $(\xi,\zeta) \in D^{J+K}$ satisfying Equations (56) – (57), there exists a unique $(x,u,y) \in D^{2J+K}$ satisfying Equations (58) – (62).*

The following is now immediate from Proposition 1:

Corollary 1. *There exists a unique tuple of functions $(f,g,h) : D^{J+K} \to D^{2J+K}$ such that whenever $(\xi,\zeta) \in D^{J+K}$ satisfies Equations (56) – (57), $(f(\xi,\zeta), g(\xi,\zeta), h(\xi,\zeta))$ satisfies Equations (58) – (62).*

See Appendix B for a characterization of these functions f, g, and h. The final result of this section establishes the continuity of the mappings defined in Corollary 1:

Proposition 2. *The mapping $(f,g,h) : D^{J+K} \to D^{2J+K}$ from Corollary 1 is continuous when both the domain and range spaces are endowed with the Skorokhod J_1 topology.*

5 Proof of Theorem 1

This section contains the main convergence results of this paper, culminating with a proof of Theorem 1. To begin, Section 5.1 proves convergence of the fluid scaled processes. These results are necessary because several of the fluid scaled processes serve as random time changes in the diffusion scaled equations. Finally, in Section 5.2 convergence of the process $\left(\hat{\xi}^n, \hat{\zeta}^n\right)$ [given in Equations (38) – (39)] is proven. This, combined with a continuous mapping argument, allows us to complete the proof of Theorem 1.

5.1 Convergence of Fluid Scaled Processes

We begin by establishing convergence of the fluid scaled processes.

Lemma 1. *As* $n \to \infty$, $\left(\bar{\xi}^n, \bar{\zeta}^n\right) \Rightarrow \mathbf{0} \in D^{J+K}$.

Proof. To prove that $\left(\bar{\xi}^n, \bar{\zeta}^n\right) \Rightarrow \mathbf{0}$, it suffices to show [by, for example, Theorem 11.4.5 in Whitt (2002)] that $\bar{\xi}^n_j \Rightarrow 0$ and $\bar{\xi}^n_k \Rightarrow 0$ for all $j \in [J]$ and $k \in [K]$. On the other hand, to prove that $\bar{\xi}^n_j \Rightarrow 0$ and $\bar{\xi}^n_k \Rightarrow 0$, it is sufficient to show that for all $T > 0$,

$$\|\bar{\xi}^n_j\|_T \Rightarrow 0 \qquad \text{and} \qquad \|\bar{\zeta}^n_k\|_T \Rightarrow 0 \tag{63}$$

as random variables; see Lemma 5 in Appendix D for a proof of this claim. By Equations (38) – (39), the triangle inequality, and the facts that $\int_0^t \bar{\bar{V}}^n_k(s)\, ds \le t$ and $T^n_j(t) \le t$ for all $t \ge 0$, one readily checks that for all $T > 0$,

$$\|\hat{\xi}^n_j\|_T \le \|\hat{Q}^n_j(0)\|_T + \sum_{k=1}^K \|\hat{\Psi}^n_{kj}\left(\bar{M}^n_k(\eta_k \cdot)\right)\|_T + \|\hat{N}^n_j(\mu_j \cdot)\|_T + \sum_{k=1}^K \|\hat{M}^n_k(\eta_k \cdot)\|_T, \tag{64}$$

$$\|\hat{\zeta}^n_k\|_T \le \|\hat{V}^n_k(0)\|_T + \sum_{j=1}^J \|\hat{\Phi}^n_{jk}\left(\bar{N}^n_j(\mu_j \cdot)\right)\|_T + \|\hat{M}^n_k(\eta_k \cdot)\|_T + \sum_{j=1}^J \|\hat{N}^n_j(\mu_j \cdot)\|_T. \tag{65}$$

A straightforward application of Donsker's theorem, the functional central limit theorem for renewal processes, the random time change theorem, and the continuous mapping theorem can be used to show that the right hand sides of Equations (64) and (65) converge weakly to a nondegenerate limit. By this and the fact that $\bar{\xi}^n_j = n^{-1/2}\hat{\xi}^n_j$ and $\bar{\zeta}^n_k = n^{-1/2}\hat{\zeta}^n_k$, we obtain Equation (63). See Appendix D for additional details. □

Lemma 2. *As* $n \to \infty$, $\left(\bar{Q}^n, I^n, \bar{V}^n\right) \Rightarrow \mathbf{0} \in D^{J+K}$

Proof. Recall that [see Equations (42) – (43)]

$$\bar{Q}^n_j(t) = \bar{\xi}^n_j(t) + \sum_{k=1}^K q_{kj}\eta_k \int_0^t \bar{V}^n_k(s)\, ds + \mu_j I^n_j(t), \tag{66}$$

$$\bar{V}^n_k(t) = \bar{\zeta}^n_k(t) - \eta_k \int_0^t \bar{V}^n_k(s)\, ds - \sum_{j=1}^J p_{jk}\mu_j I^n_j(t). \tag{67}$$

Note that the process $(\bar{\xi}^n, \bar{\zeta}^n)$ satisfies

$$\xi^n_j(0) = \bar{Q}^n_j(0) \ge 0, \quad \text{a.s. for all } j \in [J]. \tag{68}$$

Recall that by Equation (17) the process I^n_j is nondecreasing with $I^n_j(0) = 0$. Furthermore, by

Equation (18),

$$\int_0^\infty \mathbb{1}_{\left\{\bar{Q}_j^n(t)>0\right\}}\, dI_j^n(t) = \int_0^\infty \mathbb{1}_{\left\{n^{-1}Q_j^n(t)>0\right\}}\, dI_j^n(t) = \int_0^\infty \mathbb{1}_{\left\{Q_j^n(t)>0\right\}}\, dI_j^n(t) = 0. \tag{69}$$

By Equations (66) – (69) and the uniqueness in Proposition 1 it follows that

$$\left(\bar{Q}^n, I^n, \bar{V}^n\right) = \left(f\left(\bar{\xi}^n, \bar{\zeta}^n\right), g\left(\bar{\xi}^n, \bar{\zeta}^n\right), h\left(\bar{\xi}^n, \bar{\zeta}^n\right)\right). \tag{70}$$

So, by Proposition 2, Lemma 1, the continuous mapping theorem, and Equation (70) we have

$$\left(\bar{Q}^n, I^n, \bar{V}^n\right) = \left(f\left(\bar{\xi}^n, \bar{\zeta}^n\right), g\left(\bar{\xi}^n, \bar{\zeta}^n\right), h\left(\bar{\xi}^n, \bar{\zeta}^n\right)\right) \Rightarrow \left(f\left(\mathbf{0}\right), g\left(\mathbf{0}\right), h\left(\mathbf{0}\right)\right). \tag{71}$$

To complete the proof, we must show that $f(\mathbf{0}) = \mathbf{0}$, $g(\mathbf{0}) = \mathbf{0}$, and $h(\mathbf{0}) = \mathbf{0}$. If we can show that $\bar{V} := h(\mathbf{0}) = \mathbf{0}$, then by definition of $\bar{Q} := f(\mathbf{0})$ and $I := g(\mathbf{0})$,

$$\begin{aligned}
\bar{Q}_j &= \phi\left(\pi_j \circ \mathbf{0} + \sum_{k=1}^K q_{kj}\eta_k \int_0^\cdot \bar{V}_k(s)\,ds\right) = \phi\left(0 + \sum_{k=1}^K q_{kj}\eta_k \int_0^\cdot 0\,ds\right) = \phi(0+0) = 0,\\
I_j &= \mu_j^{-1}\psi\left(\pi_j \circ \mathbf{0} + \sum_{k=1}^K q_{kj}\eta_k \int_0^\cdot \bar{V}_k(s)\,ds\right) = \mu_j^{-1}\psi\left(0 + \sum_{k=1}^K q_{kj}\eta_k \int_0^\cdot 0\,ds\right) = \mu_j^{-1}\psi(0+0) = 0,
\end{aligned}$$

and the proof would be complete. By definition of $\bar{V} := h(\mathbf{0})$, for any $t \ge 0$ we have

$$\bar{V}_k(t) = -\eta_k \int_0^t \bar{V}_k(s)\,ds - \sum_{j=1}^J p_{jk}\psi\left(\sum_{l=1}^K q_{lj}\eta_l \int_0^\cdot \bar{V}_l(s)\,ds\right)(t), \tag{72}$$

for all $k = 1, \ldots, K$. But, for $T > 0$ fixed and any $0 \le t \le T$,

$$\begin{aligned}
\|\bar{V}_k\|_t &\le \eta_k \left\| \int_0^\cdot \bar{V}_k(s)\,ds \right\|_t + \sum_{j=1}^J \left\| \psi\left(\sum_{l=1}^K q_{lj}\eta_l \int_0^\cdot \bar{V}_l(s)\,ds\right)\right\|_t\\
&\le \eta \int_0^t \|\bar{V}_k\|_s\,ds + \sum_{j=1}^J\sum_{l=1}^K \eta_l \left\|\int_0^\cdot \bar{V}_l(s)\,ds\right\|_t\\
&\le \eta \int_0^t \max_{1\le k\le K}\|\bar{V}_k\|_s\,ds + \eta JK \int_0^t \max_{1\le k\le K}\|\bar{V}_k\|_s\,ds\\
&\le 2\eta JK \int_0^t \max_{1\le k\le K}\|\bar{V}_k\|_s\,ds,
\end{aligned} \tag{73}$$

where $\eta := \max_{1\le k\le K}\eta_k$. Therefore, by Equation (73),

$$\max_{1\le k\le K}\|\bar{V}_k\|_t \le 2\eta JK \int_0^t \max_{1\le k\le K}\|\bar{V}_k\|_s\,ds. \tag{74}$$

By Gronwall's inequality [see Lemma 4.1 in Pang et al. (2007)] and Equation (74), it follows that

$$\max_{1\le k\le K} \|\bar{V}_k\|_T = 0. \tag{75}$$

Since T was arbitrary, Equation (75) implies that $\bar{V} \equiv 0$. The proof is complete. □

Corollary 2. *As $n \to \infty$, $T^n \Rightarrow e \in D^J$, where $e(t) = (t, \dots, t)$ for all $t \ge 0$.*

Proof. By definition, $T^n = e - I^n$. The result then follows by Lemma 2 since $I^n \Rightarrow \mathbf{0}$. □

5.2 Convergence of Diffusion Scaled Processes

Lemma 3. *As $n \to \infty$, $\left(\hat{\xi}^n, \hat{\zeta}^n\right) \Rightarrow (\xi^*, \zeta^*) \in D^{J+K}$, where (ξ^*, ζ^*) is a $(0, \Sigma)$ Brownian motion with initial state $(Q(0), V(0))$ and covariance matrix Σ given by Equations (47) – (51).*

Proof. For this proof, let $e : [0,\infty) \to [0,\infty)$ denote the one-dimensional identity map $e(t) = t$. Recall from Equations (38) – (39) that $\hat{\xi}^n_j$ and $\hat{\zeta}^n_k$ are given by the following:

$$\hat{\xi}^n_j(t) := \hat{Q}^n_j(0) + \sum_{k=1}^K \hat{\Psi}^n_{kj}\left(\bar{M}^n_k\left(\eta_k \int_0^t \bar{\bar{V}}^n_k(s)\, ds\right)\right) - \hat{N}^n_j\left(\mu_j T^n_j(t)\right) + \sum_{k=1}^K q_{kj}\hat{M}^n_k\left(\eta_k \int_0^t \bar{\bar{V}}^n_k(s)\, ds\right),$$

$$\hat{\zeta}^n_k(t) := \hat{V}^n_k(0) + \sum_{j=1}^J \hat{\Phi}^n_{jk}\left(\bar{N}^n_j\left(\mu_j T^n_j(t)\right)\right) - \hat{M}^n_k\left(\eta_k \int_0^t \bar{\bar{V}}^n_k(s)\, ds\right) + \sum_{j=1}^J p_{jk}\hat{N}^n_j\left(\mu_j T^n_j(t)\right).$$

