



Working Paper 12-17 (11)
Statistics and Econometrics Series
June, 2012

Departamento de Estadística
Universidad Carlos III de Madrid
Calle Madrid, 126
28903 Getafe (Spain)
Fax (34) 91 624-98-49

CLOSED QUEUEING NETWORKS UNDER CONGESTION: NON-BOTTLENECK INDEPENDENCE AND BOTTLENECK CONVERGENCE

Jonatha Anselmi¹, Bernardo D'Auria² and Neil Walton³

Abstract

We analyze the behavior of closed product-form queueing networks when the number of customers grows to infinity and remains proportionate on each route (or class). First, we focus on the stationary behavior and prove the conjecture that the stationary distribution at non-bottleneck queues converges weakly to the stationary distribution of an ergodic, open product-form queueing network. This open network is obtained by replacing bottleneck queues with per-route Poissonian sources whose rates are determined by the solution of a strictly concave optimization problem. Then, we focus on the transient behavior of the network and use fluid limits to prove that the amount of fluid, or customers, on each route eventually concentrates on the bottleneck queues only, and that the long-term proportions of fluid in each route and in each queue solve the dual of the concave optimization problem that determines the throughputs of the previous open network.

Keywords: closed queueing networks; product-form; asymptotic independence; fluid limit; large population.

Acknowledgements: The second author is partially supported by the Spanish Ministry of Education and Science Grants MTM2010-16519, SEJ2007-64500 and RYC-2009-04671.

¹ Basque Center for Applied Mathematics (BCAM), Alameda de Mazarredo 14, E-48009 Bilbao, Basque Country, Spain, anselmi@bcamath.org

² Universidad Carlos III de Madrid, Avda. Universidad 30, 28911 Leganés, Madrid, Spain, bernardo.dauria@uc3m.es

³ University of Amsterdam, Science Park 904, 1098 XH, Amsterdam, The Netherlands, n.s.walton@uva.nl

Closed queueing networks under congestion: non-bottleneck independence and bottleneck convergence

J. Anselmi*

B. D'Auria†

N. Walton‡

June 2012

Abstract

We analyze the behavior of closed product-form queueing networks when the number of customers grows to infinity and remains proportionate on each route (or class). First, we focus on the stationary behavior and prove the conjecture that the stationary distribution at non-bottleneck queues converges weakly to the stationary distribution of an ergodic, open product-form queueing network. This open network is obtained by replacing bottleneck queues with per-route Poissonian sources whose rates are determined by the solution of a strictly concave optimization problem. Then, we focus on the transient behavior of the network and use fluid limits to prove that the amount of fluid, or customers, on each route eventually concentrates on the bottleneck queues only, and that the long-term proportions of fluid in each route and in each queue solve the dual of the concave optimization problem that determines the throughputs of the previous open network.

1 Introduction

Complex systems such as communication and computer networks are composed of a number of interacting particles (or customers) that exhibit important congestion phenomena as their level of interaction grows. The dynamics of such systems are affected by the randomness of their underlying events, e.g., arrivals of units of work, and can be described stochastically in terms of queueing network models. Provided that these are tractable, they allow one to make predictions on the performance achievable by the system, to optimize the network configuration and to perform capacity-planning studies. These objectives are usually difficult to achieve without a mathematical model because real systems are huge in size; e.g., Urgaonkar et al. [34].

The focus of this paper is on the well-known class of closed queueing network models introduced in Kelly [18] and Baskett et al. [3]. Specifically, a fixed number of customers circulate in a network following given *routes*. A route is a sequence of queues (or stations) that forms a cycle in the network. In terms of the amount of service required on each queue, users belonging to the same route are statistically equivalent. In contrast, users belonging to different routes can be statistically different. The stationary probability distribution of the numbers of customers in each queue of this class of queueing networks has the *product-form* property, which formally means that it can be written as the product of simple terms associated to each queue up to a normalizing constant. This surprising property represents a big step forward for the understanding of the stationary behavior of these queueing networks. However, due to the quick growth of the state space and despite the attention devoted to this problem during the last decades, the computation of the normalizing constant remains a notoriously-difficult task especially when the number of customers is large. This limits the application of these models to optimization and dimensioning studies of real systems.

During the last decades, several approaches have been investigated to assess the stationary behavior of closed product-form queueing networks with a large number of customers. A large body of the literature aims at developing exact algorithms for the efficient computation of the normalizing constant or stationary performance indices such as throughputs and queue lengths; see, e.g., Reiser and Kobayashi [26], Reiser and Lavenberg [27], Harrison and Coury [14], Casale [6] and the references therein. To the best of our knowledge, no exact algorithm has a running time that is polynomial with respect to the numbers of routes, customers and queues. Motivated by this difficulty, a number of alternative analyses emerged in the literature for the stationary behavior. These mainly consist in:

- Using the mean value analysis (MVA) by Reiser and Lavenberg [27] to develop iterative or fixed-point algorithms; e.g., Schweitzer [29], Chandy and Neuse [8], Pattipati et al. [24], Wang et al. [36]. While these techniques improve the running time of MVA, there is no guarantee that they converge to the exact solution except for particular cases.
- Developing efficient bottleneck identification techniques; e.g., Schweitzer [30], Schweitzer et al. [31], Casale and Serazzi [7], Anselmi and Cremonesi [1]. This approach aims at reducing the network size by ignoring the impact of the stations that have a minor influence on the overall performance.

*BCAM, Alameda de Mazarredo 14, E-48009 Bilbao, Basque Country, Spain, anselmi@bcamath.org

†Universidad Carlos III de Madrid, Avda. Universidad 30, 28911 Leganés, Madrid, Spain, bernardo.dauria@uc3m.es

‡University of Amsterdam, Science Park 904, 1098 XH, Amsterdam, The Netherlands, n.s.walton@uva.nl

- Studying the network behavior in some limiting regime, e.g. for the instance of Jackson networks see Goodman and Massey [12]. Our closed queueing context, one can assume the existence of a $M/M/\infty$ queue within the network and allow the total number of jobs, say n , grow in proportion to its service times; McKenna and Mitra [23], McKenna [22], Berger et al. [4]. Another possibility is to let n grow to infinity in proportion to the number of stations; Knessl and Tier [19, 20]. Finally, one can also study the network behavior when n grows to infinity keeping fixed the proportions of customers in each route; Pittel [25], Schweitzer [30], Balbo and Serazzi [2], Walton [35], Walton et al. [16], Anselmi and Cremonesi [1], George et al. [11].

The focus of this paper is on the last approach where $n \rightarrow \infty$ keeping fixed the proportions of customers in each route. In this limiting regime, it is known that some queues, called *bottlenecks* in the following, increase their backlog proportionally to n , see Pittel [25] and Walton et al. [16], and uniquely determine the throughput of customers along each route by means of a concave optimization problem; Schweitzer [30], Walton [35]. Interestingly, this optimization problem coincides with the utility optimization problem that determines the fractions of bandwidth (or rates) allocated to multiple classes of concurrent Internet flows (or end-to-end Internet connections); see Kelly et al. [17], Srikant [33]. On the other hand, the amount of backlog in each *non*-bottleneck is strictly bounded by $O(n)$ but in general its limiting behavior is not known. There is numerical evidence to support the conjecture that the limiting stationary distribution of each non-bottleneck queue is geometric; Balbo and Serazzi [2]. Such convergence was proved by Pittel [25] under the assumption that the mode of the stationary distribution is unique; see also Gordon and Newell [13], Lipsky et al. [21], Anselmi and Cremonesi [1]. This assumption, for instance, is not satisfied when the number of bottlenecks is less than the number of routes but greater than one, which is a natural scenario in real networks. As we shall explain, one of the principle contributions of this paper is to prove this conjecture by removing the unimodal assumption.

The above conjecture provides an accurate and efficient approximation to the per-route mean delay (or cycle time) of a network with n customers that takes into account the contribution of non-bottlenecks (a number of techniques only consider the mean delays at the bottlenecks; see, e.g., Schweitzer et al. [31] for an overview). The impact on mean delay of non-bottlenecks becomes non-negligible if their number prevails significantly, and this is often the case; see Casale and Serazzi [7], Anselmi and Cremonesi [1]. This approximation allows for the direct development of efficient optimization frameworks able to address, for instance, data-center consolidation problems, where the objective is to reduce the cost and the size of a data-center while guaranteeing a given performance level. Furthermore, it provides a good initial guess for the iterative or fixed-point algorithms mentioned above.

1.1 Our contribution.

Following the approach considered by a large body of the literature, we are interested in the behavior of closed product-form queueing networks when the number of customers in each route grows to infinity proportionally. This is mainly motivated because real networks are populated by a large number of customers; e.g., Urgaonkar et al. [34]. Our objective is two-fold and consists in analyzing the problem from two contrasting standpoints.

The first part of this paper focuses on the *stationary* behavior of these networks. We prove the conjecture that the stationary distribution of non-bottlenecks converges weakly to the stationary distribution of an ergodic, open product-form queueing network. This open network is obtained by replacing bottlenecks with per-route Poissonian sources whose rates are determined by the solution of a strictly concave optimization problem.

In contrast to the first part, the second part of this paper focuses on the *transient* behavior. Starting from any arbitrary distribution of customers on queues and using justified fluid-limit equations, in our second contribution we prove that the amount of fluid, or customers, on each route eventually concentrates, as time increases, on the bottleneck queues only and that the (long-term) proportions of fluid in each route and in each bottleneck solve the dual of the concave optimization problem that determines the throughputs of the open network described above. Our proof for closed queueing networks uses an entropy Lyapunov function similar to the one used by Bramson [5] to establish convergence properties of the fluid-limit equations of open queueing networks.

The technical difference behind the two contributions above is the order in which the limit in the number of customers and the limit in time are taken. In the second part, the limit in the number of customers is taken *before* the limit in time, and vice versa. In stochastic systems, these two limits do not commute in general, but for the class of queueing networks investigated in this paper we prove that they do. Taking the limit in the number of customers first provides a natural way to look at the evolution of a network populated by a large number of customers and, by subsequently taking a limit in time, we justify fluid model arguments within the queueing literature. The second result proven in this paper, thus, increases the robustness of the approach taken in the first part, which has been followed by several researchers as reference above, and furthermore, it can be also seen as a queueing theoretic analysis of the utility optimization found in congestion control protocols; e.g., Kelly et al. [17] and Srikant [33].

1.2 Organization.

In Section 2, we introduce the models considered in this paper. In particular, these include two Markov descriptions of a closed queueing network, relevant quantities such as bottleneck and non-bottleneck queues are defined, and an expression for a fluid model of a closed queueing network is given. In Section 3, we present the three main results of this paper: Section 3.1 shows the asymptotic independence of non-bottleneck queues in the large-population limit; Section 3.2 shows the convergence of the Markov closed queueing network to a fluid solution; and Section 3.3 states that a fluid solution converges to the set of bottlenecks in a way that minimizes a certain entropy Lyapunov function. In Section 4, we prove the main results stated in Section 3. We respectively prove the results of Sections 3.1, 3.2 and 3.3 in Sections 4.1, 4.2 and 4.3.

2 Closed queueing networks models.

We consider closed, multi-class queueing networks in the sense of Kelly [18] and Baskett et al. [3]. The set $\mathcal{J} \subset \mathbb{N}$ denotes the set of queues (or stations) and we let $J = |\mathcal{J}|$. The set $\mathcal{I} \subset 2^{\mathcal{J}}$ denotes the set of routes (or classes) and we let $I = |\mathcal{I}|$. A route is a sequence of queues visited by a customer during one cycle of the network. We assume, for simplicity, that each customer visits each queue at most once within a cycle of the network. Within each route $i = \{j_1^i, \dots, j_{k_i}^i\}$, we associate a route order $(j_1^i, \dots, j_{k_i}^i)$. For $k = 1, \dots, k_i - 1$, a customer departing queue j_k^i will next join queue j_{k+1}^i and a customer departing queue $j_{k_i}^i$ will join queue j_1^i . Unless otherwise specified, i will be used to index routes and j will be used to index queues. We assume that a constant number of customers circulate along each route of the network. We denote by $n = (n_i : i \in \mathcal{I}) \in \mathbb{N}^I$ the *population vector*, i.e., the total number of customers on each route. When joining queue j , we assume that route- i customers require amounts of service that are independent and exponentially distributed with mean μ_{ji}^{-1} . At each queue, we assume that customers are served at rate 1 according to a processor sharing discipline and customers joining a queue take a position uniformly at random in the queue. Thus, if a route- i customer does not join a queue j , i.e. $j \notin i$, we may assume $\mu_{ji}^{-1} = 0$.

2.1 Two Markov models of closed queueing networks.

Firstly, we could describe the exact location of each customer in the queue according to its route type. Here the *explicit state* of a queue j would be a vector $s_j = (i_j(k) : k = 1, \dots, m_j)$, where m_j is the number of customers in the queue and $i_j(k)$ is the route type of the customer in the k -th position. The explicit state of the network would then be the vector of each queue's state, $s = (s_j : j \in \mathcal{J})$. We then let $\mathcal{X}(n)$ be the set of the explicit states where the number of customers of each route type i is n_i .

Secondly, we could ignore positional information about customers within a queue, and instead, just consider the number of each route type at a queue. Here we let $m = (m_{ji} : j \in \mathcal{J}, i \in \mathcal{I}, j \in i)$ be a network *state*, where m_{ji} represents the number of route i customers in queue j . Thus, $\mathcal{S}(n) = \{m : \sum_j m_{ji} = n_i, i \in \mathcal{I}\}$ is the state space of this Markov chain. It may be verified in a straightforward manner that this state space is finite and irreducible.

