arXiv:submit/2047812 [cs.IT] 26 Oct 2017

Interference Queuing Networks on Grids

A Framework for Performance Analysis of Large Wireless Networks

Abishek Sankararaman*, François Baccelli †, Sergey Foss ‡

Abstract

Motivated by applications in large scale wireless networks, we introduce and study queuing systems of *interacting queues* where the interactions mimic wireless interference. Our network model is an abstraction of an ad-hoc wireless network and is intended to capture the interplay between the geometry of wireless links, which determine how the links interfere with each other, and their temporal traffic dynamics. Specifically, when a link accesses the spectrum, it causes interference to other nearby links, and thus their rate of communication is lowered. This in turn implies they access the spectrum longer and cause interference to nearby links for a larger duration. This coupling between the geometry and temporal dynamics is one of the main challenges in assessing the performance of wireless systems. Most prior work have modeled this phenomenon in a cellular network setting by generalizing the so called *coupled processors model*. These models are difficult to analyze and even in the simplest case, and one has to typically resort to bounds and approximations to derive insight. We consider the ad-hoc network setting and propose a new model of queues, which interact through interference. We show that this model is tractable, in the sense we can compute certain performance metrics even when there is an *infinite number of queues*. Specifically, we prove a stability phase-transition for such queuing systems. Furthermore, we also derive *exactly* the mean number of customers in any queue in steady state. Thanks to Little's Law, this also yields an exact expression for mean delay of a typical link in steady state. The techniques and results in this paper hold both for large finite networks as well as infinite networks. The key to our results is the analysis of certain Rate Conservation equations along with monotonicity and tightness arguments, which are new and interesting in their own right. Our model and results thus provide a basis to evaluate the performance of various protocols on large networks.

1 Introduction

In the present paper, we introduce and study a new model that consists of a large or infinite collection of queues whose evolutions are coupled together by their interactions. The model we propose is motivated by design and scalability questions stemming from emerging applications such as Device-to-Device (D2D) networks and Internet-of-Things (IoT), where there is no centralized entity that coordinates the spectrum access of devices. Although these types of ad-hoc networks have been extensively studied in the past, many fundamental questions, especially those pertaining

*Department of ECE, The University of Texas at Austin. Email - abishek@utexas.edu.

†Department of Mathematics and ECE, The University of Texas at Austin. Email - baccelli@math.utexas.edu.

‡School of MACS, Heriot-Watt University, Edinburgh and Dept of Mechanics and Mathematics, Novosibirsk State University. Email - S.Foss@hw.ac.uk

to large networks are still not understood satisfactorily. One of the key challenges in assessing performance of such wireless networks lies in the complex interplay induced by the spatial location of links that govern the interference they cause to one another and the temporal traffic dynamics. More specifically, if a transmitter of a link is transmitting a file to its receiver, it will cause interference to other transmitting links. Thus the neighboring transmitters will transmit data to their respective receivers at a slower rate because of the interference. This in turn causes these transmitters to transmit longer, thereby causing interference for an extended period of time at the initial link in consideration. Understanding and studying this positive correlations between the geometry of nodes and their dynamics is crucial in order to design and evaluate spectrum access algorithms. The aim of this paper is to present a framework that captures the correlations in time and space in such networks thereby allowing one to investigate the performance analytically.

The most popular method to study these correlations has been by representing the wireless network as a *flow based* queuing network (see for ex. [14],[22]). In these models, the randomness in traffic is represented as 'flows' that interact. The analysis of this framework is then used to design 'optimal protocols' in various contexts. In particular, these models have been successfully applied in modeling spatio-temporal correlations in the cellular network setting ([6],[7]), by extending the so called coupled processors model ([10],[9]). However, these models are not tractable and rely on bounds and approximations, even to study 'small' networks consisting of a few transmitting stations. Due to increasingly large-scale deployments of wireless networks, Stochastic Geometry ([1],[13]) has recently become popular as a tractable method to assess the performance of different protocols in large networks with many transmitting stations. Stochastic Geometric models allow one to compare the performance of different protocols which provides some basis in choosing certain protocols as optimal. Nevertheless, the main drawback in these models is that they do not allow one to represent traffic flows as they typically assume a "full-buffer" condition, i.e., every link always has a packet to transmit. Our work proposes a model to address the shortcomings of both approaches.

Specifically, we make three contributions in this paper. First, we develop a novel model for a network consisting of a large and possibly infinite set of interacting queues which combines the geometric representation along with the temporal dynamics due to traffic. Our model captures both the large-scale nature of wireless networks which is missing in the coupled queues models of [6], along with the temporal interactions that are missing in stochastic geometric models. Second, we bring new analytic techniques to study such queuing networks, namely Rate Conservation and coupling arguments which handle the correlations without resorting to any independence based bounds or asymptotics. Our method is applicable both to the problem defined on the infinite grid, as well as to large finite networks restricted to a torus. Third, we compute an *exact* formula for the mean queue length of any queue in the network. This is expected to provide a basis to asses the gains obtained by using more sophisticated physical and MAC layer algorithms compared to the baseline protocol studied in the present paper.