Note that by Lemma 2, Corollary 2, and derivations in Appendix D, we have the following convergence for the fluid scaled processes: $\bar{M}^n_k(\eta_k\,\cdot) \Rightarrow \eta_k e$, $\bar{N}^n_j(\mu_j\,\cdot) \Rightarrow \mu_j e$, $T^n_j \Rightarrow e$, and $\bar{\bar{V}}^n_k \Rightarrow m_k$ (note that $\bar{\bar{V}}^n_k(t) = \bar{V}^n_k(t) + m_k$). In addition, we have the following convergence for the diffusion scaled processes: $\hat{\Psi}^n_{kj} \Rightarrow \sqrt{q_{kj}(1-q_{kj})}B_{kj}$, $\hat{\Phi}^n_{kj} \Rightarrow \sqrt{p_{jk}(1-p_{jk})}B_{jk}$, $\hat{M}^n_k(\eta_k\,\cdot) \Rightarrow \eta_k^{1/2}B_k$, and $\hat{N}^n_j(\mu_j\,\cdot) \Rightarrow \mu_j^{1/2}B_j$, where B_{kj}, B_{jk}, B_k, and B_j are standard Brownian motions. Moreover, note that the function $H : D \to D$ defined by

$$H(x)(t) = \int_0^t x(s)\, ds \tag{76}$$

is continuous in the Skorokhod topology [see page 229 in Pang et al. (2007)], which implies that

$$H(\bar{\bar{V}}^n_k) \Rightarrow H(m_k) = m_k e. \tag{77}$$

By the above convergence results, Theorems 11.4.4 and 11.4.5 in Whitt (2002), and Equation (46) (and the fact that all stochastic primitives are independent), the joint processes

$$\Big(\hat{Q}_1^n(0),\ldots,\hat{Q}_J^n(0),\hat{V}_1^n(0),\ldots,\hat{V}_K^n(0)\hat{\Psi}_{11}^n,\ldots,\hat{\Psi}_{1J}^n,\ldots,\hat{\Psi}_{K1}^n,\ldots,\hat{\Psi}_{KJ}^n,$$
$$\hat{\Phi}_{11}^n,\ldots,\hat{\Phi}_{1K}^n,\ldots,\hat{\Phi}_{J1}^n,\ldots,\hat{\Phi}_{JK}^n\hat{N}_1^n\ldots,\hat{N}_J^n,\hat{M}_1^n,\ldots,\hat{M}_K^n\Big) \tag{78}$$

and

$$\left(T_1^n,\ldots,T_J^n,\bar{\bar{V}}_1^n,\ldots,\bar{\bar{V}}_K^n,\bar{N}_1^n,\ldots,\bar{N}_J^n,\bar{M}_1^n,\ldots,\bar{M}_K^n\right) \tag{79}$$

converge weakly to their appropriate limits. By the convergence of Equations (78) and (79), the random time change theorem, and the continuous mapping theorem, we get the joint weak convergence of

$$\left(\hat{\xi}_1^n,\ldots,\hat{\xi}_J^n,\hat{\zeta}_1^n,\ldots,\hat{\zeta}_K^n\right)\Rightarrow\left(\xi_1^*,\ldots,\xi_J^*,\zeta_1^*,\ldots,\zeta_K^*\right). \tag{80}$$

It is straightforward, but tedious, to derive the covariance matrix Σ, so we omit those details here. □

Finally, we conclude with a proof of Theorem 1:

Proof of Theorem 1. It is straightforward to show that the process $\left(\hat{\xi}^n,\hat{\zeta}^n\right)$ satisfies

$$\hat{\xi}_j^n(0)=\hat{Q}_j^n(0)\geq 0,\qquad \text{a.s. for all } j\in[J], \tag{81}$$

$$\sum_{j=1}^J\hat{\xi}_j^n(t)+\sum_{k=1}^K\hat{\zeta}_k^n(t)=0,\qquad \text{a.s. for all } t\geq 0. \tag{82}$$

Similar to the proof of Lemma 2, it is elementary to show that $\hat{I}_j^n$ is nondecreasing with $\hat{I}_j^n(0)=0$ and that

$$\int_0^\infty \mathbb{1}_{\left\{\hat{Q}_j^n(t)>0\right\}}\,d\hat{I}_j^n(t)=0,$$

implying (by the uniqueness in Proposition 1) that

$$\left(\hat{Q}^n,\hat{I}^n,\hat{V}^n\right)=\left(f\left(\hat{\xi}^n,\hat{\zeta}^n\right),g\left(\hat{\xi}^n,\hat{\zeta}^n\right),h\left(\hat{\xi}^n,\hat{\zeta}^n\right)\right). \tag{83}$$

By Lemma 3, Proposition 2, and the continuous mapping theorem,

$$\left(\hat{Q}^n,\hat{I}^n,\hat{V}^n\right)=\left(f\left(\hat{\xi}^n,\hat{\zeta}^n\right),g\left(\hat{\xi}^n,\hat{\zeta}^n\right),h\left(\hat{\xi}^n,\hat{\zeta}^n\right)\right)\Rightarrow\left(f\left(\xi^*,\zeta^*\right),g\left(\xi^*,\zeta^*\right),h\left(\xi^*,\zeta^*\right)\right). \tag{84}$$

Since inequalities are preserved under weak convergence, Equations (46) and (81) imply that

$$\xi_j^*(0) = Q_j(0) \geq 0, \quad \text{a.s. for all } j \in [J]. \tag{85}$$

It also follows from Lemma 3 and Equation (82) that

$$\sum_{j=1}^{J} \xi_j^*(t) + \sum_{k=1}^{K} \zeta_k^*(t) = 0, \quad \text{a.s. for all } t \geq 0. \tag{86}$$

Letting $(Q^*, I^*, V^*) := (f(\xi^*, \zeta^*), g(\xi^*, \zeta^*), h(\xi^*, \zeta^*))$, the proof is complete by Equations (85) – (86) and Proposition 1. (The distributional description of (ξ^*, ζ^*) comes from Lemma 3.) $\square$

6 Conclusion

This paper proposes a closed queueing system to model the movement of cars in a ride-hailing network. Under the assumption that the supply of cars and customer demand is perfectly balanced, our results show that the distribution of cars in the network can be approximated by a diffusion process. Crucially, this paper incorporates travel times into the ridesharing model. Modeling travel times is important because, as is often the case in large cities, drivers spend a non-trivial amount of time on the road delivering customers to their destinations. Ignoring these travel times can lead to inaccuracies when tracking cars in a city.

The results in this paper effectively assume that the cars in the ride-hailing platform are self-driving because we do not model strategic driver behavior. Therefore, a worthwhile extension to this work would be to incorporate strategic behavior into the model because, in many settings, drivers are autonomous and forward looking.

References

Afèche, P., Liu, Z., and C. Maglaras (2018), "Ride-Hailing Networks with Strategic Drivers: The Impact of Platform Control Capabilities on Performance," *Working Paper*.

Afèche, P., Liu, Z., and C. Maglaras (2020), "Surge Pricing an Dynamic Matching for Hotspot Demand Shock in Ride-hailing Networks," *Working Paper*.

Agatz, N., Erera, A., and X. Wang (2017), "Stable Matching for Dynamic Ride-sharing Systems," *Transportation Science*, **52** (4), 850–867.

Ata, B. and S. Kumar (2005), "Heavy Traffic Analysis of Open Processing Systems with Complete Resource Pooling: Asymptotic Optimality of Discrete Review Policies," *The Annals of Applied Probability*, **15** (1A), 331–391.

Ata, B. and W. Lin (2008), "Heavy Traffic Analysis of Maximum Pressure Policies for Stochastic Processing Networks with Multiple Bottlenecks," *Queueing Systems*, **59** (3–4), 191–235.

Ata, B. and T. Olsen (2009), "Near-Optimal Dynamic Lead-Time Quotation and Scheduling Under Convex-Concave Customer Delay Costs," *Operations Research*, **57** (3), 753–768.

Ata, B. and T. Olsen (2013), "Congestion-Based Leadtime Quotation and Pricing for Revenue Maximization with Heterogeneous Customers," *Queueing Systems*, **73** (1), 35–78.

Ata, B., Barjesteh, N., and S. Kumar (2019), "Spatial Pricing: An Empirical Analysis of Taxi Rides in New York City," *Working Paper*.

Ata, B., Barjesteh, N., and S. Kumar (2020), "Dynamic Matching and Centralized Relocation in Ridesharing Platforms," *Working Paper*.

Banerjee, S., Riquelme, and R. Johari (2015), "Pricing in Ride-share Platforms: A Queueing-Theoretic Approach," *Working Paper*.

Banerjee, S., Freund, D., and T. Lykouris (2017), "Pricing and Optimization in Shared Vehicle Systems: An Approximation Framework," *Working Paper*.

Banerjee, S., Kanoria, Y., and Qian, P. (2019), "State Dependent Control of Closed Queueing Networks with Application to Ride-Hailing," *Working Paper*.

Bertsimas, D., Jaillet, P., and S. Martin (2019), "Online Vehicle Routing: The Edge of Optimization in Large-Scale Applications," *Operations Research*, **67** (1), 143–162.

Besbes, O., Castro, F., and I. Lobel (2019), "Surge Pricing and Its Spatial Supply Response," *Working Paper*.

Billingsley, P. (1999), "Convergence of Probability Measures (Second Edition)," John Wiley & Sons, Inc., New York.

Bimpikis, K., Candogan, O., and D. Saban (2019) "Spatial Pricing in Ride-Sharing Networks," *Operations Research*, **67** (3), 744–769.

Braverman, A., Dai, J.G., Liu, X., and L. Ying (2019), "Empty-Car Routing in Ridesharing Systems," *Operations Research*, **67** (5), 1437–1452.

Cachon, G., Daniels, K., and R. Lobel (2017), "The Role of Surge Pricing on a Service Platform with Self-Scheduling Capacity," *Manufacturing & Service Operations Management*, **19** (3), 337–507.

Castillo, J.C., Knoepfle, D., and G. Weyl (2018), "Surge Pricing Solves the Wild Goose Chase," *Working Paper*.

Chen, M.K. and M. Sheldon (2016), "Dynamic Pricing in a Labor Market: Surge Pricing and Flexible Work on the Uber Platform," *Proceedings of the 2016 ACM Conference on Economics and Computation*.

Ethier, S. and T. Kurtz (2005), "Markov Processes: Characterization and Convergences," John Wiley & Sons, Inc., New York.

Garg, N. and H. Nazerzadeh (2019), "Driver Surge Pricing," *Working Paper*.

P.W. Glynn (1990), "Chapter 4: Diffusion Approximations" in *Handbooks on OR & MS, Vol. 2, Stochastic Models*, 145–198, North-Holland, Amsterdam.

Gokpinar, B. and C. Selcuk (2019), "The Selection of Prices and Commissions in a Spatial Model of Ride-hailing," *Working Paper*.

Guda, H. and U. Subramanian (2019), "Your Uber is Arriving: Managing On-Demand Workers Through Surge Pricing, Forecast Communication, and Worker Incentives," *Management Science*, **65** (5), 1995–2014.

J.M. Harrison (2003), "A Broader View of Brownian Networks," *The Annals of Applied Probability*, **13** (3), 1119–1150.

He, L., Hu, Z., and M. Zhang (2019), "Robust Repositioning for Vehicle Sharing," *Manufacturing & Service Operations Management*, Forthcoming.

Hosseini, M., Milner, J., and G. Romero (2020), "Efficient Strategic-Level Repositioning in Vehicle-Sharing Networks," *Working Paper*.

Hu, B., Hu, M., and H. Zhu (2019), "Surge Pricing and Two-Sided Temporal Responses in Ride-hailing," *Working Paper*.

Hu, M. and Y. Zhou (2019), "Price, Wage and Fixed Commission in On-Demand Matching," *Working Paper*.

Kanoria, Y. and P. Qian (2019), "Near Optimal Control of a Ride-hailing Platform via Mirror Backpressure," *Working Paper*.

Kogan, Y. and R. Lipster (1993), "Limit Non-Stationary Behavior of Large Closed Queueing Networks with Bottlenecks," *Queueing Systems*, **14** (1–2), 33–55.

Kogan, Y., Liptser, R., and A.V. Smorodinskii (1986), "Gaussian Diffusion Approximation of Closed Markov Models of Computer Networks," *Problems of Information Transmission*, **22** (1), 38–51.

Korolko, N., Woodard, D., Yan, C., and H. Zhu (2019), "Dynamic Pricing and Matching in Ride-hailing Platforms," *Naval Research Logistics*, Forthcoming.

Krichagina, A.A. and E.V. Puhalskii (1997), "A Heavy-Traffic Analysis of a Closed Queueing System with a GI/∞ Service Center," *Queueing Systems*, **25** (1–4), 235–280.

R. Lagos (2000), "An Alternative Approach to Search Frictions," *Journal of Political Economy*, **108** (5), 851–873.

Lam, C.T. and M. Liu (2017), "Demand and Consumer Surplus in the On-Demand Economy: The Case of Ride Sharing," *Working Paper*.