Under the above assumptions, it is known (Kelly [18] and Baskett et al. [3]) that the stationary distribution of being in state c , which we denote by $\pi(s|n)$, or being in state m , which we denote $\pi(m|n)$, is respectively

$$\pi(s|n) = \frac{1}{B(n)} \prod_{j \in \mathcal{J}} \prod_{k=1}^{m_j} \mu_{j i_j(k)}^{-1} = \frac{1}{B(n)} \prod_{j \in \mathcal{J}} \prod_{i: j \in i} \mu_{ji}^{-m_{ji}}, \quad s \in \mathcal{X}(n), \quad (2.1)$$

$$\pi(m|n) = \frac{1}{B(n)} \prod_{j \in \mathcal{J}} \left(\binom{m_j}{m_{ji} : i \ni j} \prod_{i: j \in i} \mu_{ji}^{-m_{ji}} \right), \quad m \in \mathcal{S}(n), \quad (2.2)$$

where $B(n)$ is the normalizing constant

$$B(n) \stackrel{\text{def}}{=} \sum_{m \in \mathcal{S}(n)} \prod_{j \in \mathcal{J}} \left(\binom{m_j}{m_{ji} : i \ni j} \prod_{i: j \in i} \mu_{ji}^{-m_{ji}} \right). \quad (2.3)$$

To count the possible orderings of customers inside a queue, we use the multinomial coefficient

$$\binom{m_j}{m_{ji} : i \ni j} \stackrel{\text{def}}{=} \frac{m_j!}{\prod_{i: j \in i} (m_{ji}!)}.$$

In the following, we mainly consider the stationary distribution (2.2), while the expression (2.1) will be used in proofs. We remark that there are a large number of generalizations of our queueing network where (2.2) still gives the stationary number of customers of each route at each queue. Thus, our results generalize to such cases. For instance, we could generalize our processor-sharing discipline to the class of symmetric queueing disciplines with unit service capacity, see Kelly [18, Section 3.3]. We could also generalize service requirements to be independent with mean μ_{ji}^{-1} and having a rational Laplace transform. If we keep the

assumption that service requirements are exponential and we assume that the state space of our Markov description is irreducible, then we could generalize to allow any service discipline that allocates service amongst customers in a way that does not discriminate between the route types of customers at the queue. Such service disciplines are described by Kelly [18, Section 3.1]. Finally, in all cases, we may elaborate the routing structure of our network; we may allow a route i customer to be routed through the network as a Markov chain (with finite expected exit time) whose states are determined by the set of queues so far visited. Although generalizations leading to stationary distribution (2.2) are abundant, for simplicity of exposition, we do not explore these in further detail.

2.2 Throughputs and bottlenecks.

A key quantity of interest is the rate at which customers complete service on each route with respect to a reference queue (say j_1^i for route i). Let

$$\Lambda_i(n) \stackrel{\text{def}}{=} \sum_{\substack{m \in \mathcal{S}(n): \\ m_{j_1^i} > 0}} \mu_{j_1^i} \frac{m_{j_1^i}}{m_{j_1^i}} \pi(m|n) \quad (2.4)$$

be the *throughput* of customers on route i observed at queue j_1^i . In addition, let

$$\Lambda_j(n) \stackrel{\text{def}}{=} \sum_{i: j \in i} \mu_{j_1^i}^{-1} \Lambda_i(n) \quad (2.5)$$

be the *load* (or utilization) at queue j , i.e., the mean amount of work arriving at queue j per unit of time.

Remark 1 The expression (2.4) gives the service completion rate of route i customers at queue j_1^i , i.e., $\mu_{j_1^i}$, times the service rate devoted to these customers at that queue, i.e., $m_{j_1^i}/m_{j_1^i}$, times the stationary probability that there are m customers in the network, i.e., $\pi(m|n)$. Thus, this expression gives the mean rate (or throughput) for which route i customers leave queue j_1^i . Note the route i throughputs at each queue $j \in i$ must be equal, without loss of generality we consider j_1^i .

A more concise expression for the per-route throughput, which is verified in Appendix A, is given in the following lemma.

Lemma 1

$$\Lambda_i(n) = \frac{B(n - e_i)}{B(n)}, \quad i \in \mathcal{I}$$

where $B(n)$ is the normalizing constant (2.3) and e_i is the i^{th} unit vector in \mathbb{R}_+^I .

In agreement with other works (e.g., Anselmi and Cremonesi [1], Balbo and Serazzi [2]), we define a bottleneck queue as a queue whose service is saturated.

Definition 1 Queue $j \in \mathcal{J}$ is called *bottleneck* if and only if

$$\lim_{c \rightarrow \infty} \Lambda_j(cn) = 1. \quad (2.6)$$

We define the set $\bar{\mathcal{J}} \subseteq \mathcal{J}$ to be the set of bottlenecks and let $|\bar{\mathcal{J}}|$. Similarly, we define $\mathcal{J}^\circ = \mathcal{J} \setminus \bar{\mathcal{J}}$ to be the set of non-bottleneck queues and let $J^\circ = |\mathcal{J}^\circ|$.

In other words, a bottleneck is a queue whose load approaches the queue's unit service rate as $c \rightarrow \infty$. We will think of \mathcal{J}° as the open part of the network and of $\bar{\mathcal{J}}$ as the closed part of the network.

2.3 Non-bottleneck queues and open queueing networks.

We are interested in the probability distribution of non-bottleneck queues. For this reason, we consider queue-size vectors $m^\circ = (m_{j_i}^\circ : j \in \mathcal{J}^\circ, i \in \mathcal{I}, j \in i) \in \mathbb{Z}_+^{\mathcal{J}^\circ \times \mathcal{I}}$ and

$$\pi^\circ(m^\circ|n) \stackrel{\text{def}}{=} \sum_{\substack{\bar{m} \in \mathbb{Z}_+^{\bar{\mathcal{J}}}: \\ (\bar{m}, m^\circ) \in \mathcal{S}(n)}} \pi((\bar{m}, m^\circ)|n), \quad j \in \mathcal{J}^\circ, i \in \mathcal{I}, \quad (2.7)$$

which defines the stationary probability that non-bottleneck queues are in state m° .

We also define

$$\pi_\Lambda^\circ(m^\circ) \stackrel{\text{def}}{=} \prod_{j \in \mathcal{J}^\circ} \pi_{j, \Lambda}^\circ(m^\circ), \quad (2.8a)$$

$m^\circ \in \mathbb{Z}_+^{\mathcal{J}^\circ \times \mathcal{I}}$, $\Lambda \in \mathbb{R}_+^I$, where

$$\pi_{j, \Lambda}^\circ(m^\circ) \stackrel{\text{def}}{=} \left(1 - \sum_{i: j \in i} \frac{\Lambda_i}{\mu_{j_i}}\right) \left(\begin{matrix} m_j \\ m_{j_i} : i \ni j \end{matrix} \right) \prod_{i: j \in i} \left(\frac{\Lambda_i}{\mu_{j_i}} \right)^{m_{j_i}}. \quad (2.8b)$$

The distribution π , given by (2.2), refers to the stationary distribution of a closed queueing network. In contrast, the distribution π_Λ° can be shown to be the stationary distribution of an *open* queueing network constructed on queues \mathcal{J}° . Here customers arrive on each route as a Poisson process of rate $\Lambda = (\Lambda_i : i \in \mathcal{I})$ and depart the network after receiving service at each queue on their route $i \in \mathcal{J}^\circ$, see Kelly [18] and Baskett et al. [3].

2.4 Fluid Model.

In order to study the transient behavior of our closed queueing network, we will analyze the following fluid model.

Definition 2 (Closed queueing network fluid model) *The processes $m(t) = (m_{ji}(t) : j \in \mathcal{J}, i \in \mathcal{I}, j \in i)$ and $\Lambda(t) = (\Lambda_{ji}(t) : j \in \mathcal{J}, i \in \mathcal{I}, j \in i)$ form a fluid solution (or fluid limit) of our closed queueing network if they satisfy the following conditions:*

$$m_{j_k^i}(t) = \Lambda_{j_{k-1}^i}(t) - \Lambda_{j_k^i}(t), \quad (2.9a)$$

$$\sum_{i:j \in i} \frac{1}{\mu_{ji}} (\Lambda_{ji}(t) - \Lambda_{ji}(s)) \leq t - s, \quad (2.9b)$$

$$\Lambda_{ji}(t) \text{ is increasing,} \quad (2.9c)$$

$$\text{if } m_j(t) > 0 \text{ then } \frac{d\Lambda_{ji}(t)}{dt} = \frac{m_{ji}}{m_j} \mu_{ji}, \quad (2.9d)$$

$$\sum_{j:j \in i} m_{ji}(t) = n_i. \quad (2.9e)$$

Here, $j \in \mathcal{J}, i \in \mathcal{I}, j \in i, t \geq s \geq 0$. Also, j_k^i is the k^{th} queue on route i and j_{k-1}^i is the queue before the k^{th} queue (we use the convention that the queue before j_1^i is $j_{k_i}^i$).

The conditions (2.9), thus, are the defining properties of a fluid solution and we observe that they are analogous to the ones used by Bramson [5, Formulas (2.3)-(2.6)], which hold for the open versions of the considered closed queueing networks¹. In particular, conditions (2.9a), (2.9c), (2.9e) are basic and relate queue lengths in an obvious manner, condition (2.9d) is the property that defines a processor-sharing discipline, and condition (2.9b) takes into account the maximal processing rate of the system.

We note that the condition (2.9b) implies that Λ_{ji} is Lipschitz continuous. By (2.9a), this is also true for m_{ji} . Lipschitz continuity implies absolute continuity, and therefore the processes Λ_{ji} and m_{ji} are differentiable almost everywhere with respect to the Lebesgue measure. Throughout this document, the term *for almost every* will refer a set of real numbers whose complement has Lebesgue measure zero. Shortly, we will prove that the limit of the closed queueing network described previously satisfies the fluid model (2.9).

3 Results on closed queueing networks.

In this section, we present the three main results of this article. These are: Theorem 1, which states the independence of non-bottlenecks under a large-population limit; Theorem 2, which states that the stochastic process limit of a closed queueing network is a solution to the fluid equations (2.9); Theorem 3, which shows the convergence of the fluid solutions (2.9) to the set of bottleneck queues via a Lyapunov function argument.

3.1 Independence of non-bottleneck queues.

We now consider the limiting behavior of our queueing network with cn customers when $c \rightarrow \infty$. Given the constraints on the number of customers on each route, the distribution of customers at queues are certainly dependent variables; however, in classical open queueing systems, where all customers arrive and depart the network, the stationary distribution of queues can be shown to be independent; see Kelly [18] and Baskett et al. [3]. We wish to demonstrate that non-bottleneck queues become independent as our closed queueing network congests. This fact has been conjectured and used in Balbo and Serazzi [2], and in this paper we present a proof.

In our closed queueing network context and given (2.7) and (2.8), in informal terms we wish to prove that for each $n \in \mathbb{R}_+^I, m \in \mathbb{Z}_+^{J^\circ \times I}$

$$\pi^\circ(m^\circ | cn) \xrightarrow{c \rightarrow \infty} \pi_{\Lambda^\circ}^\circ(m^\circ), \quad (3.1)$$

for some throughput vector $\Lambda \in \mathbb{R}_+^I$ which we must identify and for \mathcal{J}° the set of non-bottleneck queues which must also be identified. Since the vector cn needs not to belong to \mathbb{Z}_+^I , we consider $cn + k_c$ where $\{k_c\}_{c \in \mathbb{N}}$ is any uniformly bounded sequence such that $cn + k_c \in \mathbb{Z}_+^I$. Our theorem can then be written as follows.

Theorem 1 *For $n \in \mathbb{R}_+^I, m \in \mathbb{Z}_+^{J^\circ \times I}$*

$$\pi^\circ(m^\circ | cn + k_c) \xrightarrow{c \rightarrow \infty} \pi_{\Lambda^*(n)}^\circ(m^\circ) \quad (3.2)$$

¹Actually, the open queueing networks defined in Bramson [5], called head-of-the-line processor sharing networks, appear different from ours. In fact, upon completion of service at one queue, a class- i_1 customer becomes of class i_2 with probability $p_{i_1 i_2}$. However, one can easily build a mapping from one representation to the other.

where $(\Lambda_1^*(n), \dots, \Lambda_I^*(n))$ is the unique minimizer of the strictly-concave optimization problem

$$\text{maximize} \quad \sum_{i \in \mathcal{I}} n_i \log \Lambda_i \quad (3.3a)$$

$$\text{subject to} \quad \sum_{i: j \in i} \Lambda_i / \mu_{ji} \leq 1, \quad j \in \mathcal{J} \quad (3.3b)$$

$$\text{over} \quad \Lambda_i \geq 0, \quad i \in \mathcal{I}, \quad (3.3c)$$

and where \mathcal{J}° , the set of non-bottlenecks, is given by the set of queues j such that

$$\sum_{i: j \in i} \frac{\Lambda_i^*(n)}{\mu_{ji}} < 1.$$

The solution of optimization problem (3.3), $\Lambda^*(n)$, is known in the literature as the *proportionally-fair* allocation (see Kelly [15]) and, interestingly, emerged independently as a model for the sharing of bandwidth among Internet connections (e.g., Srikant [33]). For a detailed treatment of the relationship between closed queueing networks and the proportionally-fair allocation, see Walton [35].