Lu, A., Frazier, P., and O. Kislev (2018), "Surge Pricing Moves Uber's Driver Partners," *Proceedings of the 2018 ACM Conference on Economics and Computation*.

E. Özkan (2018), "Joint Pricing and Matching in Ridesharing Systems," *Working Paper.*

Özkan, E. and A.R. Ward (2019), "Dynamic Matching for Real-time Ridesharing," *Stochastic Systems*, Forthcoming.

Pang, G., Talreja, R., and W. Whitt (2007), "Martingale Proofs of Many-Server Heavy-Traffic Limits for Markovian Queues," *Probability Surveys*, **4**, 193–267.

W.R. Pestman (1995), "Measurability of Linear Operators in the Skorokhod Topology," *Bull. Belg. Math. Soc. Simon Stevin*, **2** (4), 381–388.

Reed, J. and A.R. Ward (2004), "A Diffusion Approximation for a Generalized Jackson Network with Reneging," *Proceedings of the 42nd Annual Conference on Communication, Control, and Computing.*

Reed, J. and A.R. Ward (2008), "Approximating the GI/GI/1+GI Queue with a Nonlinear Drift Diffusion: Hazard Rate Scaling in Heavy Traffic," *Mathematics of Operations Research*, **33** (3), 606–644.

A.V. Smorodinskii (1986), "Asymptotic Distribution of the Queue Length of One Service System" (in Russian), *Avtomatika i Telemekhanika*, **2**, 92-99.

Whitt, W. (2002), "Stochastic-Process Limits," Springer-Verlag, New York.

A Notation and Technical Preliminaries

For a function $f : X \to Y$ and a subset $S \subseteq X$, we denote by $f|_S$ the restriction of f to S. The indicator function for a subset $S \subseteq X$ is written as $\mathbb{1}_S$. The set of positive integers is denoted by $\mathbb{N} = \{1, 2, 3, \dots\}$ and we write $[n] = \{1, 2, \dots, n\}$ for $n \in \mathbb{N}$. For $k \in [n]$, the kth unit basis vector in $\mathbb{R}^n$ is denoted by e_k, which has one in the kth component and zeros elsewhere. For $a, b \in \mathbb{R}$, we let $a \vee b = \max\{a, b\}$ and $a \wedge b = \min\{a, b\}$, and let $\lfloor a \rfloor$ denote the largest integer less than or equal to a. For $l = 1, \dots, k$, the lth projection map $\pi_l : \mathbb{R}^k \to \mathbb{R}$ is given by $\pi_l(x) = x_l$, where x_l is the lth component of $x \in \mathbb{R}^k$. The abbreviation a.s. stands for "almost surely" and the notation $\overset{p}{\to}$ means "converges in probability."

For each positive integer $k \geq 1$, we denote by $D^k \equiv D\big([0, \infty), \mathbb{R}^k\big)$ the set of all functions $x : [0, \infty) \to \mathbb{R}^k$ that are right continuous on $[0, \infty)$ and have left limits on $(0, \infty)$. The identically zero function in D^k will be denoted by $\mathbf{0}$. Similarly, for each positive integer $k \geq 1$ and positive real number $T > 0$, we denote by $D_T^k \equiv D\big([0, T], \mathbb{R}^k\big)$ the set of all functions $x : [0, T] \to \mathbb{R}^k$ that are right continuous on $[0, T)$ and have left limits on $(0, T]$. When the space D_T^k is endowed with the norm

$$\|x\|_{T,k} := \max_{1 \leq l \leq k} \sup_{0 \leq t \leq T} |x_l(t)|, \tag{87}$$

it is a Banach space; for example, see Pestman (1995). When $k = 1$, we write $D^1 = D$, $D_T^1 = D_T$, and $\|\cdot\|_{T,1} = \|\cdot\|_T$. The one-sided reflection map on D is given by the pair of functions $(\psi, \phi) : D \to D^2$ defined as follows:

$$\psi(x)(t) := \sup_{0 \leq s \leq t} [-x(s)]^+, \tag{88}$$

$$\phi(x)(t) := x(t) + \psi(x)(t). \tag{89}$$

For $x, y \in D$ and $T > 0$ [see Lemma 13.5.1 in Whitt (2002)], the following inequalities holds:

$$\|\psi(x) - \psi(y)\|_T \leq \|x - y\|_T, \tag{90}$$

$$\|\phi(x) - \phi(y)\|_T \leq 2\|x - y\|_T. \tag{91}$$

We regard D^k as a topological space with the Skorokhod J_1 topology. It is indeed a Polish space [see Billingsley (1999)], however, for our purposes, the reader only needs to know what it means for a sequence of functions to converge in D^k. No further properties of the underlying topology are used. To that end, a sequence $\{x^n\}_{n=1}^{\infty}$ in D^k converges to an element $x \in D^k$, written $x^n \to x$, as $n \to \infty$, if for all continuity points $T > 0$ of x,

$$d_T^k \left(x^n|_{[0,T]}, x|_{[0,T]} \right) \to 0 \quad \text{as} \quad n \to \infty, \tag{92}$$

where $d_T^k : D_T^k \times D_T^k \to [0, \infty)$ is given by

$$d_T^k (x, y) := \inf_{\lambda \in \Lambda_T} \left\{ \|x \circ \lambda - y\|_{T,k} \vee \|\lambda - e\|_T \right\}, \tag{93}$$

where $e : [0, T] \to [0, T]$ is the identity map, i.e., $e(t) = t$, and

$$\Lambda_T = \{\lambda : [0, T] \to [0, T] \,|\, \lambda \text{ is an increasing homeomorphism}\}. \tag{94}$$

Equivalently, Equation (92) can be written with $\widetilde{d}_T^k$ in place of d_T^k [for example, see page 226 of Pang et al. (2007) and Billingsley (1999)], where $\widetilde{d}_T^k : D_T^k \times D_T^k \to [0, \infty)$ is given by

$$\widetilde{d}_T^k (x, y) := \inf_{\lambda \in \widetilde{\Lambda}_T} \left\{ \|x \circ \lambda - y\|_{T,k} \vee \|\dot{\lambda} - 1\|_T \right\}, \tag{95}$$

where $\dot{\lambda}$ is the derivative of λ, 1 is the constant function taking the value one everywhere, and

$$\widetilde{\Lambda}_T = \{\lambda : [0, T] \to [0, T] \,|\, \lambda \in \Lambda_T \text{ and is absolutely continuous w.r.t. Lebesgue measure}\}. \tag{96}$$

As before, when $k = 1$, we write $d_T^1 = d_T$ and $\widetilde{d}_T^1 = \widetilde{d}_T$.

All random variables in this paper are assumed to live on a common probability space $(\Omega, \mathcal{F}, P)$. We denote by $\mathcal{M}^k$ the Borel σ-algebra on D^k induced by the Skorokhod J_1 topology, so that $(D^k, \mathcal{M}^k)$ forms a measurable space. Each stochastic process in this paper is assumed to be a measurable function from $(\Omega, \mathcal{F}, P)$ to $(D^k, \mathcal{M}^k)$, with appropriate dimension k. For a sequence of stochastic processes $\{\xi^n\}_{n=1}^{\infty}$ in D^k, where $\xi^n = \{\xi^n(t) : t \geq 0\}$, we write

$$\xi^n \Rightarrow \xi \quad \text{as} \quad n \to \infty \tag{97}$$

to mean that the sequence of probability measures on $(D^k, \mathcal{M}^k)$ induced by the ξ^n converge weakly to the probability measure on $(D^k, \mathcal{M}^k)$ induced by the stochastic process ξ; see Billingsley (1999) and Whitt (2002) for further details.

B Proofs for Results in Section 4

The following lemma is useful in the proof of Proposition 1; its proof can be found in Appendix C. To state it, let $(\xi,\zeta) \in D^{J+K}$ and for $k = 1,\ldots,K$ and $t \geq 0$ consider the following equations:

$$y_k(t) = \zeta_k(t) - \eta_k \int_0^t y_k(s)\,ds - \sum_{j=1}^{J} p_{jk}\psi\left(\xi_j + \sum_{l=1}^{K} q_{lj}\eta_l \int_0^{\cdot} y_l(s)\,ds\right)(t). \tag{98}$$

Lemma 4. *For each $(\xi,\zeta) \in D^{J+K}$, there exists a unique $y \in D^K$ satisfying Equation (98).*

Below is a proof of Proposition 1:

Proof of Proposition 1. Fix $(\xi,\zeta) \in D^{J+K}$ satisfying Equations (56) – (57). We first prove existence. By Lemma 4, there exists a $y \in D^K$ satisfying Equation (98). For $j = 1,\ldots,J$, define

$$u_j := \mu_j^{-1}\psi\left(\xi_j + \sum_{l=1}^{K} q_{lj}\eta_l \int_0^{\cdot} y_l(s)\,ds\right), \tag{99}$$

$$x_j := \phi\left(\xi_j + \sum_{l=1}^{K} q_{lj}\eta_l \int_0^{\cdot} y_l(s)\,ds\right). \tag{100}$$

Since $y \in D^K$, it follows that $u \in D^J$ and $x \in D^J$, so that $(x,u,y) \in D^{2J+K}$. It remains to show that (x,u,y) satisfy Equations (58) – (62). Equation (58) follows from the definitions of u and x in Equations (99) and (100), respectively, as well as definition of the one-sided reflection map (ψ,ϕ) in Equations (88) – (89). Equation (59) holds by the fact that y satisfies Equation (98) and the definition of u in Equation (99). Equation (60) follows from Equations (56), (58) – (59), and the fact that

$$\sum_{k=1}^{K} p_{jk} = 1 \quad \text{for all} \quad j = 1,\ldots,J,$$
$$\sum_{j=1}^{J} q_{kj} = 1 \quad \text{for all} \quad k = 1,\ldots,K.$$

Equation (61) follows from the definition of u in Equation (99), the definition of ψ in Equation (88), and Equation (57). Finally, if we let

$$z_j := \xi_j + \sum_{l=1}^{K} q_{lj}\eta_l \int_0^{\cdot} y_l(s)\,ds, \tag{101}$$

we see that $u_j = \mu_j^{-1}\psi(z_j)$ and $x_j = \phi(z_j)$. Therefore, Equation (62) says that

$$\int_0^{\infty} \mathbb{1}_{\{\phi(z_j)>0\}}\, d\left(\mu_j^{-1}\psi(z_j)\right)(t) = 0, \tag{102}$$

which is equivalent to

$$\int_0^\infty \mathbb{1}_{\{\phi(z_j)>0\}}\, d\psi(z_j)(t) = 0. \tag{103}$$

But Equation (103) holds true by the definition of (ψ,ϕ) in Equations (88) – (89). Thus, Equation (62) holds.

We now prove uniqueness. Let $(x,u,y),(\widetilde{x},\widetilde{u},\widetilde{y}) \in D^{2J+K}$ both satisfy Equations (58) – (62). By Equation (58),

$$x_j = z_j + \mu_j u_j \geq 0, \tag{104}$$

where z_j is given by Equation (101). Similarly, we can write

$$\widetilde{x}_j = \widetilde{z}_j + \mu_j \widetilde{u}_j \geq 0, \tag{105}$$

where $\widetilde{z}_j$ is given by Equation (101) (where we replace y_l by $\widetilde{y}_l$). Since (x,u) and $(\widetilde{x},\widetilde{u})$ satisfy Equations (61) – (62), it follows that $(x,\mu_j u_j)$ and $(\widetilde{x},\mu_j\widetilde{u}_j)$ satisfy Equations (61) – (62). By this and Equations (104) – (105), it follows that

$$\mu_j u_j = \psi(z_j), \tag{106}$$

$$\mu_j \widetilde{u}_j = \psi\left(\widetilde{z}_j\right). \tag{107}$$

It then follows by Equations (59), (101), (106) – (107), and Lemma 4 that $y_k = \widetilde{y}_k$ for all $k = 1,\ldots,K$. By uniqueness of y, it then follows by Equations (101) and (106) – (107) that $u_j = \widetilde{u}_j$ for all $j = 1,\ldots,J$. Finally, by uniqueness of y and u, it follows by Equations (101) and (104) – (105) that $x_j = \widetilde{x}_j$ for all $j = 1,\ldots,J$. This completes the proof. □

Before a proof of Proposition 2, we provide a description of the mappings f, g, and h in the statement of Corollary 1. First, by Lemma 4, we can indirectly write $h : D^{J+K} \to D^K$ as the unique mapping sending $(\xi,\zeta) \in D^{J+K}$ to $h(\xi,\zeta) \in D^K$ satisfying Equation (98). For our purposes, this indirect description of h will be enough. On the other hand, the proof of Proposition 1 shows that the mappings $f : D^{J+K} \to D^J$ and $g : D^{J+K} \to D^J$ are uniquely given by the following:

$$f := (\xi,\zeta) \mapsto \left(\phi\left(\pi_j \circ \xi + \sum_{l=1}^{K} q_{lj}\eta_l \int_0^{\cdot} \left(\pi_l \circ h(\xi,\zeta)\right)(s)\, ds\right)\right)_{j=1,\ldots,J}, \tag{108}$$

$$g := (\xi, \zeta) \mapsto \left(\mu_j^{-1} \psi \left(\pi_j \circ \xi + \sum_{l=1}^{K} q_{lj} \eta_l \int_0^{\cdot} \left(\pi_l \circ h(\xi, \zeta) \right)(s)\, ds \right) \right)_{j=1,\ldots,J}. \tag{109}$$

Proof of Proposition 2. We first prove continuity of f, g, and h separately. Then we use these results to argue that the joint mapping (f, g, h) is continuous.

Continuity of h. Recall that for each $(\xi, \zeta) \in D^{J+K}$ the element $y = h(\xi, \zeta) \in D^K$ satisfies

$$y_k(t) = \zeta_k(t) - \eta_k \int_0^t y_k(s)\, ds - \sum_{j=1}^{J} p_{jk} \psi \left(\xi_j + \sum_{l=1}^{K} q_{lj} \eta_l \int_0^{\cdot} y_l(s)\, ds \right)(t) \tag{110}$$

for all $k = 1, \ldots, K$ and $t \geq 0$. Suppose that $(\xi^n, \zeta^n) \to (\xi, \zeta)$ in D^{J+K} as $n \to \infty$ and let $\widetilde{T} > 0$ be a continuity point of $f(\xi, \zeta)$. To complete the proof, we must show that

$$d_{\widetilde{T}}^K \left(h(\xi^n, \zeta^n)|_{[0,\widetilde{T}]}, h(\xi, \zeta)|_{[0,\widetilde{T}]} \right) \to 0 \quad \text{as } n \to \infty, \tag{111}$$

where $d_{\widetilde{T}}^K$ is given by Equation (93). However, by Lemma 5.1 in Ethier and Kurtz (2005), since $(\xi, \zeta) \in D^{J+K}$, it has at most countably many points of discontinuity. As a result, there exists a point $T > \widetilde{T}$ that is a continuity point of (ξ, ζ). By Lemma 1 on page 167 in Billingsley (1999), it suffices to show that

$$d_T^K \left(h(\xi^n, \zeta^n)|_{[0,T]}, h(\xi, \zeta)|_{[0,T]} \right) \to 0 \quad \text{as } n \to \infty, \tag{112}$$

for then Equation (112) would imply Equation (111). To avoid cumbersome notation, we write

$$d_T^K \left(h(\xi^n, \zeta^n), h(\xi, \zeta) \right) \to 0 \quad \text{as } n \to \infty \tag{113}$$

to mean Equation (112). The remainder of the proof aims at proving Equation (113). However, since the metrics d_T^K and $\widetilde{d}_T^K$ are topologically equivalent [see page 22 and Equation (95)], it suffices to prove Equation (113) with $\widetilde{d}_T^K$ in place of d_T^K.

Since $T > 0$ is a continuity point of (ξ, ζ) and $(\xi^n, \zeta^n) \to (\xi, \zeta)$ in D^{J+K}, there exists a sequence of homeomorphisms $\lambda^n \in \widetilde{\Lambda}_T$ such that

$$\| (\xi, \zeta) \circ \lambda^n - (\xi^n, \zeta^n) \|_{T, J+K} \vee \| \dot{\lambda}^n - 1 \|_T \to 0 \quad \text{as } n \to \infty. \tag{114}$$

Letting $y = h(\xi, \zeta)$ and $y^n = h(\xi^n, \zeta^n)$, for any $0 \leq t \leq T$ we have:

$$\max_{1 \leq k \leq K} \| y_k \circ \lambda^n - y_k^n \|_t = \max_{1 \leq k \leq K} \Bigg[\Bigg\| \Bigg(\zeta_k \circ \lambda^n - \eta_k \int_0^{\lambda^n(\cdot)} y_k(s)\, ds$$

$$
\begin{aligned}
&- \sum_{j=1}^{J} p_{jk} \psi \left(\xi_j + \sum_{l=1}^{K} q_{lj} \eta_j \int_0^{\cdot} y_l(s)\, ds \right) \circ \lambda^n \Bigg) \\
&\quad - \left(\zeta_k^n - \eta_k \int_0^{\cdot} y_k^n(s)\, ds - \sum_{j=1}^{J} p_{jk} \psi \left(\xi_j^n + \sum_{l=1}^{K} q_{lj} \eta_l \int_0^{\cdot} y_l^n(s)\, ds \right) \right) \Big\|_t \Bigg] \\
&\le \max_{1 \le k \le K} \Bigg[\| \zeta_k \circ \lambda^n - \zeta_k^n \|_t + \eta \Big\| \int_0^{\lambda^n(\cdot)} y_k(s)\, ds - \int_0^{\cdot} y_k^n(s)\, ds \Big\|_t \\
&\quad + \sum_{j=1}^{J} \Big\| \psi \left(\xi_j + \sum_{l=1}^{K} q_{lj} \eta_j \int_0^{\cdot} y_l(s)\, ds \right) \circ \lambda^n \\
&\qquad - \psi \left(\xi_j^n + \sum_{l=1}^{K} q_{lj} \eta_l \int_0^{\cdot} y_l^n(s)\, ds \right) \Big\|_t \Bigg],
\end{aligned} \tag{115}
$$

where $\eta := \max_{1 \le k \le K} \eta_k$. We now bound each of the terms in Equation (115). To that end, first let

$$
M_T := \max_{1 \le k \le K} \| y_k \|_T < \infty. \tag{116}
$$

Then by Equation (116) and the chain rule,

$$
\begin{aligned}
\Big\| \int_0^{\lambda^n(\cdot)} y_k(s)\, ds - \int_0^{\cdot} y_k^n(s)\, ds \Big\|_t &= \Big\| \int_0^{\cdot} y_k(\lambda^n(s)) \dot{\lambda}^n(s)\, ds - \int_0^{\cdot} y_k^n(s)\, ds \Big\|_t \\
&\le \Big\| \int_0^{\cdot} y_k(\lambda^n(s)) \left(\dot{\lambda}^n(s) - 1 \right) ds \Big\|_t \\
&\quad + \Big\| \int_0^{\cdot} \left(y_k(\lambda^n(s)) - y_k^n(s) \right) ds \Big\|_t \\
&\le T M_T \| \dot{\lambda} - 1 \|_T + \int_0^t \| y_k \circ \lambda^n - y_k^n \|_s \, ds.
\end{aligned} \tag{117}
$$

Similar to Equation (117), we have

$$
\begin{aligned}
&\Big\| \psi \left(\xi_j + \sum_{l=1}^{K} q_{lj} \eta_j \int_0^{\cdot} y_l(s)\, ds \right) \circ \lambda^n - \psi \left(\xi_j^n + \sum_{l=1}^{K} q_{lj} \eta_l \int_0^{\cdot} y_l^n(s)\, ds \right) \Big\|_t \\
&\le \| \xi_j \circ \lambda^n - \xi_j^n \|_T + \eta \sum_{l=1}^{K} \Big\| \int_0^{\lambda^n(\cdot)} y_l(s)\, ds - \int_0^{\cdot} y_l^n(s)\, ds \Big\|_t \\
&\le \| \xi_j \circ \lambda^n - \xi_j^n \|_T + \eta \sum_{l=1}^{K} \Big\| \int_0^{\cdot} y_l\left(\lambda^n(s)\right) \dot{\lambda}^n(s)\, ds - \int_0^{\cdot} y_l^n(s)\, ds \Big\|_t \\
&\le \| \xi_j \circ \lambda^n - \xi_j^n \|_T + \eta K T M_T \| \dot{\lambda}^n - 1 \|_T + \eta \sum_{l=1}^{K} \int_0^t \| y_l \circ \lambda^n - y_l^n \|_s \, ds,
\end{aligned} \tag{118}
$$

where the first inequality holds by Equation (90) and the fact that [see Lemma 13.5.2 in Whitt (2002)]

$$
\psi \left(\xi_j + \sum_{l=1}^{K} q_{lj} \eta_j \int_0^{\cdot} y_l(s)\, ds \right) \circ \lambda^n = \psi \left(\xi_j \circ \lambda^n + \sum_{l=1}^{K} q_{lj} \eta_j \int_0^{\lambda^n(\cdot)} y_l(s)\, ds \right). \tag{119}
$$

By Equations (115) and (117) – (118), it follows that for all $0 \le t \le T$,

$$\begin{aligned}
\max_{1\le k\le K} \|y_k \circ \lambda^n - y_k^n\|_t \le \max_{1\le k\le K} \Bigg[& \|\zeta_k \circ \lambda^n - \zeta_k^n\|_T + \eta T M_T \|\dot{\lambda} - 1\|_T + \eta \int_0^t \|y_k \circ \lambda^n - y_k^n\|_s \, ds \\
& + \sum_{j=1}^{J} \left(\|\xi_j \circ \lambda^n - \xi_j^n\|_T + \eta K T M_T \|\dot{\lambda}^n - 1\|_T + \eta \sum_{l=1}^{K} \int_0^t \|y_l \circ \lambda^n - y_l^n\|_s \, ds \right) \Bigg] \\
\le \max_{1\le k\le K} & \|\zeta_k \circ \lambda^n - \zeta_k^n\|_T + \eta T M_T \|\dot{\lambda} - 1\|_T + \eta \int_0^t \max_{1\le k\le K} \|y_k \circ \lambda^n - y_k^n\|_s \, ds \\
& + J \max_{1\le j\le J} \|\xi_j \circ \lambda^n - \xi_j^n\|_T + \eta J K T M_T \|\dot{\lambda}^n - 1\|_T \\
& + \eta J K \int_0^t \max_{1\le k\le K} \|y_k \circ \lambda^n - y_k^n\|_s \, ds \\
\le 2J & \| (\xi, \zeta) \circ \lambda^n - (\xi^n, \zeta^n) \|_{D_T^{J+K}} + (JK+1)\eta T M_T \|\dot{\lambda}^n - 1\|_T \\
& + 2\eta J K \int_0^t \max_{1\le k\le K} \|y_k \circ \lambda^n - y_k^n\|_s \, ds.
\end{aligned} \tag{120}$$

For $\epsilon > 0$ fixed, let n_0 be large enough such that for all $n \ge n_0$,

$$2J \| (\xi, \zeta) \circ \lambda^n - (\xi^n, \zeta^n) \|_{T, J+K} < \frac{\epsilon}{2e^{2\eta J K T}} \tag{121}$$

and that

$$(JK+1)\eta T M_T \|\dot{\lambda}^n - 1\|_T < \frac{\epsilon}{2e^{2\eta J K T}}. \tag{122}$$

Then by Equations (120) – (122), it follows that for all $n \ge n_0$,

$$\max_{1\le k\le K} \|y_k \circ \lambda^n - y_k^n\|_t \le \frac{\epsilon}{e^{2\eta J K T}} + 2\eta J K \int_0^t \max_{1\le k\le K} \|y_k \circ \lambda^n - y_k^n\|_s \, ds \quad \text{for all } 0 \le t \le T. \tag{123}$$

By Gronwall's inequality [see Lemma 4.1 in Pang et al. (2007)] and Equation (123), it follows that

$$\max_{1\le k\le K} \|y_k \circ \lambda^n - y_k^n\|_t \le \frac{\epsilon}{e^{2\eta J K T}} e^{2\eta J K t} \quad \text{for all } 0 \le t \le T. \tag{124}$$

In particular, using Equation (124) with $t = T$, it follows that

$$\|y \circ \lambda^n - y^n\|_{T,K} \equiv \max_{1\le k\le K} \|y_k \circ \lambda^n - y_k^n\|_T \le \epsilon. \tag{125}$$

Finally, by Equation (114) let n_1 be large enough such that for all $n \ge n_1$,

$$\|\dot{\lambda}^n - 1\|_T \le \epsilon. \tag{126}$$

Therefore, by Equations (125) – (126), for all $n \geq \max\{n_0, n_1\}$ we have

$$\|y \circ \lambda^n - y^n\|_{T,K} \vee \|\dot{\lambda}^n - 1\|_T \leq \epsilon, \tag{127}$$

completing the proof.