3.2 Existence of fluid limits for closed queueing networks.

In Section 3.1, we analyzed the *stationary* probability distribution (2.2) of the closed queueing networks under investigation in the large-population limit. Now, we focus on the *transient* probability distribution in the large-population limit and then study the evolution in time of the system. In other words, the limit in time is now taken before the limit in the number of customers. In stochastic systems it is known that both limits are not interchangeable in general. The fluid limit, see Definition 2, is a natural framework that allows for the analysis of such scenario.

We consider a sequence of the closed queueing networks described in Section 2. In this sequence, the only variables that change are the number of customers on each route (the number of queues, routes, service distributions are kept fixed). We let the vector $n \in \mathbb{R}_+^I$ be the proportion of customers on each route of the network. In the c^{th} network of this sequence of closed queueing networks, there are $cn + k_c$ customers on each route, where $\{k_c\}_{c \in \mathbb{N}}$ is a bounded sequence of variables in \mathbb{R}_+^I such that $cn + k_c \in \mathbb{Z}_+^I$.

We let $M_{ji}^c(t)$ be the number of route- i customers in queue j at time t of the c^{th} closed queueing network. We let $\Lambda_{ji}^c(t)$ give the total number of route- i customers served by queue j by time t in the c^{th} closed queueing network. From this, we define the rescaled processes

$$\bar{M}_{ji}^c(t) \stackrel{\text{def}}{=} \frac{M_{ji}^c(ct)}{c}, \quad \bar{\Lambda}_{ji}^c(t) \stackrel{\text{def}}{=} \frac{\Lambda_{ji}^c(ct)}{c}, \quad j \in \mathcal{J}, \quad i \in \mathcal{I}, \quad j \in i.$$

We wish to show that the vector processes \bar{M}^c and $\bar{\Lambda}^c$ converge to a fluid solution.

Theorem 2 *The sequence of stochastic processes $\{(\bar{M}^c, \bar{\Lambda}^c)\}_{c \in \mathbb{N}}$ converges in probability on uniforms on compact time intervals to a process $(\bar{m}, \bar{\Lambda})$ that satisfies the fluid solution equations (2.9).*

3.3 Convergence of the fluid solution.

Having established the existence of a fluid solution, say (m, Λ) , of our closed queueing networks, now our goal is to study its evolution in the long-term, i.e., $\dot{m}_{ij} \stackrel{\text{def}}{=} dm_{ij}(t)/dt$ when $t \rightarrow \infty$. In our main result, we show that the amount of fluid m eventually concentrates on the bottleneck queues only and that the long-term proportions of fluid in each route and in each queue solve the dual of optimization problem (3.3).

From the stationary distribution $\pi(m|n)$, given by (2.2), and under the same premise of Theorem 1, we can show that

$$\lim_{c \rightarrow \infty} \frac{1}{c} \log \pi(cm|cn) = -\beta(m)$$

where

$$\beta(m) = \sum_{j \in \mathcal{J}} \sum_{i: j \in i} m_{ji} \log \frac{m_{ji} \mu_{ji}}{m_j} \quad (3.4)$$

$$= \sum_{j \in \mathcal{J}} m_j \sum_{i: j \in i} p_{ji} \log \frac{p_{ji}}{\mu_{ij}^{-1}}. \quad (3.5)$$

Here, we have used the notation $p_{ji} \stackrel{\text{def}}{=} \frac{m_{ji}}{m_j}$.

Remark 2 (Relative Entropy) *A key quantity that will be useful in our proofs is the unnormalized relative entropy²*

$$D(p||q) \stackrel{\text{def}}{=} \sum_i p_i \log \frac{p_i}{q_i}, \quad p, q \in \mathbb{R}_+^I.$$

²This entropy is unnormalized because we do not enforce the condition that its arguments are probability distributions.

We notice that the function β , given by (3.5), is a linear combination of these unnormalized relative entropies for each queue. Furthermore, we note that if $\sum_i p_i = \sum_i q_i$, then we can renormalize and treat p and q as probability distributions. One can show that $D(p||q) \geq 0$, using Jensen's inequality, and that $D(p||q)$ is minimized if $p = q$.

Remark 3 It is an interesting observation that the rate of decrease of $\beta(m)$, the entropy of states, will be determined by the relative entropy between the rates of service; see Proposition 7 below. Although distinct, this argument is similar to those argued for Markov processes by Spitzer [32] and more recently by Dupuis and Fischer [10]. A further important reference is Bramson [5], which shows, for the open versions of the closed queueing networks considered in this paper, that the amount of fluid in non-bottlenecks converges to zero.

In Bramson [5], the main argument behind the fluid convergence to zero for non-bottleneck stations relies on closing the open network with an additional (artificial) queue whose service rate equals the overall throughput of the network. All customer routes enter this additional queue and it is assumed that this queue never empties. This plays a significant role in regulating traffic to match that of the open network. So although a closed queueing network appears in that analysis, this additional queue introduces an important loss of generality.

The function $\beta(m)$ forms a natural candidate for a Lyapunov function. We need to show that, for $n \in \mathbb{R}_+^I$ fixed, $\beta(m(t))$ decreases to its minimal value

$$\beta^* \stackrel{\text{def}}{=} \text{minimize } \beta(m) \quad \text{subject to } \sum_{j \in i} m_{ji} = n_i, \quad i \in \mathcal{I}. \quad (3.6)$$

We will consider the set of points attaining this minimal value

$$\mathcal{M} \stackrel{\text{def}}{=} \text{argmin } \beta(m) \quad \text{subject to } \sum_{j \in i} m_{ji} = n_i, \quad i \in \mathcal{I}. \quad (3.7)$$

The following theorem shows how the amount of fluid eventually distributes among network stations.

Theorem 3 For $m(t), \Lambda(t)$ a fluid solution (2.9),

$$\beta(m(t)) \searrow \beta^*, \quad \text{as } t \rightarrow \infty$$

and, moreover,

$$\min_{m^* \in \mathcal{M}} |m(t) - m^*| \rightarrow 0, \quad \text{as } t \rightarrow \infty.$$

In the above theorem, for the norm $|m|$ is the euclidean norm in $bR^{\mathcal{I} \times \mathcal{J}}$.

4 Proofs of main results.

We now focus on proving the main results of this paper, namely, Theorem 1, Theorem 2 and Theorem 3. These are respectively proven in Sections 4.1, 4.2 and 4.3.

4.1 Analysis of non-bottleneck queues.

We develop a proof of Theorem 1. Recalling the informal statement (3.1), we need to verify that for each $n \in \mathbb{R}_+^I, m \in \mathbb{Z}_+^{\mathcal{J} \times \mathcal{I}}$

$$\pi^\circ(m^\circ | cn) \xrightarrow{c \rightarrow \infty} \pi_\Lambda^\circ(m^\circ), \quad (4.1)$$

for some $\Lambda \in \mathbb{R}_+^I$ which we must identify and for \mathcal{J}° the set of non-bottleneck queues which we must identify. Before verifying such a statement, we must identify the relevant throughput vector Λ and non-bottleneck queues \mathcal{J}° . Fortunately, the following result, which was proven by Walton [35], characterizes Λ and \mathcal{J}° .

Proposition 1 ([35])

$$\Lambda_i^*(n) = \lim_{c \rightarrow \infty} \Lambda_i(cn + k_c), \quad i \in \mathcal{I}$$

where $\Lambda_i^*(n)$ is the unique minimizer of the concave optimization problem (3.3) and so that $\Lambda_i(cn + k_c)$ is defined on a point in its domain, $\{k_c\}_{c \in \mathbb{N}}$ is any bounded sequence in \mathbb{R}_+^I such that $cn + k_c \in \mathbb{Z}_+^I$.

Consequently, $j \in \mathcal{J}^\circ$ iff

$$\sum_{i: j \in i} \frac{\Lambda_i^*(n)}{\mu_{ji}} < 1.$$

Now, we show that (3.2) holds. Before proceeding with this proof, we introduce a specific closed queueing network, which will help us in the proof. Recall the closed queueing network defined on set \mathcal{J} introduced in Section 2. Consider a queueing network defined exactly as in Section 2, except that queues \mathcal{J}° are removed. The resulting queueing network has states $\bar{m} = (\bar{m}_{ji} : j \in \bar{\mathcal{J}}, i \in \mathcal{I}, j \in i) \in \mathbb{Z}_+^{\bar{\mathcal{J}} \times \mathcal{I}}$; if there are

$n \in \mathbb{Z}_+^I$ customers on each route, this network has state space $\bar{\mathcal{S}}(n) = \{\bar{m} : \sum_j \bar{m}_{ji} = n_i, i \in \mathcal{I}\}$, stationary distribution

$$\bar{\pi}(\bar{m}|n) = \frac{1}{\bar{B}(n)} \prod_{j \in \bar{\mathcal{J}}} \left(\binom{\bar{m}_j}{\bar{m}_{ji} : j \in i} \prod_{i:j \in i} \mu_{ji}^{-\bar{m}_{ji}} \right), \quad \bar{m} \in \bar{\mathcal{S}}(n), \quad (4.2)$$

where

$$\bar{B}(n) = \sum_{\bar{m} \in \bar{\mathcal{S}}(n)} \prod_{j \in \bar{\mathcal{J}}} \left(\binom{\bar{m}_j}{\bar{m}_{ji} : j \in i} \prod_{i:j \in i} \mu_{ji}^{-\bar{m}_{ji}} \right),$$

and stationary throughput

$$\bar{\Lambda}(n) = \frac{\bar{B}(n - e_i)}{\bar{B}(n)}.$$

Proposition 1 holds for this network and, here, we would consider $\bar{\Lambda}^*(n)$, the solution to the optimization

$$\text{maximize } \sum_{i \in \mathcal{I}} n_i \log \Lambda_i \quad \text{subject to } \sum_{i:j \in i} \Lambda_i / \mu_{ji} \leq 1, \quad j \in \bar{\mathcal{J}} \quad \text{over } \Lambda_i \geq 0, \quad i \in \mathcal{I}.$$

In this optimization, all constraints that are not relevant to our solution $\Lambda^*(n)$ are removed. Thus, it is not surprising that the following lemma holds.

Lemma 2 $\bar{\Lambda}^*(n) = \Lambda^*(n)$.

We prove Lemma 2 in Appendix A. A direct consequence of Lemma 2 and Proposition 1 is the following.

Lemma 3

$$\bar{\Lambda}_i(n) = \frac{\bar{B}(n - e_i)}{\bar{B}(n)} \xrightarrow{c \rightarrow \infty} \Lambda_i^*(n), \quad i \in \mathcal{I}.$$

We can now proceed to demonstrate that (3.2) holds. Let us consider the equilibrium distribution $\pi^\circ(m^\circ|n)$. We have

$$\begin{aligned} \pi^\circ(m^\circ|n) &= \sum_{\substack{\bar{m} \in \mathbb{Z}_+^{\bar{\mathcal{J}}}: \\ (\bar{m}, m^\circ) \in \mathcal{S}(n)}} \pi((\bar{m}, m^\circ)|n) \\ &= \frac{1}{B(n)} \times \prod_{j \in \mathcal{J}^\circ} \left(\binom{m_j^\circ}{m_{ji}^\circ : j \in i} \prod_{i:j \in i} \mu_{ji}^{-m_{ji}^\circ} \right) \times \sum_{\substack{\bar{m} \in \mathbb{Z}_+^{\bar{\mathcal{J}}}: \\ (\bar{m}, m^\circ) \in \mathcal{S}(n)}} \prod_{j \in \mathcal{J}} \left(\binom{\bar{m}_j}{\bar{m}_{ji} : j \in i} \prod_{i:j \in i} \mu_{ji}^{-\bar{m}_{ji}} \right) \\ &= \frac{1}{B(n)} \times \prod_{j \in \mathcal{J}^\circ} \left(\binom{m_j^\circ}{m_{ji}^\circ : j \in i} \prod_{i:j \in i} \mu_{ji}^{-m_{ji}^\circ} \right) \times \sum_{\bar{m} \in \bar{\mathcal{S}}(n - n^\circ)} \prod_{j \in \mathcal{J}} \left(\binom{\bar{m}_j}{\bar{m}_{ji} : j \in i} \prod_{i:j \in i} \mu_{ji}^{-\bar{m}_{ji}} \right) \\ &= \underbrace{\frac{\bar{B}(n)}{B(n)}}_{(c)} \times \underbrace{\frac{\bar{B}(n - n^\circ)}{\bar{B}(n)}}_{(b)} \times \underbrace{\prod_{j \in \mathcal{J}^\circ} \left(\binom{m_j^\circ}{m_{ji}^\circ : j \in i} \prod_{i:j \in i} \mu_{ji}^{-m_{ji}^\circ} \right)}_{(a)}. \end{aligned} \quad (4.3)$$

Here, $n^\circ = (n_i^\circ : i \in \mathcal{I})$ is the number of route- i customers in non-bottleneck queues, i.e. $n_i^\circ = \sum_{j \in i \cap \mathcal{J}^\circ} m_{ji}^\circ$. The third equality above follows by observing that the summation is over all the states where the number of non-bottleneck customers is $n - n^\circ$.

We now consider how the terms (a), (b) and (c), converge as we keep m° fixed and let n increase. Term (a) is easily dealt with as it does not depend on n . Term (b) will be shown to converge in the next proposition. Subsequently, term (c) will take a more in depth analysis.

Term (a) represents the correct expression for the unnormalized stationary distribution of our open queueing network (see (2.8)), except that we do not include the multiplicative term

$$\prod_{j \in \mathcal{J}^\circ} \prod_{i:j \in i} \Lambda_i^*(n)^{m_{ji}^\circ} = \prod_{i \in \mathcal{I}} \Lambda_i^*(n)^{n_i^\circ}.$$

As the following proposition shows, this is the limit of the term (b).