Continuity of g. The continuity proof for g [see Equation (109) for its expression] proceeds in the some way as in the continuity proof for f. Suppose $(\xi^n, \zeta^n) \to (\xi, \zeta)$ in D^{J+K} as $n \to \infty$ and let $T > 0$ be a continuity point of (ξ, ζ). Then there exists a sequence of homeomorphisms $\lambda^n \in \widetilde{\Lambda}_T$ such that Equation (114) holds. Letting $u = g(\xi, \zeta)$, $u^n = g(\xi^n, \zeta^n)$, $y = h(\xi, \zeta))$ and $y^n = h(\xi^n, \zeta^n)$, we have:

$$\begin{aligned}
\max_{1\leq j\leq J} \|u_j \circ \lambda^n - u_j^n\|_T &= \max_{1\leq j\leq J} \left\| \mu_j^{-1} \psi\left(\xi_j + \sum_{l=1}^K q_{lj}\eta_l \int_0^{\cdot} y_l(s)\,ds\right) \circ \lambda^n - \mu_j^{-1}\psi\left(\xi_j^n + \sum_{l=1}^K q_{lj}\eta_l \int_0^{\cdot} y_l^n(s)\,ds\right) \right\|_T \\
&\leq \mu^{-1} \max_{1\leq j\leq J} \left\| \psi\left(\xi_j \circ \lambda^n + \sum_{l=1}^K q_{lj}\eta_l \int_0^{\lambda^n(\cdot)} y_l(s)\,ds\right) - \psi\left(\xi_j^n + \sum_{l=1}^K q_{lj}\eta_l \int_0^{\cdot} y_l^n(s)\,ds\right) \right\|_T \\
&\leq \mu^{-1} \max_{1\leq j\leq J} \|\xi_j \circ \lambda^n - \xi_j^n\|_T + \mu^{-1}\eta K T M_T \|\dot{\lambda}^n - 1\|_T \\
&\qquad + \mu^{-1}\eta K T \max_{1\leq k\leq K} \|y_k \circ \lambda^n - y_k^n\|_T \\
&\leq \mu^{-1} \|(\xi,\zeta) \circ \lambda^n - (\xi^n, \zeta^n)\|_{T,J+K} + \mu^{-1}\eta K T M_T \|\dot{\lambda}^n - 1\|_T \\
&\qquad + \mu^{-1}\eta K T \max_{1\leq k\leq K} \|y_k \circ \lambda^n - y_k^n\|_T,
\end{aligned} \tag{128}$$

where $\mu := \min_{1\leq j\leq J} \mu_j$ and where the second inequality follows from Equation (90). Since $(\xi^n, \zeta^n) \to (\xi, \zeta)$ in D^{J+K}, for $\epsilon > 0$ fixed there exists an n_0 such that for all $n \geq n_0$,

$$\mu^{-1} \|(\xi,\zeta) \circ \lambda^n - (\xi^n, \zeta^n)\|_{T,J+K} \leq \frac{\epsilon}{3} \tag{129}$$

and that

$$\mu^{-1}\eta K T M_T \|\dot{\lambda}^n - 1\|_T \leq \frac{\epsilon}{3}. \tag{130}$$

Furthermore, by Equation (125), there exists an n_1 such that for all $n \geq n_1$,

$$\mu^{-1}\eta K T \max_{1\leq k\leq K} \|y_k \circ \lambda^n - y_k^n\|_T \leq \frac{\epsilon}{3}. \tag{131}$$

Finally, let n_2 be large enough such that Equation (126) holds for all $n \geq n_2$. Then by Equations

(128) − (131), it follows that for all $n \geq \max\{n_0, n_1, n_2\}$,

$$\|u \circ \lambda^n - u^n\|_{T,J} \vee \|\dot{\lambda}^n - 1\|_T \leq \epsilon, \tag{132}$$

which is the desired result.

Continuity of f. The continuity proof for f is nearly identical to the continuity proof for g [see Equation (108) for its expression]. Suppose $(\xi^n, \zeta^n) \to (\xi, \zeta)$ in D^{J+K} as $n \to \infty$ and let $T > 0$ be a continuity point of (ξ, ζ). Then there exists a sequence of homeomorphisms $\lambda^n \in \widetilde{\Lambda}_T$ such that Equation (114) holds. Letting $x = f(\xi, \zeta)$, $x^n = f(\xi^n, \zeta^n)$, $y = h(\xi, \zeta))$ and $y^n = h(\xi^n, \zeta^n)$, we have:

$$\begin{aligned}
\max_{1\leq j\leq J} \|x_j \circ \lambda^n - x_j^n\|_T &= \max_{1\leq j\leq J} \left\| \phi\left(\xi_j \circ \lambda^n + \sum_{l=1}^{K} q_{lj}\eta_l \int_0^{\lambda^n(\cdot)} y_l(s)\,ds\right) - \phi\left(\xi_j^n + \sum_{l=1}^{K} q_{lj}\eta_l \int_0^{\cdot} y_l^n(s)\,ds\right)\right\|_T \\
&\leq 2\max_{1\leq j\leq J} \|\xi_j \circ \lambda^n - \xi_j^n\|_T + 2\eta \sum_{l=1}^{K} \left\| \int_0^{\cdot} y_l\left(\lambda^n(s)\right)\dot{\lambda}^n(s)\,ds - \int_0^{\cdot} y_l^n(s)\,ds \right\|_T \\
&\leq 2\|(\xi,\zeta)\circ\lambda^n - (\xi^n,\zeta^n)\|_{T,J+K} + 2\eta K T M_T \|\dot{\lambda}^n - 1\|_T \\
&\qquad + 2\eta K T \max_{1\leq k\leq K} \|y_k \circ \lambda^n - y_k^n\|_T,
\end{aligned} \tag{133}$$

where the first inequality follows from Equation (91). We see that the right hand side Equation (133) is the same as the right hand side Equation (128), except with 2 instead of μ^{-1}. Thus, the same final arguments used in the continuity proof of g can be used to complete the continuity proof for h.

Continuity of (f, g, h). We regard (f, g, h) as a function from $D^{J+K} \to D^{2J+K}$ defined by $(\xi, \zeta) \mapsto (f(\xi,\zeta), g(\xi,\zeta), h(\xi,\zeta))$. Suppose that $(\xi^n, \zeta^n) \to (\xi, \zeta)$ in D^{J+K} and let $T > 0$ be a continuity point of (ξ, ζ). Therefore, there exists a sequence of homeomorphisms $\lambda^n \in \widetilde{\Lambda}_T$ such that Equation (114) holds. For notational purposes, let $F^n = (f(\xi^n,\zeta^n), g(\xi^n,\zeta^n), h(\xi^n,\zeta^n))$, $F = (f(\xi,\zeta), g(\xi,\zeta), h(\xi,\zeta))$, $x^n = f(\xi^n,\zeta^n)$, $x = f(\xi,\zeta)$, $u^n = g(\xi^n,\zeta^n)$, $u = g(\xi,\zeta)$, $y^n = h(\xi^n,\zeta^n)$, and $y = h(\xi,\zeta)$. Using this notation, to prove continuity of (f, g, h) we it suffices to show that

$$\|F \circ \lambda^n - F^n\|_{T,2J+K} \vee \|\dot{\lambda}^n - 1\| \to 0 \quad \text{as} \quad n \to \infty. \tag{134}$$

But note that

$$\|F \circ \lambda^n - F^n\|_{T,2J+K} \equiv \max_{1\leq j\leq J} \|x_j \circ \lambda^n - x_j^n\|_T \vee \max_{1\leq j\leq J} \|u_j \circ \lambda^n - u_j^n\|_T \vee \max_{1\leq k\leq K} \|y_k \circ \lambda^n - y_k^n\|_T. \tag{135}$$

Since each of the three terms on the right hand side of Equation (135) converge to zero (by continuity of f, g, and h separately), Equation (134) will follow. The proof is complete. □

C Proof of Lemma 4

The general proof technique we use parallels that of Lemma 1 in Reed and Ward (2004). We prove that for each $T > 0$, there exists a unique $y \in D_T^K$ satisfying Equation (98), then extend this to D^K in the obvious way. To improve the readability of this proof, we break it up into a few separate steps.

C.1 Existence of an Element in D_K^T Satisfying Equation (98)

We prove existence via the method of successive approximations. In particular, we show that the sequence formed by this method is Cauchy in D_K^T, and then argue that the limit of the sequence (by completeness of D_K^T) satisfies Equation (98). To that end, let $y_k^0 \equiv 0 \in D$ and let $y_k^n \in D$, $n \geq 1$, be defined by

$$y_k^n := \xi_k - \eta_k \int_0^{\cdot} y_k^{n-1}(s)\, ds - \sum_{j=1}^{J} p_{jk} \psi \left(\xi_j + \sum_{l=1}^{K} q_{lj} \eta_l \int_0^{\cdot} y_l^{n-1}\, ds \right), \tag{136}$$

for each $k = 1, \ldots, K$. We claim that the sequence $\left\{ \left(y_1^n|_{[0,T]}, \ldots, y_K^n|_{[0,T]} \right) : n = 0, 1, \ldots \right\}$ defined by Equation (136) is Cauchy in D_T^K; see Claim 1 at the end of Appendix C for a proof of this claim. By completeness of $(D_T^K, \|\cdot\|_{T,K})$, it follows that

$$\left(y_1^n|_{[0,T]}, \ldots, y_K^n|_{[0,T]} \right) \to \left(y_{1,T}^{\infty}, \ldots, y_{K,T}^{\infty} \right) \in D_T^K \quad \text{as} \quad n \to \infty. \tag{137}$$

To show that $(y_{1,T}^{\infty}, \ldots, y_{K,T}^{\infty})$ satisfies (98), define the mapping $L : D_T^K \to D_T^K$ by

$$(y_1, \ldots, y_K) \mapsto \left(\zeta_k - \eta_k \int_0^{\cdot} y_k(s)\, ds - \sum_{j=1}^{J} p_{jk} \psi \left(\xi_j + \sum_{l=1}^{K} q_{lj} \eta_l \int_0^{\cdot} y_l(s)\, ds \right) \right)_{k=1,\ldots,K}. \tag{138}$$

Then for $y, \widetilde{y} \in D_T^K$ we have:

$$\begin{aligned} \|L(y) - L(\widetilde{y})\|_{T,K} = \max_{1 \leq k \leq K} \Bigg\| \eta_k \int_0^{\cdot} \left(y_k(s) - \widetilde{y}_k(s) \right) ds + \sum_{j=1}^{J} p_{jk} \Bigg[\psi \left(\xi_j + \sum_{l=1}^{K} q_{lj} \eta_l \int_0^{\cdot} y_l(s)\, ds \right) \\ - \psi \left(\xi_j + \sum_{l=1}^{K} q_{lj} \eta_l \int_0^{\cdot} \widetilde{y}_l(s)\, ds \right) \Bigg] \Bigg\|_T \end{aligned}$$

$$\begin{aligned}
&\leq \max_{1\leq k\leq K}\left\{\eta T\|y_k-\widetilde{y}_k\|_T+\sum_{j=1}^{J}\sum_{l=1}^{K}\eta T\|y_l-\widetilde{y}_l\|_T\right\}\\
&\leq \eta T\|y-\widetilde{y}\|_{T,K}+\eta TJK\|y-\widetilde{y}\|_{T,K}\\
&\leq \eta TJK\|y-\widetilde{y}\|_{T,K}. \qquad (139)
\end{aligned}$$

Equation (139) shows that L is Lipschitz continuous, so by Equations (136) – (137) we have

$$\left(y^{\infty}_{1,T},\ldots,y^{\infty}_{K,T}\right)\leftarrow\left(y^{n+1}_{1}|_{[0,T]},\ldots,y^{n+1}_{K}|_{[0,T]}\right)=L\left(y^{n}_{1}|_{[0,T]},\ldots,y^{n}_{K}|_{[0,T]}\right)\to L\left(y^{\infty}_{1,T},\ldots,y^{\infty}_{K,T}\right), \quad (140)$$

as $n\to\infty$. By uniqueness of limits in metric spaces, it follows that

$$L\left(y^{\infty}_{1,T},\ldots,y^{\infty}_{K,T}\right)=\left(y^{\infty}_{1,T},\ldots,y^{\infty}_{K,T}\right), \quad (141)$$

implying that $\left(y^{\infty}_{1,T},\ldots,y^{\infty}_{K,T}\right)$ satisfies Equation (98).