Proposition 2 For $n^\circ \in \mathbb{Z}_+^I$, $n \in \mathbb{R}_+^I$ and $\{k_c\}_{c \in \mathbb{N}}$ some bounded sequence such that $cn + k_c \in \mathbb{Z}_+^I$

$$\frac{B(cn + k_c - n^\circ)}{B(cn + k_c)} \xrightarrow{c \rightarrow \infty} \prod_{i \in \mathcal{I}} \Lambda_i^*(n)^{n_i^\circ}.$$

Proof: From Proposition 1, we have that

$$\frac{B(cn + k'_c - e_i)}{B(cn + k'_c)} \xrightarrow{c \rightarrow \infty} \Lambda_i^*(n), \quad (4.4)$$

for any $n \in \mathbb{R}_+^I$ and any bounded sequence $\{k'_c\}_{c \in \mathbb{N}}$ such that $cn + k'_c \in \mathbb{Z}_+^I$. Let $K = \sum_i n_i^\circ$, Let $e_{i(1)}, \dots, e_{i(K)}$ be a finite sequence of unit vectors and $n(0), \dots, n(K)$ be a sequence of vectors in \mathbb{Z}_+^I such that

$$\begin{aligned} n(0) &= 0, \quad n(K) = n^\circ \\ n(k) &= n(k-1) + e_{i(k)}, \quad k = 1, \dots, K. \end{aligned}$$

Applying (4.4), we have that

$$\frac{B(cn + k_c - n^\circ)}{B(cn + k_c)} = \prod_{k=0}^{K-1} \frac{B(cn + k_c - n(k) - e_{i(k+1)})}{B(cn + k_c - n(k))} \xrightarrow{c \rightarrow \infty} \prod_{k=1}^K \Lambda_{i(k)}^*(n) = \prod_{i \in \mathcal{I}} \Lambda_i^*(n)^{n_i^\circ},$$

as required. ■

As the above proposition holds for any closed queueing network and because Lemma 3 holds, we can say that

$$\frac{\bar{B}(cn + k_c - n^\circ)}{\bar{B}(cn + k_c)} \xrightarrow{c \rightarrow \infty} \prod_{i \in \mathcal{I}} \Lambda_i^*(n)^{n_i^\circ} = \prod_{j \in \mathcal{J}^\circ} \prod_{i: j \in i} \Lambda_i^*(n)^{m_{ji}^\circ}. \quad (4.5)$$

We now study term (c) in expression (4.3). We can note that this term is exactly the probability that all non-bottleneck queues are empty.

Lemma 4

$$\pi(\{m_j = 0, \forall j \in \mathcal{J}^\circ\} | n) = \frac{\bar{B}(n)}{B(n)}.$$

Proof: We have the following equations:

$$\begin{aligned} \pi(\{m_j = 0, \forall j \in \mathcal{J}^\circ\} | n) &= \sum_{\substack{m \in \mathcal{S}(n): \\ m_j = 0, j \in \mathcal{J}^\circ}} \frac{1}{B(n)} \prod_{j \in \mathcal{J}} \left(\binom{m_j}{m_{ji} : j \in i} \prod_{i: j \in i} \mu_{ji}^{-m_{ji}} \right) \\ &= \sum_{(0, \bar{m}) \in \mathcal{S}(n)} \frac{1}{B(n)} \prod_{j \in \bar{\mathcal{J}}} \left(\binom{\bar{m}_j}{\bar{m}_{ji} : j \in i} \prod_{i: j \in i} \mu_{ji}^{-\bar{m}_{ji}} \right) \\ &= \sum_{\bar{m} \in \mathcal{S}(n)} \frac{1}{B(n)} \prod_{j \in \bar{\mathcal{J}}} \left(\binom{\bar{m}_j}{\bar{m}_{ji} : j \in i} \prod_{i: j \in i} \mu_{ji}^{-\bar{m}_{ji}} \right) = \frac{\bar{B}(n)}{B(n)}. \end{aligned}$$
■

Although it is difficult to directly deal with events of the form $\{m_j = 0, \forall j \in \mathcal{J}^\circ\}$, we can deal with events of the form $\{m_j > 0, \forall j \in \mathcal{J}'\}$ where $\mathcal{J}' \subset \mathcal{J}^\circ$.

Lemma 5 For $\mathcal{J}' \subset \mathcal{J}^\circ$,

$$\pi(\{m_j > 0, \forall j \in \mathcal{J}'\} | cn + k_c) \xrightarrow{c \rightarrow \infty} \prod_{j \in \mathcal{J}'} \Lambda_j^*(n)$$

where we define

$$\Lambda_j^*(n) \stackrel{\text{def}}{=} \sum_{i: j \in i} \frac{\Lambda_i^*(n)}{\mu_{ji}}.$$

Proof: To prove this lemma, we consider the explicit stationary distribution of a closed multi-class queueing network (2.1).

Recall $s = (s_j : j \in \mathcal{J})$ where $s_j = (i_j(k) : k = 1, \dots, m_j) \in \mathcal{I}^{m_j}$ keeps track of the exact position and route type of each customer within a queue. Within this state representation, we can calculate the probability that the customer at the head of each queue $j \in \mathcal{J}'$ is from a specific route type $r(j) \in \mathcal{I}$. In particular, if we let $r(j)$ be the route type of the customer at the head of queue $j \in \mathcal{J}'$ and we let the vector n' give the number of customers of each route type at the heads of these queues then we can see that

$$\begin{aligned} \pi(\{i_j(1) = r(j), j \in \mathcal{J}'\} | n) &= \sum_{\substack{c \in \mathcal{X}(n): \\ i_j(1) = r(j), j \in \mathcal{J}'}} \frac{1}{B(n)} \prod_{j \in \mathcal{J}} \prod_{i: j \in i} \mu_{ji}^{-m_{ji}} \\ &= \left[\prod_{j \in \mathcal{J}'} \frac{1}{\mu_{jr(j)}} \right] \times \sum_{c' \in \mathcal{X}(n-n')} \frac{1}{B(n)} \prod_{j \in \mathcal{J}} \prod_{i: j \in i} \mu_{ji}^{-m'_{ji}} \\ &= \left[\prod_{j \in \mathcal{J}'} \frac{1}{\mu_{jr(j)}} \right] \times \frac{B(n-n')}{B(n)}. \end{aligned}$$

In the second inequality above, we factor out the multiplicative terms corresponding to the heads of each queue in \mathcal{J}' . We then notice the remaining states that must be summed over are all the states where there are $n - n'$ customers on each route. The resulting sum then gives the normalizing constant when there are $n - n'$ customers on each route.

Next, by Proposition 2, we have that

$$\pi(\{i_j(1) = r(j), j \in \mathcal{J}'\} | n) = \left[\prod_{j \in \mathcal{J}'} \frac{1}{\mu_{jr(j)}} \right] \times \frac{B(n - n')}{B(n)} \xrightarrow{c \rightarrow \infty} \prod_{j \in \mathcal{J}'} \frac{\Lambda_{r(j)}^*(n)}{\mu_{jr(j)}}. \quad (4.6)$$

The events $\{i_j(1) = r(j), j \in \mathcal{J}'\}$ are disjoint for different choices of $r' = (r(j) : j \in \mathcal{J}')$. For a queue to be nonempty, there must be some customer at its head. Thus,

$$\bigcup_{r' \in \mathcal{I}^{\mathcal{J}'}} \{i_j(1) = r(j), j \in \mathcal{J}'\} = \{m_j > 0, j \in \mathcal{J}\},$$

and consequently using (4.6), we have

$$\begin{aligned} \pi(\{m_j > 0, j \in \mathcal{J}\} | n) &= \sum_{r' \in \mathcal{I}^{\mathcal{J}'}} \pi(\{i_j(1) = r(j), j \in \mathcal{J}'\} | n) \\ &= \sum_{r' \in \mathcal{I}^{\mathcal{J}'}} \left[\prod_{j \in \mathcal{J}'} \frac{1}{\mu_{jr(j)}} \right] \times \frac{B(n - n')}{B(n)} \\ &\xrightarrow{c \rightarrow \infty} \sum_{r' \in \mathcal{I}^{\mathcal{J}'}} \prod_{j \in \mathcal{J}'} \left(\frac{\Lambda_{r(j)}^*(n)}{\mu_{jr(j)}} \right) = \prod_{j \in \mathcal{J}'} \left(\sum_{r \in \mathcal{I}} \frac{\Lambda_r^*(n)}{\mu_{jr}} \right) = \prod_{j \in \mathcal{J}'} \Lambda_j^*(n) \end{aligned}$$

as required. ■

Now we are in the position to prove the convergence of term (c) in expression (4.3).

Proposition 3

$$\frac{\bar{B}(cn + k_c)}{B(cn + k_c)} \xrightarrow{c \rightarrow \infty} \prod_{j \in \mathcal{J}^\circ} (1 - \Lambda_j^*(n))$$

Proof: In the following expression, we use Lemma 4; we apply the Inclusion-Exclusion Principle (4.8); we apply Lemma 5 (4.9); and then we notice the resulting summation is $\prod_{j \in \mathcal{J}^\circ} (1 - \Lambda_j^*(n))$ expanded. Formally, we obtain

$$\frac{\bar{B}(cn + k_c)}{B(cn + k_c)} = \pi(\{m_j = 0, \forall j \in \mathcal{J}^\circ\} | cn + k_c) \quad (4.7)$$

$$\begin{aligned} &= 1 - \pi\left(\bigcup_{j \in \mathcal{J}^\circ} \{m_j > 0\} \mid cn + k_c\right) \\ &= \sum_{k=0}^{J^\circ} \sum_{\substack{j_1, \dots, j_k \in \mathcal{J}^\circ: \\ j_1 < \dots < j_k}} (-1)^k \pi(\{m_j > 0, j = j_1, \dots, j_k\} | cn + k_c) \end{aligned} \quad (4.8)$$

$$\xrightarrow{c \rightarrow \infty} \sum_{k=0}^{J^\circ} \sum_{\substack{j_1, \dots, j_k \in \mathcal{J}^\circ: \\ j_1 < \dots < j_k}} (-1)^k \prod_{j=j_1, \dots, j_k} \Lambda_j^*(n) = \prod_{j \in \mathcal{J}^\circ} (1 - \Lambda_j^*(n)). \quad (4.9)$$

In expression (4.8), the $k = 0$ summand is understood to equal 1. ■

We have now discovered the limiting behavior of terms (a), (b), and (c), and we can prove Theorem 1.

Proof of Theorem 1: The result holds by applying Proposition 2 and equality (4.5), which we derived from Proposition 3, to our derived equality (4.3):

$$\begin{aligned} \pi^\circ(m^\circ | cn + k_c) &= \frac{\bar{B}(cn + k_c)}{B(cn + k_c)} \times \frac{\bar{B}(cn + k_c - n^\circ)}{\bar{B}(cn + k_c)} \times \prod_{j \in \mathcal{J}^\circ} \left(\left(\frac{m_j^\circ}{m_{j_i}^\circ : j \in i} \right) \prod_{i: j \in i} \frac{1}{\mu_{ji}^\circ} \right) \\ &\xrightarrow{c \rightarrow \infty} \prod_{j \in \mathcal{J}^\circ} (1 - \Lambda_j^*(n)) \times \prod_{j \in \mathcal{J}^\circ} \prod_{i: j \in i} \Lambda_i^*(n)^{m_{j_i}^\circ} \times \prod_{j \in \mathcal{J}^\circ} \left(\left(\frac{m_j^\circ}{m_{j_i}^\circ : j \in i} \right) \prod_{i: j \in i} \left(\frac{\Lambda_i(n)}{\mu_{ji}} \right)^{m_{j_i}^\circ} \right) \\ &= \prod_{j \in \mathcal{J}^\circ} (1 - \Lambda_j^*(n)) \times \prod_{j \in \mathcal{J}^\circ} \left(\left(\frac{m_j^\circ}{m_{j_i}^\circ : j \in i} \right) \prod_{i: j \in i} \left(\frac{\Lambda_i(n)}{\mu_{ji}} \right)^{m_{j_i}^\circ} \right) \\ &= \pi_{\Lambda^*(n)}^\circ(m^\circ). \end{aligned}$$

This proves that our network's non-bottleneck queues become independent. ■

4.2 Proof of fluid limit.

Let $\mathcal{I}(j) \stackrel{\text{def}}{=} \{i \in \mathcal{I} : j \in i\}$ and $I(j) \stackrel{\text{def}}{=} |\mathcal{I}(j)|$. Assuming that set $\mathcal{I}(j)$ is ordered, we also denote by $i_j(l)$, with $l = 1, \dots, I(j)$, the function that enumerates its elements. We define for each $j \in \mathcal{J}$, an independent Poisson marked point process N_j with intensity $\mu_j dt \otimes du$ on $\mathbb{R} \times [0, 1]$, where $\mu_j = \max\{i \in \mathcal{I}(j) : \mu_{ji}\}$ and the function $\chi_j(m_j, u)$ on $\mathbb{N}^{I(j)} \times (0, 1)$ in the following way:

$$\chi_j(m_j, u) = \begin{cases} 0 & \text{if } m_j = 0 \\ i_j(l) & \text{if } \sum_{h=0}^{l-1} m_{jh} \mu_{jh} \leq m_j \mu_j \times u \leq \sum_{h=0}^l m_{jh} \mu_{jh} \\ 0 & \text{if } \sum_{h=0}^{I(j)} m_{jh} \mu_{jh} \leq m_j \mu_j \times u \leq m_j \mu_j. \end{cases}$$

Therefore, the network process can be written as

$$dM_{ji}(t) = 1\{\chi_{j'}(M_j'(t-), U_{N_{j'}(t)}) = i\} dN_{j'}(t) - 1\{\chi_j(M_j(t-), U_{N_j(t)}) = i\} dN_j(t)$$

where j' denotes the queue just before j on route i . In the following we will use also the notations \hat{j} and \hat{j}' to denote the first queue on route i before and including the queues j and j' respectively that are non empty. In integral form, we have

$$M_{ji}(t) = M_{ji}(0) + \mathcal{L}_{j'i}(t) - \mathcal{L}_{ji}(t) + \Lambda_{j'i}(t) - \Lambda_{ji}(t)$$

where $\mathcal{L}_{ji}^1(t)$ is the martingale

$$\mathcal{L}_{ji}(t) = \int_0^t \int_0^1 1\{\chi_j(M_j(s-), u) = i\} [dN_j(s, u) - \mu_j du ds],$$

and the process $\Lambda_{ji}(t)$ is given by

$$\Lambda_{ji}(t) = \int_0^t \int_0^1 \mu_j 1\{\chi_j(M_j(s-), u) = i\} du ds.$$

From the equation above, noticing that $\int_0^1 1\{\chi_j(M_j(s, u) = i\} du = (\mu_{ji} M_{ji}(s))/(\mu_j M_j(s))$ and with the assumption that $0/0 = 0$, we get that

$$\frac{1}{\mu_{ji}} \Lambda_{ji}(t) = \int_0^t \frac{\mu_j}{\mu_{ji}} \int_0^1 1\{\chi_j(M_j(s-), u) = i\} du ds = \int_0^t \frac{M_{jh}(s-)}{M_j(s-)} ds.$$

Summing over $i : j \in i$ we have that for $t' \leq t''$

$$\sum_{i:j \in i} \frac{1}{\mu_{ji}} (\Lambda_{ji}(t'') - \Lambda_{ji}(t')) = \int_{t'}^{t''} \sum_{i:j \in i} \frac{M_{jh}(s-)}{M_j(s-)} ds = \int_{t'}^{t''} \sum_{i:j \in i} 1\{M_j(s-) \neq 0\} ds \leq t'' - t', \quad (4.10)$$

which gives the Lipschitz condition for the process $\Lambda_{ji}(t)$.

We define the scaled process $\bar{M}(nc, t) = c^{-1} M(nc, ct)$, such that, having $M_{ji}(0) = cn_{ji}$, it has the ji -component given by

$$\bar{M}_{ji}(nc, t) = n_{ji} + \frac{\mathcal{L}_{j'i}(ct) - \mathcal{L}_{ji}(ct)}{c} + \frac{L_{j'i}(ct) - L_{ji}(ct)}{c}.$$

Using the same steps as before, we can rewrite the process $c^{-1} \Lambda_{ji}(ct)$ in the following way

$$\begin{aligned} \frac{\Lambda_{ji}(ct)}{c} &= \frac{1}{c} \int_0^{ct} \int_0^1 \mu_j 1\{\chi_j(M_j(s), u) = i\} du ds \\ &= \int_0^t \int_0^1 \mu_j 1\{\chi_j(M_j(cs), u) = i\} du ds = \int_0^t \frac{M_{ji}(cs)}{M_j(cs)} \mu_{ji} ds. \end{aligned}$$

Proposition 4 *Given $n \in \mathbb{N}^{J \times I}$, the processes $\{\bar{M}(nc, t), t > 0\}_{c>0}$, with $M(nc, 0) = nc$, are tight.*

Proof: By the triangular inequalities for metrics, to prove tightness of the multidimensional process $\{\bar{M}(nc, t), t > 0\}_{c>0}$, it is enough to show that any of its coordinate processes are tight. By Theorem C.9 in Robert [28], this can be done by showing that given ij , for any fixed $T, \eta > 0$, there exists $\delta > 0$ such that

$$\Pr\{\omega_{\bar{M}_{ji}(nc, \cdot)}(\delta) > \eta\} < \epsilon \quad (4.11)$$

where, for a given function f , the modulus of continuity on $[0, T]$ is defined as

$$\omega_f(\delta) = \sup\{s, t \leq T, |t - s| < \delta : |f(t) - f(s)|\}.$$

Since $\bar{M}_{ji}(nc, t) = n + c^{-1} \Delta \mathcal{L}_{ji}(ct) + c^{-1} \Delta \Lambda_{ji}(ct) - (\mathcal{L}_{ji}(ct))$ with $\Delta \mathcal{L}_{ji}(t) = \mathcal{L}_{j'i}(t) - \mathcal{L}_{ji}(t)$ and $\Delta \Lambda_{ji}(t) = \Lambda_{j'i}(t) - \Lambda_{ji}(t)$, it is enough to prove that relation (4.11) is valid separately for the processes $c^{-1} \Delta \mathcal{L}_{ji}(ct)$ and $c^{-1} \Delta \Lambda_{ji}(ct)$. We have

$$\begin{aligned} & \Pr \left\{ \sup_{s, t \leq T; |t-s| < \delta} \left| \frac{\Delta \Lambda_{ij}(ct)}{c} - \frac{\Delta \Lambda_{ij}(cs)}{c} \right| > \eta \right\} \tag{4.12} \\ &= \Pr \left\{ \sup_{s, t \leq T; |t-s| < \delta} \left| \int_s^t \left(\frac{M_{j'i}(cu)}{M_{j'}(cu)} \mu_{j'i} - \frac{M_{ji}(cu)}{M_j(cu)} \mu_{ji} \right) du \right| > \eta \right\} \\ &\leq \Pr \left\{ \sup_{s, t \leq T; |t-s| < \delta} \int_s^t \left| \frac{M_{j'i}(cu)}{M_{j'}(cu)} \mu_{j'i} - \frac{M_{ji}(cu)}{M_j(cu)} \mu_{ji} \right| du > \eta \right\} \\ &\leq \Pr \left\{ \sup_{s, t \leq T; |t-s| < \delta} 2|t-s|\mu > \eta \right\} = 0 \tag{4.13} \end{aligned}$$

with $\delta < \eta/(2\mu)$ and where $\mu = \max\{i, j : \mu_{ji}\}$.

For the martingale $c^{-1} \Delta \mathcal{L}_{ji}(ct)$, we have that

$$\Pr \left\{ \sup_{s, t \leq T; |t-s| < \delta} \left| \frac{\Delta \mathcal{L}_{ji}(ct)}{c} - \frac{\Delta \mathcal{L}_{ji}(cs)}{c} \right| > \eta \right\} = \Pr \left\{ \sup_{t \leq T} \frac{\Delta \mathcal{L}_{ji}(ct)}{c} > \frac{\eta}{2} \right\} \leq \frac{4}{c^2 \eta^2} \mathbb{E} [(\Delta \mathcal{L}_{ji}(cT))^2]$$

where the last inequality follows by applying the Doob's inequality. Having that $\mathbb{E} [(\Delta \mathcal{L}_{ji}(cT))^2] \leq 2\mu cT$ we have that for $c > 8\mu T/(\epsilon \eta^2)$ the probability is bounded above by ϵ , as required. \blacksquare

The tightness property ensures the relative compactness, therefore from every sequence $\{(\bar{M}(nc, t), t > 0)\}_{c>0}$, with $M(nc, 0) = nc$ and n fixed, it is possible to extract a convergent subsequence. The following proposition ensures that any limit process will be given by a fluid solution.

Proposition 5 *Assume that a sequence of processes $\{(\bar{M}(nc, t), t > 0)\}_{c>0}$ converges to a limit process $m(t)$ as $c \rightarrow \infty$. Then, $m(t)$ is almost surely continuous and it is a fluid solution, i.e., it satisfies the following conditions*

$$m_{ji}(t) = \Lambda_{j'i}^*(t) - \Lambda_{ji}^*(t), \tag{4.14a}$$

$$\sum_{i:j \in i} \frac{1}{\mu_{ji}} (\Lambda_{j'i}^*(t) - \Lambda_{ji}^*(s)) \leq t - s, \tag{4.14b}$$

$$\Lambda_{ji}^*(t) \text{ is increasing,} \tag{4.14c}$$

$$\text{if } m_j > 0 \text{ then } \frac{d\Lambda_{ji}^*(t)}{dt} = \frac{m_{ji}}{m_j} \mu_{ji}, \tag{4.14d}$$

$$\sum_{j:i \in j} m_{ji}(t) = n_i. \tag{4.14e}$$

Proof: Using the Skorohod's Representation theorem, see Robert [28, Theorem C.8], we can assume that all the elements of the sequence are random processes defined on the same probability space with probability \mathbb{P} and the convergence is \mathbb{P} -a.s.

Since we have that for any c

$$\bar{M}_{ji}(nc, t) = n_{ji} + \frac{\Delta \mathcal{L}_{ji}(ct)}{c} + \frac{\Lambda_{j'i}(ct)}{c} - \frac{\Lambda_{ji}(ct)}{c},$$

passing to the limit and using bounded convergence for the integral we get

$$\bar{m}_{ji}(t) = n_{ji} + \Lambda_{j'i}^*(t) - \Lambda_{ji}^*(t)$$

where

$$\Lambda_{ji}^*(t) = \lim_{c \rightarrow \infty} \int_0^t \frac{M_{ji}(cu)}{M_j(cu)} \mu_{ji} du,$$

which exists by (4.12). In particular, if $m_j(t) > 0$ for some t , it will be positive in a neighborhood $B(t)$ of t by continuity. It follows that for $t' < t''$ and $t', t'' \in B(t)$,

$$\Lambda_{ji}^*(t'') - \Lambda_{ji}^*(t') = \lim_{c \rightarrow \infty} \int_{t'}^{t''} \frac{M_{ji}(cu)}{M_j(cu)} \mu_{ji} du = \int_{t'}^{t''} \lim_{c \rightarrow \infty} \frac{M_{ji}(cu)}{M_j(cu)} \mu_{ji} du = \int_{t'}^{t''} \frac{m_{ji}(u)}{m_j(u)} \mu_{ji} du,$$

where in the second equality we have used the bounded convergence theorem, which implies

$$\partial_t \Lambda_{ji}^*(t) = \mu_{ji} \frac{m_{ji}(t)}{m_j(t)} \quad \text{as } m_j(t) > 0.$$

The additional conditions satisfied by $\Lambda_{ji}^*(t)$ easily follow by the property of the approximation processes $\{c^{-1} \Lambda_{ji}(ct), c > 0\}$, in particular the Lipschitz condition follows by (4.10). \blacksquare

4.3 Proof of convergence of bottleneck queues.

We now focus on proving Theorem 3. Our proof is quite involved and requires a number of lemmas.

For a fluid solution $m(t), \Lambda(t)$, let $\Lambda'_{ji}(t)$ be the derivative of $\Lambda_{ji}(t)$, when it exists. Let also l_{ji} be the queue before queue j on route i and x_{ji} be the next queue after queue j on route i .

The following proposition and lemma helps us differentiate our Lyapunov function.

Proposition 6 *For any time interval $[t_0, t]$ with $t > t_0 > 0$, there exists a constant $D > 0$ such that for any $t_1, t_2 \in [t_0, t]$*

$$|\beta(t_2) - \beta(t_1)| \leq D|t_2 - t_1|. \quad (4.15)$$

Lemma 6 *For a function $x : \mathbb{R}_+ \rightarrow \mathbb{R}_+$, if $t > 0$ is such that the derivative of x exists at t , the derivative of $x \log x$ exists at t and $x(t) = 0$ then*

$$\frac{dx}{dt} = 0 \quad \text{and} \quad \frac{dx \log x}{dt} = 0.$$

Proposition 6 is based on Proposition 4.2 of Bramson [5]. Both Proposition 6 and Lemma 6 are proven in Appendix B. Recalling that x_{ji} is the next queue on route i after j and that l_{ji} is the queue before queue j on route i , we can now prove the following proposition.