C.2 Uniqueness of the Element in D^T_K Satisfying Equation (98)

We show that $\left(y^{\infty}_{1,T},\ldots,y^{\infty}_{K,T}\right)$ is the unique element in D^K_T satisfying Equation (98). To that end, suppose that $(y_1,\ldots,y_K),(\widetilde{y}_1,\ldots,\widetilde{y}_K)\in D^K_T$ both satisfy Equation (98). Define

$$m:=\inf\left\{n\geq 1:\frac{n}{2}\left(2JK\eta\right)^{-1}>T\right\}. \quad (142)$$

Then for $t_1=\frac{1}{2}\left(2JK\eta\right)^{-1}$ we have

$$\begin{aligned}
\max_{1\leq k\leq K}\|y_k-\widetilde{y}_k\|_t &= \max_{1\leq k\leq K}\left\|\eta_k\int_0^{\cdot}\left(y_k(s)-\widetilde{y}(s)\right)ds+\sum_{j=1}^{J}p_{jk}\left[\psi\left(\xi_j+\sum_{l=1}^{K}q_{lj}\eta_l\int_0^{\cdot}y_l(s)\,ds\right)\right.\right.\\
&\qquad\left.\left.-\psi\left(\xi_j+\sum_{l=1}^{K}q_{lj}\eta_l\int_0^{\cdot}\widetilde{y}_l(s)\,ds\right)\right]\right\|_{t_1}\\
&\leq \max_{1\leq k\leq K}\eta\int_0^{t_1}\|y_k-\widetilde{y}_k\|_{t_1}ds+\sum_{j=1}^{J}\sum_{l=1}^{K}\eta\int_0^{t_1}\|y_l-\widetilde{y}_l\|_{t_1}ds\\
&\leq 2JK\eta t_1\max_{1\leq k\leq K}\|y_k-\widetilde{y}_k\|_{t_1}. \qquad (143)
\end{aligned}$$

Since $2JK\eta t_1=\frac{1}{2}<1$, Equation (143) implies that $\max_{1\leq k\leq K}\|y_k-\widetilde{y}_k\|_{t_1}=0$, so that

$$y=\widetilde{y}\quad\text{on}\quad[0,t_1]. \quad (144)$$

Now for $t_2 = (2JK\eta)^{-1}$ we have

$$\begin{aligned}
\max_{1\le k\le K} \|y_k - \widetilde{y}_k\|_{t_2} &\le \eta \left\| \int_0^{\cdot} |y_k(s) - \widetilde{y}_k(s)|\, ds \right\|_{t_2} + \sum_{j=1}^{J}\sum_{l=1}^{K} \eta \left\| \int_0^{\cdot} |y_l(s) - \widetilde{y}_l(s)|\, ds \right\|_{t_2} \\
&= \max_{1\le k\le K} \eta \left\| \int_0^{t_1} |y_k(s) - \widetilde{y}_k(s)|\, ds + \int_{t_1}^{\cdot} |y_k(s) - \widetilde{y}_k(s)|\, ds \right\|_{t_2} \\
&\qquad + \eta J \sum_{l=1}^{K} \left\| \int_0^{t_1} |y_l(s) - \widetilde{y}_l(s)|\, ds + \int_{t_1}^{\cdot} |y_l(s) - \widetilde{y}_l(s)|\, ds \right\|_{t_2} \\
&\le \eta t_1 \max_{1\le k\le K} \|y_k - \widetilde{y}_k\|_{t_1} + (t_2 - t_1) \max_{1\le k\le K} \|y_k - \widetilde{y}_k\|_{t_2} \\
&\qquad + \eta J \sum_{l=1}^{K} \left[t_1 \|y_l - \widetilde{y}_l\|_{t_1} + (t_2 - t_1) \|y_l - \widetilde{y}_l\|_{t_2} \right] \\
&\le 2\eta JK\,(t_2 - t_1) \max_{1\le k\le K} \|y_k - \widetilde{y}_k\|_{t_2},
\end{aligned} \tag{145}$$

where in the last inequality we use the fact that Equation (144) holds. Since $2\eta JK(t_2 - t_1) = \frac{1}{2} < 1$, Equation (145) implies that $\|y_k - \widetilde{y}_k\|_{t_2} = 0$, so that

$$y = \widetilde{y} \quad \text{on} \quad [0, t_2]. \tag{146}$$

We can continue in this iterative manner to show that for each $n = 1, \ldots, m-1$,

$$y = \widetilde{y} \quad \text{on} \quad [0, t_n], \tag{147}$$

where

$$t_n = \frac{n}{2}(2JK\eta)^{-1} \quad \text{for all} \quad n = 1, \ldots, m-1. \tag{148}$$

If in fact $t_{m-1} = T$, then we are done by Equation (147). If on the other hand $t_{m-1} < T$, then we can take $t_m = T$ and show using the same argument that

$$y = \widetilde{y} \quad \text{on} \quad [0, t_m], \tag{149}$$

which would then complete the proof.

C.3 Extension to a Unique Element in D^K Satisfying Equation (98)

We use the constructed unique solution in D_T^K to define a $(y_1^\infty, \ldots, y_K^\infty)$ that is the unique element in D^K satisfying Equation (98). To that end, define $(y_1^\infty, \ldots, y_K^\infty) \in D^K$ by

$$\left(y_1^\infty, \ldots, y_K^\infty\right)(t) := \left(y_{1,T}^\infty, \ldots, y_{K,T}^\infty\right)(t) \quad \text{for} \quad t \in [0, T]. \tag{150}$$

To complete the proof, we must show that $(y_1^\infty, \ldots, y_K^\infty)$ is well-defined and is the unique element in D^K satisfying Equation (98). To prove that $(y_1^\infty, \ldots, y_K^\infty)$ is well-defined, we must show that whenever $t \in [0, T_1] \cap [0, T_2]$ that

$$\left(y_{1,T_1}^\infty, \ldots, y_{K,T_1}^\infty\right)(t) = \left(y_{1,T_2}^\infty, \ldots, y_{K,T_2}^\infty\right)(t). \tag{151}$$

Without loss of generality, suppose that $T_1 \leq T_2$. Then $\left(y_{1,T_2}^\infty, \ldots, y_{K,T_2}^\infty\right)|_{[0,T_1]} \in D_{T_1}^K$ satisfies Equation (98) for all $t \leq T_1$. By uniqueness,

$$\left(y_{1,T_1}^\infty, \ldots, y_{K,T_1}^\infty\right) = \left(y_{1,T_2}^\infty, \ldots, y_{K,T_2}^\infty\right)|_{[0,T_1]},$$

so we have that

$$\left(y_{1,T_1}^\infty, \ldots, y_{K,T_1}^\infty\right)(t) = \left(y_{1,T_2}^\infty, \ldots, y_{K,T_2}^\infty\right)|_{[0,T_1]}(t) = \left(y_{1,T_2}^\infty, \ldots, y_{K,T_2}^\infty\right)(t),$$

proving Equation (151) holds. Next we show that $(y_1^\infty, \ldots, y_K^\infty)$ satisfies Equation (98). Indeed, for any $t \in [0, \infty)$, there exists a $T > 0$ such that Equation (150) holds. But since the right hand side of Equation (150) satisfies Equation (98) at t, so does $(y_1^\infty, \ldots, y_K^\infty)$, which gives the desired result. Finally, we show that $(y_1^\infty, \ldots, y_K^\infty)$ is unique. To that end, let $(y_1, \ldots, y_K) \in D^K$ also satisfy Equation (98) and fix $t \in [0, \infty)$. Let $T > 0$ be such that $t \in [0, T]$. Then $(y_1, \ldots, y_K)|_{[0,T]} \in D_T^K$ satisfies Equation (98) for all $t \in [0, T]$. Since $(y_1^\infty, \ldots, y_K^\infty)|_{[0,T]}$ also satisfies Equation (98) for all $t \in [0, T]$, uniqueness implies that

$$(y_1^\infty, \ldots, y_K^\infty)(t) = (y_1^\infty, \ldots, y_K^\infty)|_{[0,T]}(t) = (y_1, \ldots, y_K)|_{[0,T]}(t) = (y_1, \ldots, y_K)(t).$$

Therefore, $(y_1^\infty, \ldots, y_K^\infty) = (y_1, \ldots, y_K)$, as desired. This completes the proof. □

We conclude Appendix C with a proof showing that the sequence defined by Equation (136) is Cauchy:

Claim 1. *For each $T > 0$, the sequence $\left\{\left(y_1^n|_{[0,T]}, \dots, y_K^n|_{[0,T]}\right) : n = 0, 1, \dots\right\}$ defined by Equation (136) is Cauchy in D_K^T (with respect to the norm $\|\cdot\|_{T,K}$).*

Proof. For any $t \in [0, T]$ note that

$$\|y_k^1 - y_k^0\|_t = \left\|\zeta_k - \sum_{j=1}^J p_{jk}\psi\left(\xi_j\right)\right\|_t \le \|\zeta_k\|_t + \sum_{j=1}^J \|\psi(\xi_j)\|_t$$
$$\le \max_{1\le k\le K} \|\zeta_k\|_T + \sum_{j=1}^J \|\psi(\xi_j)\|_T =: C_T. \tag{152}$$

Let $\eta := \max_{1\le k\le K} \eta_k$ and fix $\delta \in (0, T)$ such that

$$2\delta\eta JK < 1. \tag{153}$$

Then for $n \ge 2$,

$$\begin{aligned}
\|y_k^n - y_k^{n-1}\|_\delta = \Bigg\|\eta_k \int_0^\cdot \left(y_k^{n-2}(s) - y_k^{n-1}(s)\right) ds \\
+ \sum_{j=1}^J p_{jk}\left[\psi\left(\xi_j + \sum_{l=1}^K q_{lj}\eta_l \int_0^\cdot y_l^{n-2}\, ds\right) - \psi\left(\xi_j + \sum_{l=1}^K q_{lj}\eta_l \int_0^\cdot y_l^{n-1}\, ds\right)\right]\Bigg\|_\delta \\
&\le \eta_k\delta\|y_k^{n-1} - y_k^{n-2}\|_\delta + \sum_{j=1}^J p_{jk}\left\|\sum_{l=1}^K q_{lj}\eta_l \int_0^\cdot \left(y_l^{n-1}(s) - y_l^{n-2}(s)\right) ds\right\|_\delta \\
&\le \eta_k\delta\|y_k^{n-1} - y_k^{n-2}\|_\delta + \sum_{j=1}^J p_{jk} \sum_{l=1}^K q_{lj}\eta_l\delta\|y_l^{n-1} - y_l^{n-2}\|_\delta \\
&\le \delta\eta\left(\|y_k^{n-1} - y_k^{n-2}\|_\delta + J\sum_{l=1}^K \|y_l^{n-1} - y_l^{n-2}\|_\delta\right) \\
&\le 2\delta\eta J \sum_{l=1}^K \|y_l^{n-1} - y_l^{n-2}\|_\delta,
\end{aligned} \tag{154}$$

where the first inequality follows from Equation (90). Doing another iteration of Equation (154) gives

$$\|y_k^n - y_k^{n-1}\|_\delta \le 2\delta\eta J \sum_{l=1}^K \|y_l^{n-1} - y_l^{n-2}\|_\delta \le (2\delta\eta J)^2 K \sum_{l=1}^K \|y_l^{n-2} - y_l^{n-3}\|_\delta. \tag{155}$$

Continuing in the way and using Equation (152), we find that

$$\|y_k^n - y_k^{n-1}\|_\delta \le (2\delta\eta JK)^{n-1} C_T \quad \text{for all} \quad n \ge 1. \tag{156}$$

For each $k = 1, \dots, K$, we now prove that for all $m \ge 1$ and $n \ge 1$ that

$$\|y_k^n - y_k^{n-1}\|_{m\delta} \le mn^m (2\delta\eta JK)^{n-1} C_T. \tag{157}$$

We proceed by (strong) induction on m. The base case follows by Equation (156) since for any $n \geq 1$,

$$\|y_k^n - y_k^{n-1}\|_{1\cdot\delta} \leq (2\delta\eta JK)^{n-1} C_T \leq 1 \cdot n^1 (2\delta\eta JK)^{n-1} C_T, \tag{158}$$

which is the $m = 1$ case. For the inductive step assume that for all $n \geq 2$,

$$\|y_k^n - y_k^{n-1}\|_{r\delta} \leq rn^r (2\delta\eta JK)^{n-1} C_T \quad \text{for all} \quad r = 1, \ldots, m. \tag{159}$$

By Equation (159), it follows that for any $n \geq 2$,

$$\begin{aligned}
\|y_k^n - y_k^{n-1}\|_{(m+1)\delta} &= \Bigg\|\eta_k \int_0^{\cdot} \left(y_k^{n-2}(s) - y_k^{n-1}(s)\right) ds \\
&\qquad + \sum_{j=1}^{J} p_{jk}\Bigg[\psi\left(\xi_j + \sum_{l=1}^{K} q_{lj}\eta_l \int_0^{\cdot} y_l^{n-2}\, ds\right) \\
&\qquad\qquad - \psi\left(\xi_j + \sum_{l=1}^{K} q_{lj}\eta_l \int_0^{\cdot} y_l^{n-1}\, ds\right)\Bigg]\Bigg\|_{(m+1)\delta} \\
&\leq \eta_k \sum_{r=1}^{m+1} \int_{(r-1)\delta}^{r\delta} \|y_k^{n-1} - y_k^{n-2}\|_{r\delta}\, ds \\
&\qquad + \sum_{j=1}^{J} p_{jk} \Bigg\| \sum_{l=1}^{K} q_{lj}\eta_l \int_0^{\cdot} |y_l^{n-1}(s) - y_l^{n-2}(s)|\, ds \Bigg\|_{(m+1)\delta} \\
&\leq \delta\eta \sum_{r=1}^{m+1} \|y_k^{n-1} - y_k^{n-2}\| + \eta \sum_{j=1}^{J}\sum_{l=1}^{K}\sum_{r=1}^{m+1} \int_{(r-1)\delta}^{r\delta} \|y_l^{n-1} - y_l^{n-2}\|_{r\delta}\, ds \\
&= \delta\eta \left[\sum_{r=1}^{m+1} \|y_k^{n-1} - y_k^{n-2}\| + J \sum_{l=1}^{K}\sum_{r=1}^{m+1} \|y_l^{n-1} - y_l^{n-2}\|_{r\delta}\right] \\
&\leq 2\delta\eta J \sum_{l=1}^{K}\sum_{r=1}^{m+1} \|y_l^{n-1} - y_l^{n-2}\|_{r\delta} \\
&= 2\delta\eta J \sum_{l=1}^{K}\sum_{r=1}^{m} \|y_l^{n-1} - y_l^{n-2}\|_{r\delta} + 2\delta\eta J \sum_{l=1}^{K} \|y_l^{n-1} - y_l^{n-2}\|_{(m+1)\delta} \\
&\leq 2\delta\eta J \sum_{l=1}^{K}\sum_{r=1}^{m} r(n-1)^r (2\delta\eta JK)^{n-2} C_T + 2\delta\eta J \sum_{l=1}^{K} \|y_l^{n-1} - y_l^{n-2}\|_{(m+1)\delta} \\
&= (2\delta\eta JK)^{n-1} C_T \sum_{r=1}^{m} r(n-1)^r + 2\delta\eta J \sum_{l=1}^{K} \|y_l^{n-1} - y_l^{n-2}\|_{(m+1)\delta} \\
&\leq (2\delta\eta JK)^{n-1} C_T m n^m + 2\delta\eta J \sum_{l=1}^{K} \|y_l^{n-1} - y_l^{n-2}\|_{(m+1)\delta}.
\end{aligned} \tag{160}$$

By Equations (152) and (160), for $n = 2$ we have

$$\|y_k^2 - y_k^1\|_{(m+1)\delta} \leq (2\delta\eta JK)\, C_T m 2^m + 2\delta\eta J \sum_{l=1}^{K} \|y_l^1 - y_l^0\|_{(m+1)\delta}$$

$$\leq (2\delta\eta JK)\, C_T m 2^m + 2\delta\eta JK C_T$$

$$= (2\delta\eta JK)\, C_T\, (m2^m + 1)\,. \tag{161}$$

Similarly, by Equations (160) − (161), for $n = 3$ we have

$$\begin{aligned}\|y_k^3 - y_k^2\|_{(m+1)\delta} &\leq (2\delta\eta JK)^2\, C_T m 3^m + 2\delta\eta J \sum_{l=1}^{K} \|y_l^2 - y_l^1\|_{(m+1)\delta} \\ &\leq (2\delta\eta JK)^2\, C_T m 3^m + 2\delta\eta J \sum_{l=1}^{K} (2\delta\eta JK)\, C_T\, (m2^m + 1) \\ &= (2\delta\eta JK)^2\, C_T\, (m3^m + m2^m + 1)\,. \end{aligned}\tag{162}$$

Continuing in this iterative fashion, we find that for $n \geq 2$,

$$\begin{aligned}\|y_k^n - y_k^{n-1}\|_{(m+1)\delta} &\leq (2\delta\eta JK)^{n-1}\, C_T \left(m \sum_{i=2}^{n} i^m + 1 \right) \\ &\leq (2\delta\eta JK)^{n-1}\, C_T \left(m n^{m+1} + 1 \right) \\ &\leq (2\delta\eta JK)^{n-1}\, C_T (m+1) n^{m+1}. \end{aligned}\tag{163}$$

Because Equation (157) also holds for all $m \geq 1$ when $n = 1$ [by Equation (152)], this completes the inductive step. We conclude that Equation (157) holds for all $m \geq 1$ and $n \geq 1$.

By Equation (157), for all $n \geq 1$ and $k = 1, \ldots, K$ we have

$$\|y_k^n - y_k^{n-1}\|_T \leq \|y_k^n - y_k^{n-1}\|_{\lceil \delta^{-1} T\rceil \delta} \leq \lceil \delta^{-1} T \rceil n^{\lceil \delta^{-1} T \rceil} (2\delta JK)^{n-1}\, C_T, \tag{164}$$

implying that

$$\max_{1 \leq k \leq K} \|y_k^n - y_k^{n-1}\|_T \leq \lceil \delta^{-1} T \rceil n^{\lceil \delta^{-1} T \rceil} (2\delta JK)^{n-1}\, C_T. \tag{165}$$

Since by Equation (153) we have

$$\limsup_{n\to\infty} \left| \frac{\lceil \delta^{-1} T\rceil (n+1)^{\lceil \delta^{-1} T \rceil} (2\delta JK)^n\, C_T}{\lceil \delta^{-1} T\rceil n^{\lceil \delta^{-1} T \rceil} (2\delta JK)^{n-1}\, C_T} \right| = \limsup_{n\to\infty} \left[\frac{(n+1)^{\lceil \delta^{-1} T \rceil}}{n^{\lceil \delta^{-1} T \rceil}} \cdot 2\delta JK \right] = 2\delta JK < 1, \tag{166}$$

the ratio test implies that

$$\sum_{n=1}^{\infty} \lceil \delta^{-1} T \rceil n^{\lceil \delta^{-1} T \rceil} (2\delta JK)^{n-1}\, C_T < \infty, \tag{167}$$

which in turn implies that the sequence $\left\{ \left(y_1^n|_{[0,T]}, \ldots, y_K^n|_{[0,T]} \right) : n = 0, 1, \ldots \right\}$ is Cauchy in D_T^K. □

D Miscellaneous Proofs and Derivations

D.1 Derivation of Equations (40) – (41)

We begin by deriving Equation (40), the diffusion scaled system equation for the single-server stations. By Equations (8), (10), and (12) – (13), we have

$$\begin{aligned} Q_j^n(t) &= Q_j^n(0) + A_j^n(t) - D_j^n(t) \\ &= Q_j^n(0) + \sum_{k=1}^{K} \Psi_{kj}\left(M_k\left(\eta_k \int_0^t V_k^n(s)\,ds\right)\right) - N_j\left(n\mu_j T_j^n(t)\right). \end{aligned} \tag{168}$$

Elementary algebraic manipulations and the appropriate scaling in Equations (23) – (35) applied to Equation (168) yields

$$\begin{aligned} \hat{Q}_j^n(t) &= n^{-1/2}Q_j^n(0) + \sum_{k=1}^{K} n^{-1/2}\Psi_{kj}\left(M_k\left(\eta_k \int_0^t V_k^n(s)\,ds\right)\right) - n^{-1/2}N_j\left(n\mu_j T_j^n(t)\right) \\ &= \hat{Q}_j^n(0) + \sum_{k=1}^{K} n^{-1/2}\left[\Psi_{kj}\left(M_k\left(\eta_k \int_0^t V_k^n(s)\,ds\right)\right) - q_{kj}M_k\left(\eta_k \int_0^t V_k^n(s)\,ds\right)\right] \\ &\qquad + \sum_{k=1}^{K} q_{kj}n^{-1/2}M_k\left(\eta_k \int_0^t V_k^n(s)\,ds\right) - n^{-1/2}\left[N_j\left(n\mu_j T_j^n(t)\right) - n\mu_j T_j^n(t)\right] \\ &\qquad\qquad - n^{-1/2}n\mu_j T_j^n(t) \\ &= \hat{Q}_j^n(0) + \sum_{k=1}^{K} n^{-1/2}\left[\Psi_{kj}\left(nn^{-1}M_k\left(n\eta_k \int_0^t n^{-1}V_k^n(s)\,ds\right)\right) - q_{kj}nn^{-1}M_k\left(n\eta_k \int_0^t n^{-1}V_k^n(s)\,ds\right)\right] \\ &\qquad + \sum_{k=1}^{K} q_{kj}n^{-1/2}\left[M_k\left(n\eta_k \int_0^t n^{-1}V_k^n(s)\,ds\right) - n\eta_k \int_0^t n^{-1}V_k^n(s)\,ds\right] \\ &\qquad\qquad + \sum_{k=1}^{K} q_{kj}n^{-1/2}\eta_k \int_0^t V_k^n(s)\,ds \\ &\qquad\qquad\qquad - n^{-1/2}\left[N_j\left(n\mu_j T_j^n(t)\right) - n\mu_j T_j^n(t)\right] - \sqrt{n}\mu_j\left(t - I_j^n(t)\right) \\ &= \hat{Q}_j^n(0) + \sum_{k=1}^{K} n^{-1/2}\left[\Psi_{kj}\left(n\bar{M}_k^n\left(\eta_k \int_0^t \bar{\bar{V}}_k^n(s)\,ds\right)\right) - nq_{kj}\bar{M}_k^n\left(\eta_k \int_0^t \bar{\bar{V}}_k^n(s)\,ds\right)\right] \\ &\qquad + \sum_{k=1}^{K} q_{kj}n^{-1/2}\left[M_k\left(n\eta_k \int_0^t \bar{\bar{V}}_k^n(s)\,ds\right) - n\eta_k \int_0^t \bar{\bar{V}}_k^n(s)\,ds\right] \\ &\qquad\qquad + \sum_{k=1}^{K} q_{kj}n^{-1/2}\eta_k \int_0^t \left[\sqrt{n}\hat{V}_k^n(s) + nm_k\right]ds \\ &\qquad\qquad\qquad - n^{-1/2}\left[N_j\left(n\mu_j T_j^n(t)\right) - n\mu_j T_j^n(t)\right] - \sqrt{n}\mu_j t - \sqrt{n}\mu_j I_j^n(t) \end{aligned}$$

$$
\begin{aligned}
&= \hat{Q}_j^n(0) + \sum_{k=1}^K \hat{\Psi}_{kj}^n \left(\bar{M}_k^n \left(\eta_k \int_0^t \bar{\bar{V}}_k^n(s)\,ds \right)\right) + \sum_{k=1}^K q_{kj} \hat{M}_k^n \left(\eta_k \int_0^t \bar{\bar{V}}_k^n(s)\,ds \right) - \hat{N}_j^n\left(\mu_j T_j^n(t)\right) \\
&\qquad + \sum_{k=1}^K q_{kj}\eta_k \int_0^t \hat{V}_k^n(s)\,ds - \mu_j \hat{I}_j^n(t) + t\sqrt{n}\left[\sum_{k=1}^K q_{kj}\eta_k m_k - \mu_j\right] \\
&= \hat{\xi}_j^n(t) + \sum_{k=1}^K q_{kj}\eta_k \int_0^t \hat{V}_k^n(s)\,ds - \mu_j \hat{I}_j^n(t),
\end{aligned}
$$

which is Equation (40). Note that the final equality in the above calculation follows from Equation (38), the definition of m_k (given in Section 3), and the heavy traffic condition in Equation (21).