Proposition 7 *For almost every t ,*

$$\frac{d\beta}{dt} = - \sum_{i \in \mathcal{I}} \sum_{j \in i} \Lambda'_{ji}(t) \log \frac{\Lambda'_{ji}(t)}{\Lambda'_{x_{ji}i}(t)}. \quad (4.16)$$

Proof: Proof Our processes in Proposition 6 are absolutely continuous and thus almost everywhere differentiable. So for almost every t , we can differentiate the terms $\Lambda_{ji}(t)$, $m_{ij}(t)$, $m_j(t)$, $m_{ij}(t) \log m_{ji}(t)$, $m_j(t) \log m_j(t)$ and $\beta(m(t))$. We now differentiate the function

$$\beta(m(t)) = \sum_{j \in \mathcal{J}} \sum_{i: j \in i} \left(m_{ji}(t) \log m_{ji}(t) - m_{ji} \log \mu_{ji} \right) - \sum_{j \in \mathcal{J}} m_j(t) \log m_j(t).$$

We obtain

$$\frac{d\beta(m(t))}{dt} = \sum_{j \in \mathcal{J}} \sum_{i: j \in i} \left(\frac{dm_{ji}}{dt} \log m_{ji} - \frac{dm_{ji}}{dt} \log \mu_{ji} \right) - \sum_{j \in \mathcal{J}} \frac{dm_j}{dt} \log m_j \quad (4.17a)$$

$$= \sum_{j \in \mathcal{J}} \sum_{\substack{i: j \in i \\ m_{ji}(t) > 0}} \left(\frac{dm_{ji}}{dt} \log m_{ji} - \frac{dm_{ji}}{dt} \log \mu_{ji} \right) - \sum_{\substack{j \in \mathcal{J}: \\ m_j(t) > 0}} \frac{dm_j}{dt} \log m_j \quad (4.17b)$$

$$= \sum_{j \in \mathcal{J}} \sum_{\substack{i: j \in i \\ m_{ji}(t) > 0}} \left(\frac{dm_{ji}}{dt} \log m_{ji} + \frac{dm_{ji}}{dt} - \frac{dm_{ji}}{dt} \log \mu_{ji} \right) - \sum_{\substack{j \in \mathcal{J}: \\ m_j(t) > 0}} \left(\frac{dm_j}{dt} \log m_j + \frac{dm_j}{dt} \right) \quad (4.17c)$$

$$= \sum_{j \in \mathcal{J}} \sum_{\substack{i: j \in i \\ m_{ji}(t) > 0}} \frac{dm_{ji}}{dt} \log \frac{m_{ji}(t) \mu_{ji}}{m_j(t)} \quad (4.17d)$$

$$= \sum_{i \in \mathcal{I}} \sum_{j \in i} \left(\Lambda'_{jii}(t) - \Lambda'_{ji}(t) \right) \log \Lambda'_{ji}(t) \quad (4.17e)$$

$$= \sum_{i \in \mathcal{I}} \sum_{j \in i} \Lambda'_{ji}(t) \left(\log \Lambda'_{x_{ji}i}(t) - \log \Lambda'_{ji}(t) \right) \quad (4.17f)$$

$$= - \sum_{i \in \mathcal{I}} \sum_{j \in i} \Lambda'_{ji}(t) \log \frac{\Lambda'_{ji}(t)}{\Lambda'_{x_{ji}i}(t)}. \quad (4.17g)$$

In the above sequence of equalities, equality (4.17b) holds by Lemma 6. Equality (4.17d) holds by observing that $\sum_{i: j \in i} m_{ji} = m_j$ and canceling terms. For equality (4.17e), we know by our fluid model assumption (2.9a) that $m'_{ji}(t) = \Lambda'_{jii}(t) - \Lambda'_{ji}(t)$. In addition, we note that if $m_{ji} > 0$ then, using (2.9d), $\log \frac{m_{ji} \mu_{ji}}{m_j} = \log \Lambda'_{ji}(t)$ and if $m_{ji} = 0$ then $0 = m'_{ji}(t) = \Lambda'_{jii}(t) - \Lambda'_{ji}(t)$ and $\Lambda'_{ji}(t) > 0$. Thus, we may reintroduce the $m_{ji} = 0$ terms in our summation. In equality (4.17e), for each route, we re-interpolate the first term in our summation. ■

The next lemma, found in Cover and Thomas [9], will be of key importance in bounding our Lyapunov function.

Lemma 7 (Pinsker's Inequality) *For the relative entropy between two discrete probability distributions $p = (p_j)_j$ and $q = (q_j)_j$ with the same support:*

$$D(p||q) = \sum_j p_j \log \frac{p_j}{q_j},$$

the following inequality holds

$$\sqrt{D(p||q)} \geq \sum_j |p_j - q_j|.$$

Applying Pinsker's inequality to Proposition 7 gives

Lemma 8 For almost every t ,

$$\frac{d\beta(m(t))}{dt} \leq - \sum_{i \in \mathcal{I}} \frac{1}{\mu_{max} |\mathcal{J}|} \sum_{j \in i} \left(\Lambda'_{ji}(t) - \Lambda'_{x_{ji}i}(t) \right)^2, \quad (4.18)$$

where $|\mathcal{J}|$ is the size of set \mathcal{J} and we recall that $\mu_{max} = \max\{\mu_{ji} : i \in \mathcal{I}, j \in i\}$.

Proof: Proof Let t be a time for which Proposition 7 holds and let $\Lambda_i^\Sigma(t) = \sum_{j \in i} \Lambda'_{ji}(t)$, for $i \in \mathcal{I}$. For each i , let $p_j = \Lambda'_{ji}(t)/\Lambda_i^\Sigma(t)$ and $q_j = \Lambda'_{x_{ji}i}(t)/\Lambda_i^\Sigma(t)$. By Lemma 11, p and q both have the same support. For each i , we applying Pinsker's Lemma

$$\sum_{j \in i} \frac{\Lambda'_{ji}(t)}{\Lambda_i^\Sigma(t)} \log \frac{\Lambda'_{ji}(t)}{\Lambda'_{x_{ji}i}(t)} = \sum_{j \in i} p_i \log \frac{p_i}{q_i} \geq \left(\sum_{j \in i} |p_i - q_i| \right)^2 \geq \sum_{j \in i} |p_i - q_i|^2 = \frac{1}{(\Lambda_i^\Sigma(t))^2} \sum_{j \in i} \left(\Lambda'_{ji}(t) - \Lambda'_{x_{ji}i}(t) \right)^2$$

Multiplying the left and right of this inequality by $-\Lambda_i^\Sigma(t)$, summing over $i \in \mathcal{I}$ gives

$$\frac{d\beta(m(t))}{dt} = - \sum_{i \in \mathcal{I}} \sum_{j \in i} \Lambda'_{ji}(t) \log \frac{\Lambda'_{ji}(t)}{\Lambda'_{x_{ji}i}(t)} \leq - \sum_{i \in \mathcal{I}} \frac{1}{\Lambda_i^\Sigma(t)} \sum_{j \in i} \left(\Lambda'_{ji}(t) - \Lambda'_{x_{ji}i}(t) \right)^2.$$

Recall that from Lemma (11) that $\Lambda'_{ji}(t) \leq \mu_{max}$ thus $\Lambda_i^\Sigma(t) \leq |\mathcal{J}| \mu_{max}$. Applying this bound to $\Lambda_i^\Sigma(t)$ the above equation gives the required result (4.18). \blacksquare

We define m^* to be a solution to the optimization problem

$$\text{minimize } \beta(m) \quad \text{subject to } \sum_{j \in i} m_{ji} = n_i, \quad i \in \mathcal{I} \quad \text{over } m_{ji} \geq 0, \quad i \in \mathcal{I}, j \in i. \quad (4.19)$$

As we discussed we expect the path of the $m(t)$ to converge to the optimal value of the optimization. To conduct further analysis, we characterize the dual of this problem.

Lemma 9 The dual of the optimization (4.19) is

$$\text{maximize } \sum_{i \in \mathcal{I}} n_i \log \Lambda_i \quad \text{subject to } \sum_{i:j \in i} \frac{\Lambda_i}{\mu_{ji}} \leq 1 \quad \text{over } \Lambda_i \geq 0, \quad i \in \mathcal{I}.$$

Proof: Proof Taking Lagrange multipliers $\lambda \in \mathbb{R}^I$, its Lagrangian is,

$$\begin{aligned} L(m, \lambda) &= \sum_{j \in \mathcal{J}: m_j > 0} \sum_{i \in \mathcal{I}} m_{ji} \log \frac{m_{ji} \mu_{ji}}{m_j} + \sum_{i \in \mathcal{I}} \lambda_i \left(n_i - \sum_{j: j \in i} m_{ji} \right) \\ &= \sum_{j \in \mathcal{J}: m_j > 0} \sum_{i \in \mathcal{I}} m_{ji} \log \frac{m_{ji} \mu_{ji}}{m_j e^{\lambda_i}} + \sum_{i \in \mathcal{I}} \lambda_i n_i \\ &= \sum_{j \in \mathcal{J}: m_j > 0} m_j D(p^j || q^j) + \sum_{j \in \mathcal{J}: m_j > 0} m_j \log \left(\sum_{i: j \in i} e^{\lambda_i} \mu_{ji}^{-1} \right) + \sum_{i \in \mathcal{I}} \lambda_i n_i. \end{aligned}$$

In the final, inequality above we let $p^j = (m_{ji}/m_j : i \ni j)$ and $q^j = (e^{\lambda_i} \mu_{ji}^{-1} / \sum_{r: r \ni j} e^{\lambda_r} \mu_{jr}^{-1} : i \ni j)$. Recalling our Remark 2 on relative entropies, this Lagrangian is minimized by taking $p^j = q^j$ for each $j \in \mathcal{J}$ and then by minimizing over m_j . In particular, we get

$$\min_{m \in \mathbb{R}_+^K} L(m, \lambda) = \begin{cases} \sum_{i: n_i > 0} n_i \lambda_i & \text{if } \sum_{i: j \in i} \frac{e^{\lambda_i}}{\mu_{ji}} \leq 1, \quad \forall j \in \mathcal{J}, \\ -\infty & \text{otherwise,} \end{cases}$$

thus we find dual,

$$\text{maximize } \sum_{i: n_i > 0} n_i \lambda_i \quad \text{subject to } \sum_{i: j \in i} \frac{e^{\lambda_i}}{\mu_{ji}} \leq 1 \quad \text{over } \lambda \in \mathbb{R}^I.$$

Substituting $\Lambda_i = e^{\lambda_i}$ gives the required result

$$\text{maximize } \sum_{i: n_i > 0} n_i \log \Lambda_i \quad \text{subject to } \sum_{i: j \in i} \frac{\Lambda_i}{\mu_{ji}} \leq 1, \quad \forall j \in \mathcal{J} \quad \text{over } \Lambda \in \mathbb{R}_+^I. \quad \blacksquare$$

Lemma 10 If, for some $\epsilon > 0$, $\beta(m(t)) \geq \beta(m^*) + \epsilon$ then there exists $\delta > 0$, $i \in \mathcal{I}$ and $j \in i$ such that

$$|\Lambda'_{ji}(t) - \Lambda'_{x_{ji}i}(t)| \geq \delta.$$

Proof: Proof We develop a proof by contradiction. If this result was not true, as the set of queue sizes is compact, we could construct a sequence of times t^k , $k = 1, 2, 3, \dots$ such that $\beta(m(t^k)) \geq \beta(m^*) + \epsilon$ and $m(t^k) \rightarrow \tilde{m}$, as $k \rightarrow \infty$ and

$$\sum_{i \in \mathcal{I}} \sum_{j \in i} \left| \Lambda'_{ji}(t^k) - \Lambda_{ji}(t^k) \right| \xrightarrow{k \rightarrow \infty} 0.$$

For each queue $j \in \mathcal{I}$ with $\tilde{m}_j > 0$, let j_i^+ be the next non-empty queue on route i i.e. $m_{j_i^+} > 0$. We can say

$$\Lambda'_{ji}(t^k) \xrightarrow{k \rightarrow \infty} \frac{\tilde{m}_{ji} \mu_{ji}}{\tilde{m}_j} \quad \text{and} \quad \Lambda'_{j_i^+ i}(t^k) \xrightarrow{k \rightarrow \infty} \frac{\tilde{m}_{j_i^+ i} \mu_{j_i^+ i}}{\tilde{m}_{j_i^+}}.$$

We letting J^+ be the set of queues on route i between j and j_i^+ that includes j but does not include j_i^+ . Applying a triangle inequality across these queues, we can say that

$$\left| \frac{\tilde{m}_{ji} \mu_{ji}}{\tilde{m}_j} - \frac{\tilde{m}_{j_i^+ i} \mu_{j_i^+ i}}{\tilde{m}_{j_i^+}} \right| \leq \lim_{k \rightarrow \infty} \left[\left| \frac{\tilde{m}_{ji} \mu_{ji}}{\tilde{m}_j} - \Lambda'_{ji}(t^k) \right| + \left| \frac{\tilde{m}_{j_i^+ i} \mu_{j_i^+ i}}{\tilde{m}_{j_i^+}} - \Lambda'_{j_i^+ i}(t^k) \right| + \sum_{l \in J^+} \left| \Lambda'_{li}(t^k) - \Lambda'_{l_i i}(t^k) \right| \right] = 0.$$

Applying this triangle inequality once more, for any queue l on route i that is between j and j_i^+ , we see that

$$\lim_{k \rightarrow \infty} \left| \frac{\tilde{m}_{ji} \mu_{ji}}{\tilde{m}_j} - \Lambda'_{li}(t^k) \right| = 0.$$

In other words, for each route i for all queues $j \in i$, $\Lambda'_{ji}(t^k)$ converges to some value $\tilde{\Lambda}_i > 0$ where if $\tilde{m}_j > 0$ we have that

$$\tilde{\Lambda}_i = \frac{\tilde{m}_{ji} \mu_{ji}}{\tilde{m}_j}$$

for some constant $\tilde{\Lambda}_i > 0$. Observe that, by (2.9b), for each queue $j \in \mathcal{J}$

$$\sum_{i: j \in i} \frac{\tilde{\Lambda}_i}{\mu_{ji}} = \lim_{k \rightarrow \infty} \sum_{i: j \in i} \frac{\Lambda'_{ji}(t^k)}{\mu_{ji}} \leq 1$$

also

$$\beta(\tilde{m}) = \sum_{i \in \mathcal{I}} \sum_{j: j \in i} \tilde{m}_{ji} \log \tilde{\Lambda}_i = \sum_{i \in \mathcal{I}} n_i \log \tilde{\Lambda}_i. \quad (4.20)$$

Thus, the vector $\tilde{\Lambda} = (\tilde{\Lambda}_i : i \in \mathcal{I})$ is feasible for the dual problem and the vector \tilde{m} is feasible for the primal problem. We know by Weak Duality (for a minimization) that the any primal feasible solution is bigger than that of the dual. Thus, we know by (4.20) that the primal equals dual solution and so \tilde{m} must be optimal for the primal problem i.e. $\beta(\tilde{m}) = \beta(m^*)$. This must be a contradiction because by assumption $\beta(m(t^k)) \geq \beta(m^*) + \epsilon$ and thus by continuity of β , $\beta(\tilde{m}) \geq \beta(m^*) + \epsilon$. ■

We are now in a position to prove Theorem 3.