Next we derive Equation (41), the diffusion scaled system equation for the infinite-server stations. By Equations (9) and (11) – (13), we have

$$
\begin{aligned}
V_k^n(t) &= V_k^n(0) + E_k^n(t) - F_k^n(t) \\
&= V_k^n(0) + \sum_{j=1}^J \Phi_{jk}\left(N_j\left(n\mu_j T_j^n(t)\right)\right) - M_k\left(\eta_k \int_0^t V_k^n(s)\,ds\right).
\end{aligned} \tag{169}
$$

As before, elementary algebraic manipulations and the appropriate scaling in Equations (23) – (35) applied to Equation (169) yields

$$
\begin{aligned}
\hat{V}_k^n(t) &= n^{-1/2}V_k^n(0) + \sum_{j=1}^J n^{-1/2}\Phi_{jk}\left(N_j\left(n\mu_j T_j^n(t)\right)\right) - n^{-1/2}M_k\left(\eta_k \int_0^t V_k^n(s)\,ds\right) \\
&= \hat{V}_k^n(0) + \sum_{j=1}^J n^{-1/2}\left[\Phi_{jk}\left(N_j\left(n\mu_j T_j^n(t)\right)\right) - p_{jk}N_j\left(n\mu_j T_j^n(t)\right)\right] \\
&\qquad + \sum_{j=1}^J n^{-1/2}p_{jk}N_j\left(n\mu_j T_j^n(t)\right) - n^{-1/2}\left[M_k\left(\eta_k \int_0^t V_k^n(s)\,ds\right) - \eta_k \int_0^t V_k^n(s)\,ds\right] \\
&\qquad\qquad - n^{-1/2}\eta_k \int_0^t V_k^n(s)\,ds \\
&= \hat{V}_k^n(0) + \sum_{j=1}^J n^{-1/2}\left[\Phi_{jk}\left(nn^{-1}N_j\left(n\mu_j T_j^n(t)\right)\right) - p_{jk}nn^{-1}N_j\left(n\mu_j T_j^n(t)\right)\right] \\
&\qquad + \sum_{j=1}^J n^{-1/2}p_{jk}\left[N_j\left(n\mu_j T_j^n(t)\right) - n\mu_j T_j^n(t)\right] + \sum_{j=1}^J n^{-1/2}p_{jk}n\mu_j T_j^n(t) \\
&\qquad\qquad - n^{-1/2}\left[M_k\left(n\eta_k \int_0^t n^{-1}V_k^n(s)\,ds\right) - n\eta_k \int_0^t n^{-1}V_k^n(s)\,ds\right] \\
&\qquad\qquad\qquad - n^{-1/2}\eta_k \int_0^t \left[\sqrt{n}\hat{V}_k^n(s) + nm_k\right]ds \\
&= \hat{V}_k^n(0) + \sum_{j=1}^J n^{-1/2}\left[\Phi_{jk}\left(n\bar{N}_j^n\left(\mu_j T_j^n(t)\right)\right) - p_{jk}n\bar{N}_j\left(\mu_j T_j^n(t)\right)\right] \\
&\qquad + \sum_{j=1}^J p_{jk}\hat{N}_j^n\left(\mu_j T_j^n(t)\right) + \sum_{j=1}^J \sqrt{n}p_{jk}\mu_j\left(t - I_j^n(t)\right)
\end{aligned}
$$

$$
\begin{aligned}
&- n^{-1/2}\left[M_k\left(n\eta_k \int_0^t \bar{\bar{V}}_k^n(s)\,ds\right) - n\eta_k \int_0^t \bar{\bar{V}}_k^n(s)\,ds\right] \\
&- \eta_k \int_0^t \hat{V}_k^n(s)\,ds - t\sqrt{n}\eta_k m_k \\
= {}& \hat{V}_k^n(0) + \sum_{j=1}^J \hat{\Phi}_{jk}^n\left(\bar{N}_j^n\left(\mu_j T_j^n(t)\right)\right) + \sum_{j=1}^J p_{jk}\hat{N}_j^n\left(\mu_j T_j^n(t)\right) - \hat{M}_k^n\left(\eta_k \int_0^t \bar{\bar{V}}_k^n(s)\,ds\right) \\
&- \eta_k \int_0^t \hat{V}_k^n(s)\,ds - \sum_{j=1}^J p_{jk}\mu_j \hat{I}_j^n(t) + t\sqrt{n}\left[\sum_{j=1}^J p_{jk}\mu_j - \eta_k m_k\right] \\
= {}& \hat{\zeta}_k^n(t) - \eta_k \int_0^t \hat{V}_k^n(s)\,ds - \sum_{j=1}^J p_{jk}\mu_j \hat{I}_j^n(t),
\end{aligned}
$$

which is Equation (41). Note that the final equality in the above calculation follows from Equation (39) and the definition of m_k.

Below is the proof of a result used in the proof of Lemma 1 in Section 5:

Lemma 5. *Let $\{X^n\}_{n=1}^\infty$ be a random sequence in D such that $\|X^n\|_T \Rightarrow 0$ for all $T > 0$. Then $X^n \Rightarrow \mathbf{0}$ in D.*

Proof. Note that $\|X^n\|_T \Rightarrow 0$ for all $T > 0$ is equivalent to $\|X^n\|_T \xrightarrow{p} 0$ for all $T > 0$. We claim that $X^n \xrightarrow{p} \mathbf{0}$ in D. This amounts to showing that for all $0 < \epsilon < 1$,

$$
P\left(\int_0^\infty e^{-t}\left[d_t\left(X^n, 0\right) \wedge 1\right] dt > \epsilon\right) \to 0 \quad \text{as} \quad n \to \infty; \tag{170}
$$

for example, see Chapter 3, Section 3 of Whitt (2002). More explicitly, we must show that

$$
P\left(\int_0^\infty e^{-t}\left[\inf_{\lambda \in \Lambda_t}\{\|X^n \circ \lambda\|_t \vee \|\lambda - e\|_t\} \wedge 1\right] dt > \epsilon\right) \to 0 \quad \text{as} \quad n \to \infty. \tag{171}
$$

But,

$$
\begin{aligned}
&P\left(\int_0^\infty e^{-t}\left[\inf_{\lambda \in \Lambda_t}\{\|X^n \circ \lambda\|_t \vee \|\lambda - e\|_t\} \wedge 1\right] dt > \epsilon\right) \\
&\le P\left(\int_0^\infty e^{-t}\left[\|X^n\|_t \wedge 1\right] dt > \epsilon\right) \\
&\le P\left(\int_0^T e^{-t}\|X^n\|_t\,dt + \int_T^\infty e^{-t}\,dt > \epsilon\right),
\end{aligned} \tag{172}
$$

for all $T > 0$. Now fix $T > 0$ to be such that $\int_T^\infty e^{-t}\,dt = \epsilon/2$. Continuing from Equation (172),

$$
\begin{aligned}
P\left(\int_0^T e^{-t}\|X^n\|_t\,dt + \int_T^\infty e^{-t}\,dt > \epsilon\right) &= P\left(\int_0^T e^{-t}\|X^n\|_t\,dt > \frac{\epsilon}{2}\right) \le P\left(\|X^n\|_T\left(1 - \frac{\epsilon}{2}\right) > \frac{\epsilon}{2}\right) \\
&= P\left(\|X^n\|_T > \frac{\epsilon}{2}\left(1 - \frac{\epsilon}{2}\right)\right).
\end{aligned} \tag{173}
$$

Since $\|X^n\|_T \xrightarrow{p} 0$, it follows from Equation (173) that $X^n \xrightarrow{p} \mathbf{0}$ in D. Since convergence in probability implies convergence in distribution, we conclude that $X^n \Rightarrow \mathbf{0}$ in D. □

D.2 Omitted Details From the Proof of Lemma 1

We show that the right hand sides of Equations (64) and (65) converge weakly to a nondegenerate limit and conclude that $\|\bar{\xi}_j^n\|_T \Rightarrow 0$ and $\|\bar{\zeta}_k^n\|_T \Rightarrow 0$. We show this only for Equation (64); the proof for Equation (65) is identical. First, by Equation (46) and the continuous mapping theorem,

$$\|\hat{Q}_j^n(0)\|_T \Rightarrow \|Q_j(0)\|_T, \tag{174}$$

so that

$$n^{-1/2}\|\hat{Q}_j^n(0)\|_T \Rightarrow 0 \cdot \|Q_j(0)\|_T = 0. \tag{175}$$

Define $\widetilde{M}_k(t) = M_k(\eta_k t)$, so that $\widetilde{M}_k$ is a Poisson process with rate η_k. By the functional central limit theorem for renewal processes [see Theorem 14.6 in Billingsley (1999)],

$$\hat{\widetilde{M}}_k^n \Rightarrow \eta_k^{1/2} B_k, \tag{176}$$

where B_k is a standard Brownian motion and where

$$\hat{\widetilde{M}}_k^n(t) = \frac{\widetilde{M}_k(nt) - n\eta_k t}{\sqrt{n}} \equiv \frac{M_k(n\eta_k t) - n\eta_k t}{\sqrt{n}} = \hat{M}_k^n(\eta_k t). \tag{177}$$

By Equations (176) – (177) and the continuous mapping theorem, note that

$$n^{-1/2}\|\hat{M}_k^n(\eta_k \cdot)\|_T \Rightarrow 0 \cdot \|\eta^{1/2} B_k\|_T = 0. \tag{178}$$

Furthermore, observe by Equations (35) and (177) that

$$\bar{M}_k^n(\eta_k t) = \frac{\hat{M}_k^n(\eta_k t)}{\sqrt{n}} + \eta_k t. \tag{179}$$

Therefore, by Equations (176) and (179),

$$\bar{M}_k^n(\eta_k \cdot) = \frac{\hat{M}_k(\eta_k \cdot)}{\sqrt{n}} + \eta_k e \Rightarrow 0 \cdot \eta_k^{1/2} B_k + \eta_k e = \eta_k e, \tag{180}$$

where $e:[0,\infty)\to[0,\infty)$ is the identity map $e(t)=t$. By Equation (28) and Donsker's theorem [see Theorem 4 in Glynn (1990)],

$$\hat{\Phi}^n_{kj} \Rightarrow \sqrt{q_{kj}\left(1-q_{kj}\right)}B_{kj}, \tag{181}$$

where B_{kj} is a standard Brownian motion. Since $t\mapsto \bar{M}^n_k(\eta t)$ is non-decreasing function from $[0,\infty)$ to $[0,\infty)$, the random time change theorem [see Proposition 5 in Glynn (1990)] and Equations (180) – (181) yield

$$\hat{\Psi}^n_{kj}\left(\bar{M}^n_k\left(\eta_k\,\cdot\right)\right) \Rightarrow \sqrt{q_{kj}\left(1-q_{kj}\right)}B_{kj}(\eta_k\,\cdot). \tag{182}$$

By Equation (182) and the continuous mapping theorem,

$$n^{-1/2}\|\hat{\Psi}^n_{kj}\left(\bar{M}^n_k\left(\eta_k\,\cdot\right)\right)\|_T \to 0\cdot\|\sqrt{q_{kj}\left(1-q_{kj}\right)}B_{kj}(\eta_k\,\cdot)\|_T = 0. \tag{183}$$

Similar to Equation (176), we have that

$$\hat{N}^n_j\left(\mu_j\,\cdot\right) \Rightarrow \mu_j^{1/2}B_j, \tag{184}$$

where B_j is standard Brownian motion. Another application of the continuous mapping theorem with Equation (184) gives

$$n^{-1/2}\|\hat{N}^n_j(\mu_j\cdot)\|_T \Rightarrow 0\cdot\|\mu_j^{1/2}B_j\|_T = 0. \tag{185}$$

Finally, by Equations (64), (175), (178), (183), and (185), it follows that

$$\begin{aligned}\|\bar{\xi}^n_j\|_T \le n^{-1/2}\|\hat{Q}^n_j(0)\|_T + \sum_{k=1}^K n^{-1/2}\|\hat{\Psi}^n_{kj}\left(\bar{M}^n_k\left(\eta_k\cdot\right)\right)\|_T + n^{-1/2}\|\hat{N}^n_j\left(\mu_j\cdot\right)\|_T \\ + \sum_{k=1}^K n^{-1/2}\|\hat{M}^n_k\left(\eta_k\cdot\right)\|_T \Rightarrow 0,\end{aligned}$$

as desired.