Proof: Proof of Theorem 3 We found $\beta(m(t))$ was absolutely continuous in t . In Lemma 8, we found the derivative of $\beta(m(t))$ was almost everywhere negative and thus $\beta(m(t))$ must be a decreasing function.

Suppose for $s \in [0, t]$, $\beta(m(s)) \geq \beta(m^*) + \epsilon$ for some $\epsilon > 0$ then by Lemma 10 there exists an $i \in \mathcal{I}$ and a $j \in i$

$$\left| \Lambda'_{ji}(s) - \Lambda_{ji}(s) \right| \geq \delta_\epsilon,$$

for some $\delta_\epsilon > 0$. Thus applying this to our bound in Lemma for intervals of time $[0, t]$ such that $\beta(m(s)) \geq \beta(m^*) + \epsilon$, we have that

$$\beta(m(t)) \leq \beta(m(0)) - t \frac{\delta_\epsilon^2}{|\mathcal{J}| \mu_{max}}.$$

As $\beta(m(t))$ is bounded below by $\beta(m^*)$, the above inequality cannot be sustained for all times t . In other words, eventually $\beta(m(t)) \leq \beta(m^*) + \epsilon$. Thus $\beta(m(t)) \searrow \beta(m^*)$. This proves the first assertion in Theorem 3.

Now, it remains to show that $m(t)$ approaches \mathcal{M} , the set of solutions to (4.19). Take some $\epsilon_1 > 0$. Let $m = (m_{ji} : i \in \mathcal{I}, j \in i)$ be any vector with $\sum_{j: j \in i} m_{ji} = n_i$ for $i \in \mathcal{I}$ and such that

$$\min_{m^* \in \mathcal{M}} |m - m^*| \geq \epsilon_1.$$

Such an m belongs to a compact set and thus, as β is continuous, it must be that $\beta(m) \geq \beta(m^*) + \epsilon$ for some $\epsilon > 0$. Or stated differently if $\beta(m) < \beta(m^*) + \epsilon$ then it must be that

$$\min_{m^* \in \mathcal{M}} |m - m^*| < \epsilon_1.$$

As we have just shown $\beta(m(t)) < \beta(m^*) + \epsilon$ holds eventually for all fluid paths. Thus,

$$\lim_{t \rightarrow \infty} \min_{m^* \in \mathcal{M}} |m(t) - m^*| = 0. \quad \blacksquare$$

A Proof of Lemmas 1 and 2.

Proof of Lemma 1: Taking $j = j_i^i$, we have

$$\begin{aligned}\Lambda_i(n) &= \sum_{\substack{m \in \mathcal{S}(n): \\ m_j > 0}} \mu_{ji} \frac{m_{ji}}{m_j} \pi(m|n) \\ &= \sum_{\substack{m \in \mathcal{S}(n): \\ m_j > 0}} \mu_{ji} \frac{m_{ji}}{m_j} \frac{1}{B(n)} \prod_{l \in \mathcal{J}} \left(\binom{m_l}{m_{lr} : r \ni l} \right) \prod_{r: l \in r} \left(\frac{1}{\mu_{lr}} \right)^{m_{lr}} \\ &= \sum_{m' \in \mathcal{S}(n-e_i)} \frac{1}{B(n)} \prod_{l \in \mathcal{J}} \left(\binom{m'_l}{m'_{lr} : r \ni l} \right) \prod_{r: l \in r} \left(\frac{1}{\mu_{li}} \right)^{m'_{lr}} = \frac{B(n-e_i)}{B(n)}.\end{aligned}$$

In the third inequality, we canceled terms and substituted $m'_{lr} = m_{lr} - 1$ if $(l, r) = (j, i)$ and $m'_{lr} = m_{lr}$ otherwise. \blacksquare

Proof of Lemma 2: We consider both optimal solutions $\bar{\Lambda}^*(n)$ and $\Lambda^*(n)$. Let

$$G_n(\Lambda) = \sum_{i \in \mathcal{I}} n_i \log \Lambda_i$$

Since $\bar{\Lambda}^*(n)$ is the solution of an optimization with a larger feasible set $G_n(\bar{\Lambda}^*(n)) \geq G_n(\Lambda^*(n))$.

Take $v = \bar{\Lambda}^*(n) - \Lambda^*(n)$. Note $\Lambda^*(n) + \delta v$ belongs to feasible set

$$\left\{ \Lambda \geq 0 : \sum_{i: j \in i} \frac{\Lambda_i}{\mu_{ji}} \leq 1 \right\}$$

for all δ suitably small. If this was not so then there would have been some constraint/queue which we did not correctly include in the set of bottleneck links $\bar{\mathcal{J}}$. Taking the partial derivative of G_n from $\Lambda^*(n)$ in the direction of v , we can then say that

$$\sum_{i \in \mathcal{I}} v_i \frac{\partial G_n(\Lambda^*(n))}{\partial \Lambda_i} \leq 0.$$

This holds because $\Lambda^*(n)$ is optimal. Now, also, by the concavity of $G_n(\cdot)$

$$G_n(\bar{\Lambda}^*(n)) - G_n(\Lambda^*(n)) \leq \sum_{i \in \mathcal{I}} v_i \frac{\partial G_n(\Lambda^*(n))}{\partial \Lambda_i}.$$

So $G_n(\bar{\Lambda}^*(n)) \leq G_n(\Lambda^*(n))$. So $G_n(\Lambda^*(n)) = G_n(\bar{\Lambda}^*(n))$. By strict concavity of $G_n(\cdot)$, $\bar{\Lambda}^*(n)$ is the unique feasible solution to its optimization problem. Thus as $\bar{\Lambda}^*(n)$ is also feasible it must be that $\bar{\Lambda}^*(n) = \Lambda^*(n)$. \blacksquare

B Lipschitz Continuity of $\beta(m(t))$.

Before proceeding to prove the Lipschitz continuity of $\beta(m(t))$, We shall quickly give a proof of Lemma 6.

Proof of Lemma 6: We use the fact that we know that the derivative exists. Firstly, it is clear $\frac{dx}{dt} = 0$ because,

$$\frac{dx}{dt} = \lim_{h \searrow 0} \frac{x(t+h) - x(t)}{h} \geq 0, \quad \text{and} \quad \frac{dx}{dt} = \lim_{h \nearrow 0} \frac{x(t+h) - x(t)}{h} \leq 0.$$

So $\frac{dx}{dt} = 0$. Noting that $x \log(x)$ is negative for all $x < 1$. By essentially the same argument

$$\frac{dx \log x}{dt} = \lim_{h \searrow 0} \frac{x(t+h) \log x(t+h) - x(t) \log x(t)}{h} \leq 0, \quad \text{and} \quad \frac{dx \log x}{dt} = \lim_{h \nearrow 0} \frac{x(t+h) \log x(t+h) - x(t) \log x(t)}{h} \geq 0.$$

Thus $\frac{dx \log x}{dt} = 0$. \blacksquare

We now demonstrate that the function $\beta(m(t))$ is Lipschitz continuous on any compact time interval. Here $m(t)$ is any solution to the fluid equations (2.9) and $\beta(m)$ is defined by (3.5). The following arguments are adapted from Lemma 4.2 and Proposition 4.2 of Bramson [5]. All queues may empty in our network, so we have to apply some degree of care in proving the Lipschitz continuity on compact time interval.

Lemma 11 *For almost every t , $0 < \Lambda'_{ji}(t) \leq \mu_{max}$, where $\mu_{max} = \max\{\mu_{ji} : i \in \mathcal{I}, j \in i\}$.*

Proof: Suppose that $m(t)$ and $\Lambda(t)$ are differentiable at t . We may assume $\Lambda'_{ji}(t) > 0$ for some queue j on route i . Such a queue must exist because there is always some queue with $m_{ji}(t) > 0$ as $n_i > 0$ and thus by (2.9d) $\Lambda'_{ji}(t) > 0$. Now consider x_{ji} the next queue on route i , if $m_{x_{ji}i}(t) > 0$ then by (2.9d) $\Lambda'_{x_{ji}i}(t) > 0$, and if $m_{x_{ji}i}(t) = 0$ then $m'_{x_{ji}i}(t) = 0$, thus by (2.9a) $\Lambda'_{x_{ji}i}(t) = \Lambda'_{ji}(t) > 0$. Continuing inductively we see that $\Lambda'_{ji}(t) > 0$ for all queues.

From this argument we now see that the value of $\Lambda'_{ji}(t)$ on any route i is achieved by a queue j^* with $m_{j^*i} > 0$. Thus applying (2.9d), $\Lambda'_{ji}(t) \leq \mu_{ji} \leq \mu_{max}$. \blacksquare

Proposition 8 For almost every $t > t_0 > 0$, there exists a constant $T > 0$ such that if $t - t_0 < \kappa T$ for $\kappa \in \mathbb{N}$ then

$$\min_{ji} \Lambda'_{ji}(t) > \frac{\min_{ji} \Lambda'_{ji}(t_0)}{(1 + \max_{ji} \mu_{ji})^\kappa}$$

Here $c_{m(t_0)}$ is a strictly positive constant which depends on $m(t_0)$, the fluid model state at time t_0 .

In order to prove this proposition we require the following lemma

Lemma 12 For almost every time t_0 and t with $t > t_0$, if a queue j on route i , has arrival process from the queue before j , l_{ji} such that for almost every $s \in [t_0, t]$

$$\Lambda'_{l_{ji}j}(s) > c,$$

then, almost everywhere, the output of route i work from queue j satisfies

$$\Lambda'_{ji}(t) \geq c' \tag{B.1}$$

where

$$c' = \min \left\{ \Lambda'_{ji}(t_0), \frac{c}{1 + \max_{ji} \mu_{ji}} \right\}.$$

Proof: We note that $\Lambda_{ij}(t)$ is a Lipschitz function and thus is almost every where differentiable. We assume that t and t' are differentiable points where (B.1) is violated. Observe that if $m_{ji}(t) = 0$ then, by (2.9a), $0 = m'_{ji}(t) = \Lambda'_{ji}(t) - \Lambda'_{l_{ji}i}(t)$. Thus $\Lambda'_{ji}(t) = \Lambda'_{l_{ji}i}(t) > c > c'$. So it must be that $m_{ji}(t) > 0$. Consequently by (2.9d), $\Lambda'_{ji}(s) = \mu_{ji} m_{ji}(s) / m_j(s)$ must be continuous on an interval around t and there must be an open interval around t for which $\Lambda'_{ji}(s) < c'$.

For a $\Lambda'_{ji}(s)$ to get small we need the total number of departures to be comparable relative to the arrivals. So, we will next argue the contradiction that $\Lambda'_{ji}(t)$ cannot enter an interval of time for which $\Lambda'_{ji}(s) < c'$ without the average departure rate $(\Lambda'_{ji}(t) - \Lambda'_{ji}(s)) / (t - s)$ being bigger than c' .

We let \tilde{t} be the last time before t and after t_0 for which $\Lambda_{ji}(\tilde{t}) \geq c'$. Note $\Lambda_{ji}(t_0) \geq c'$, so \tilde{t} is well defined. We use the shorthand $\Lambda_{ji}(\tilde{t}, t) = \Lambda_{ji}(t) - \Lambda_{ji}(\tilde{t})$ and $\Delta_j(\tilde{t}, t) = m_j(t) - m_j(\tilde{t})$. As $m_{ij}(t) > 0$, by (2.9d), we have

$$c' > \Lambda'_{ji}(t) = \mu_{ji} \frac{m_{ji}(t)}{m_j(t)} = \mu_{ji} \frac{m_{ji}(\tilde{t}) + \Lambda_{l_{ji}i}(\tilde{t}, t) - \Lambda_{ji}(\tilde{t}, t)}{m_j(\tilde{t}) + \Delta_j(\tilde{t}, t)}.$$

Rearranging the above expression implies

$$\Lambda_{ji}(\tilde{t}, t) > \Lambda_{l_{ji}i}(\tilde{t}, t) - c' \Delta_j(\tilde{t}, t) + \mu_{ji} m_{ji}(\tilde{t}) - c' m_j(\tilde{t}) \geq \Lambda_{l_{ji}i}(t) - c' \Delta_j(\tilde{t}, t).$$

The last inequality holds because, by (2.9d), $\Lambda'_{l_{ji}i}(\tilde{t}) \geq c'$ implies $\mu_{l_{ji}i} m_{l_{ji}i}(\tilde{t}) - c' m_j(\tilde{t}) \geq 0$. Let's now look at the mean value of the terms in the above inequality:

$$\frac{\Lambda_{ji}(\tilde{t}, t)}{t - \tilde{t}} > \frac{\Lambda_{l_{ji}i}(\tilde{t}, t)}{t - \tilde{t}} - c' \frac{\Delta_j(\tilde{t}, t)}{t - \tilde{t}} \geq c - c' \max_{ji} \mu_{ji} \geq c'.$$

In the second inequality above, we use the assumption that $\Lambda'_{l_{ji}i}(s) > c$, $s \in [t_0, t_0 + T]$ and the fact the average change in $m_j(t)$, $\Delta_j(\tilde{t}, t) / (t - \tilde{t})$, is at most by the maximum service rate $\max_{ji} \mu_{ji}$. The final inequality holds by our choice of c' . This then contradicts our assumption that $\Lambda'_{ji}(s) < c'$ on the interval $(\tilde{t}, t]$. \blacksquare

The following lemma is a consequence achieved by iteratively applying the last result. We have to apply some care because, in comparison to Bramson's open network analysis [5], every queue can empty or have low arrival and departure rate.

Lemma 13 For almost every t_0 , there exists an interval of fixed length $[t_0, t_0 + T]$ such that for almost every $t \in [t_0, t_0 + T]$ and for all $i \in \mathcal{I}$ and $j \in \mathcal{J}$

$$\Lambda'_{ji}(t) > \tilde{c}(t_0) > 0.$$

Here $\tilde{c}(t_0)$ is a function of the network's state at time t_0 .

Proof: On route i there is always one queue with greater than or equal to the average amount of work. With out loss of generality, i.e. relabelling queues if necessary, we assume that this is the first queue on route i . So, we have that for any route i

$$m_{j_i^1}(t_0) \geq \frac{\min_r n_r}{I}.$$

The biggest rate that this queue could decrease by is $\max_{j_i} \mu_{j_i}$. So

$$m_{j_i^1}(t) \geq \frac{\min_r n_r}{2I} \quad \text{for } t \in [t_0, t_0 + T],$$

where we define $T = \frac{\min_r n_r}{2I \max_{j_i} \mu_{j_i}}$.

$$\Lambda_{j_i}(t_0) = \mu_{j_i} \frac{m_{j_i}(t_0)}{m_j(t_0)} > \frac{\min_{j_i} \mu_{j_i} \min_r n_r}{2I \sum_r n_r} = c_1.$$

We can now repeatedly apply Lemma (12). Starting from c_1 above for $k = 1, \dots, k_i - 1$ we define

$$c_{k+1} = \min \left\{ \Lambda'_{j_i}(t_0), \frac{c_k}{1 + \max_{j_i} \mu_{j_i}} \right\}.$$

Each time we iterate we reduce c_k by at least $1 + \max_{j_i} \mu_{j_i}$. A simple lower bound, which is sufficient for our purposes is that

$$c_k \geq \frac{\min_{j_i} \Lambda'_{j_i}(t_0)}{(1 + \max_{j_i} \mu_{j_i})^K} =: \tilde{c}(t_0)$$

where $K = \max_i k_i$ is the longest route within the queueing network. Thus given this, we have from Lemma (12) that

$$\Lambda'_{j_i}(t) \geq \tilde{c}(t_0)$$

for almost every $t \in [t_0, t_0 + T]$. ■

Here in a similar manner to Proposition 4.2 of Bramson [5] we show that $\beta(t)$ is Lipschitz on any compact time interval.

Proof of Proposition 6: Note that without loss of generality, we may assume interval $[t_0, t]$ is of length less than or equal to $T = \frac{\min_r n_r}{2I \max_{j_i} \mu_{j_i}}$, where T was derived in the last lemma, Lemma 13. If $t - t_0 > T$ then we can split the interval $[t_0, t]$ in to overlapping sub-interval of size T and then use the largest Lipschitz constant found in each sub-interval as a Lipschitz constant for $[t_0, t]$.

Note that

$$\beta(m(t)) = \sum_{j \in \mathcal{J}} \sum_{i: j \in i} m_{j_i}(t) \log \frac{m_{j_i}(t) \mu_{j_i}(t)}{m_j(t)} = \sum_{j \in \mathcal{J}} \sum_{i: j \in i} m_{j_i}(t) \log \Lambda'_{j_i}(t).$$

It is enough to prove Lipschitz continuity of each term summed above:

$$|m_{ij}(t_2) \log(\Lambda'_{j_i}(t_2)) - m_{ij}(t_1) \log(\Lambda'_{j_i}(t_1))| \leq D_1 |t_2 - t_1|.$$

By Lemma 13 for $s \in [t_0, t]$, $\log(\Lambda'_{j_i}(s))$ is bounded below by $\tilde{c}(t_0)$ and also above by C_j , the capacity of queue j . So

$$|\log(\Lambda'_{j_i}(s))| \leq D_0 := \max\{|\log(C_j)|, |\log(\tilde{c}(t_0))|\}.$$

If $m_{ij}(t_2) = m_{ij}(t_1) = 0$, the relation is trivial, so we assume that $m_{ij}(t_2) > 0$. If $m_{ij}(t_1) = 0$ we have that

$$|m_{ij}(t_2) \log(\Lambda'_{j_i}(t_2))| \leq D_0 |m_{ij}(t_2)| = D_0 |m_{ij}(t_2) - m_{ij}(t_1)| \leq D_2 D_0 |t_2 - t_1|$$

where the constant D_0 comes from Lemma 13 and constant D_2 is the Lipschitz constant of $m_{ij}(t)$.

Now assume $m_{ij}(t_2)$ and $m_{ij}(t_1)$ both positive, and without loss of generality that

$$\Lambda'_{j_i}(t_1) \leq \Lambda'_{j_i}(t_2).$$

It follows that

$$\begin{aligned} |m_{ij}(t_2) \log \Lambda'_{j_i}(t_2) - m_{ij}(t_1) \log \Lambda'_{j_i}(t_1)| &\leq |m_{ij}(t_2) - m_{ij}(t_1)| |\log \Lambda'_{j_i}(t_2)| \\ &\quad + m_{ij}(t_1) |\log \Lambda'_{j_i}(t_2) - \log \Lambda'_{j_i}(t_1)| \end{aligned}$$

Again the first term in this upperbound is less than or equal to $D_2 D_0 |t_2 - t_1|$. The second term, since the logarithm function is concave and has its derivative maximized at the left point, can be bounded in the following way

$$\begin{aligned} m_{ij}(t_1) |\log \Lambda'_{j_i}(t_2) - \log \Lambda'_{j_i}(t_1)| &\leq \frac{m_{ij}(t_1)}{\Lambda'_{j_i}(t_1)} |\Lambda'_{j_i}(t_2) - \Lambda'_{j_i}(t_1)| \\ &= \left| m_j(t_1) \frac{m_{j_i}(t_2)}{m_j(t_2)} - m_{ij}(t_1) \right| \\ &\leq \frac{m_{ij}(t_2)}{m_j(t_2)} |m_j(t_2) - m_j(t_1)| + |m_{ij}(t_2) - m_{ij}(t_1)| \\ &\leq |m_j(t_2) - m_j(t_1)| + |m_{ij}(t_2) - m_{ij}(t_1)| \\ &\leq D_3 |t_2 - t_1| \end{aligned}$$

that follows from the fact that $m_{ij}(t_2) \leq m_j(t_2)$ and the fact that both $m_{ij}(t)$ and $m_j(t)$ are Lipschitz continuous. The result follows by choosing $D_1 \geq \max\{D_3, D_2 D_0\}$. ■

References

- [1] J. Anselmi and P. Cremonesi, *A unified framework for the bottleneck analysis of multiclass queueing networks*, Perform. Eval. **67** (2010), 218–234.
- [2] Gianfranco Balbo and Giuseppe Serazzi, *Asymptotic analysis of multiclass closed queueing networks: Multiple bottlenecks*, Performance Evaluation **30** (1997), no. 3, 115–152.
- [3] F. Baskett, K.M. Chandy, R.R. Muntz, and F.G. Palacios, *Open, closed, and mixed networks of queues with different classes of customers.*, Journal of the ACM **22** (1975), no. 2, 248–260.
- [4] Arthur Berger, Lev Bregman, and Yaakov Kogan, *Bottleneck analysis in multiclass closed queueing networks and its application*, Queueing Syst. Theory Appl. **31** (1999), no. 3-4, 217–237.
- [5] Maury Bramson, *Convergence to equilibria for fluid models of head-of-the-line proportional processor sharing queueing networks*, Queueing Systems **23** (1996), 1–26.
- [6] Giuliano Casale, *A generalized method of moments for closed queueing networks*, Perform. Eval. **68** (2011), no. 2, 180–200.
- [7] Giuliano Casale and Giuseppe Serazzi, *Bottlenecks identification in multiclass queueing networks using convex polytopes*, MASCOTS '04 (Washington, DC, USA), IEEE Computer Society, 2004, pp. 223–230.
- [8] K. Mani Chandy and Doug Neuse, *Linearizer: a heuristic algorithm for queueing network models of computing systems*, Commun. ACM **25** (1982), no. 2, 126–134.
- [9] Thomas M. Cover and Joy A. Thomas, *Elements of Information Theory*, Wiley-Interscience, August 1991.
- [10] Paul Dupuis and Markus Fischer, *On the construction of lyapunov functions for nonlinear markov processes via relative entropy*, preprint (2012).
- [11] David K. George, Cathy H. Xia, and Mark S. Squillante, *Exact-order asymptotic analysis of closed queueing networks*, Journal of Applied Probability (2012).
- [12] Jonathan B. Goodman and William A. Massey, *The non-ergodic jackson network*, Journal of Applied Probability **21** (1984), no. 4, pp. 860–869.
- [13] W. J. Gordon and G.F. Newell, *Closed queueing systems with exponential servers*, Operation Research (1967), no. 15, 254–265.
- [14] Peter G. Harrison and Sergio Coury, *On the asymptotic behaviour of closed multiclass queueing networks*, Perform. Eval. **47** (2002), no. 2, 131–138.
- [15] F. P. Kelly, *Charging and rate control for elastic traffic*, European Transactions on Telecommunications **8** (1997), 33–37.
- [16] F. P. Kelly, L. Massoulié, and N. S. Walton, *Resource pooling in congested networks: proportional fairness and product form*, Queueing Syst. Theory Appl. **63** (2009), 165–194.
- [17] F. P. Kelly, A. K. Maulloo, and D. K. H. Tan, *Rate control in communication networks: shadow prices, proportional fairness and stability*, Journal of the Operational Research Society **49** (1998), 237–252.
- [18] F.P. Kelly, *Reversibility and stochastic networks*, Wiley, Chicester, 1979.
- [19] Charles Knessl and Charles Tier, *Asymptotic expansion for large closed queueing networks*, Journal of the ACM **37** (1990), no. 1, 144–174.
- [20] ———, *Asymptotic expansions for large closed queueing networks with multiple job classes*, IEEE Trans. Comput. **41** (1992), no. 4, 480–488.
- [21] Lester Lipsky, Chee-Min Henry Lieu, Abolfazl Tehranipour, and Appie van de Liefvoort, *On the asymptotic behavior of time-sharing systems*, Commun. ACM **25** (1982), no. 10, 707–714.
- [22] J. McKenna and Debasis Mitra, *Asymptotic expansions and integral representations of moments of queue lengths in closed markovian networks*, Journal of the ACM **31** (1984), no. 2, 346–360.
- [23] ———, *Asymptotic expansions for closed markovian networks with state-dependent service rates*, J. ACM **33** (1986), no. 3, 568–592.
- [24] K. R. Pattipati, M. M. Kostreva, and J. L. Teele, *Approximate mean value analysis algorithms for queueing networks: existence, uniqueness, and convergence results*, J. ACM **37** (1990), no. 3, 643–673.
- [25] B. Pittel, *Closed exponential networks of queues with saturation: The jackson-type stationary distribution and its asymptotic analysis*, Mathematics of Operations Research **4** (1979), no. 4, 357–378.
- [26] M. Reiser and H. Kobayashi, *Queueing networks with multiple closed chains: Theory and computational algorithms*, IBM J. Res. Dev. **19** (1975), no. 3, 283–294.
- [27] M. Reiser and S. S. Lavenberg, *Mean-value analysis of closed multichain queueing networks*, Journal of the ACM **27** (1980), no. 2, 313–322.

- [28] Philippe Robert, *Stochastic networks and queues*, Springer-Verlag, 2000.
- [29] P. Schweitzer, *Approximate analysis of multi-class closed networks of queues*, In Proceedings of the International Conference on Stochastic Control and Optimization (1979).
- [30] ———, *Bottleneck determination in networks of queues*, Proc. ORSA/TIMS Special Interest Conf on Appl. Probab. - Comput. Sci., The Interface, Boca Raton, FLA (1981), 471–485.
- [31] P. Schweitzer, G. Serazzi, and M. Broglia, *A survey of bottleneck analysis in closed queues*, Perf. Eval. of Comp. and Comm. Sys., LNCS, No. 729, Springer-Verlag, Berlin (1993), 491–508.
- [32] F. Spitzer and Mathematical Association of America, *Random fields and interacting particle systems: notes on lectures given at the 1971 maa summer seminar, williams college, williamstown, massachusetts*, Random Fields and Interacting Particle Systems, Mathematical Association of America, 1971.
- [33] R. Srikant, *The mathematics of internet congestion control*, Birkhauser, 2004.
- [34] Bhuvan Urgaonkar, Giovanni Pacifici, Prashant Shenoy, Mike Spreitzer, and Asser Tantawi, *An analytical model for multi-tier internet services and its applications*, ACM SIGMETRICS, ACM, 2005, pp. 291–302.
- [35] N. S. Walton, *Proportional fairness and its relationship with multi-class queueing networks*, Annals of Applied Probability **19** (2009), no. 6, 2301–2333.
- [36] Hai Wang, Kenneth C. Sevcik, Giuseppe Serazzi, and Shouhong Wang, *The general form linearizer algorithms: A new family of approximate mean value analysis algorithms*, Perform. Eval. **65** (2008), no. 2, 129–151.