In regards to our contributions, one may wonder what are the practical benefits of studying an infinite network of queues, since all real networks, no matter how large, are still finite. Moreover, as we state above and demonstrate in this paper, all our techniques and results hold for large finite networks and do not require the analysis of the infinite case. Our motivations to discuss the analysis of both large finite as well as the infinite network are two fold. First, from a mathematical point of view, the infinite model provides an elegant method of modeling dynamics on classical stochastic geometric objects, which are typically defined on an infinite domain. Second and more importantly in our context, the infinite network provides a clean mathematical abstraction to understand many phenomena pertaining to large finite networks. Just as in the Ising Model [15] for example, in our case, the infinite model could potentially admit multiple stationary distributions (See Corollary 10 and the remark following it). If it is the case, by analogy to the Ising model, it could indicate the presence of long range spatial correlations. Practically, operating the network in this 'phase' could be undesirable. Thus understanding the infinite model could pave the way for more refined analysis of large finite networks, as done for other interacting particle systems such as the Ising model. In view of this, we find it important to describe and analyze the infinite network model along with large finite networks.

The rest of this paper is organized as follows. In the remainder of this section, we discuss our model, and its interpretation. In Section 2, we survey the related work and connect it with the vast literature on wireless performance analysis and general large queuing networks. In Section 3 we describe the probabilistic framework, establish certain monotonicity properties and define precisely the notion of stochastic stability we study in this paper. The main theoretical results are stated in Section 4. In order to prove them, we consider certain space truncated systems in Section 5, where we also present the Rate Conservation argument. We finish the proof of the main theorem in Section 6. For clarity in presentation, we give only the main ideas of proofs in the paper and defer the technical details to the Appendix.

1.1 The Model Description

We present our model heuristically and defer a precise mathematical definition till Section 3. The infinite model consists of an infinite collection of processor sharing queues that evolve in continuous time, with a queue located at each grid point of $\mathbb{Z}^d$, for some $d \geq 1$. The arrival processes across queues are independent and identically distributed with the arrival process in each queue being a Poisson Point Process of intensity $\lambda > 0$. The queues interact through an *interference sequence* $\{a_i\}_{i\in\mathbb{Z}^d}$. We assume that $\{a_i\}_{i\in\mathbb{Z}^d}$ is non-negative and finitely supported, i.e. $L := \max\{||i||_\infty : a_i > 0\} < \infty$. Furthermore, we impose a natural symmetry condition of $a_{-i} = a_i$ for all $i \in \mathbb{Z}^d$. For ease of notation, we normalize the sequence and consider $a_0 = 1$. Let $\{x_k(t)\}_{k\in\mathbb{Z}^d} \in \mathbb{N}^{\mathbb{Z}^d}$ denote the queue lengths at some $t \in \mathbb{R}$ in the network. The instantaneous rate of departure from any queue at time t is then given by $\frac{x_k(t)}{\sum_{j\in\mathbb{Z}^d} a_j x_{k-j}(t)}$ provided that $x_k(t) > 0$. However, the departure processes are conditionally independent given queue lengths, i.e., the departures are only coupled by their instantaneous rate. Observe that since $\{a_i\}_{i\in\mathbb{Z}^d}$ is a non-negative sequence, the rate of departure from a queue is reduced, if its 'neighbors' have 'large' queue lengths. This is the basic coupling between geometry and temporal dynamics as alluded to earlier. The central question we study in this paper is that of the stability of the above infinite queuing system. Due to the positive correlations, it is not clear apriori whether a region of stability even exists for this model. The main result in the present paper consists in establishing that the above infinite system undergoes a stability phase-transition at a critical arrival intensity $\lambda_c := \frac{1}{\sum_{j\in\mathbb{Z}^d} a_j}$ together with similar results for large networks with finitely many queues.

Notice that we can also study a networks consisting of queues located on a large finite torus of $\mathbb{Z}^d$, which will be finite restrictions of our infinite model. Our results and analysis holds for such finite networks as well as the infinite network described above (see Section 5). The model stated above is starkly different from the coupled processors model of [10],[9] and the prior ad-hoc network models of [20], since the rate of service at any queue in our model depends on the queue lengths at all servers and not just on whether a server is busy or not. Nevertheless, our model does not admit a product form stationary distribution like Jackson networks. Even when there are infinitely many queues, our model has no asymptotic independence properties as those encountered in mean-field models (such as the supermarket model [24]) due to the positive correlations between queues (see Figure 1). The correlations across queues is intuitive, since if a queue has a large number of customers, then its neighboring queues will receive lower rates, and thus they will in turn build up. Therefore in steady state, if a particular queue is large, most likely its neighboring queues are also large. The presence of correlations also implies that there is no mean-field based approximation, even when we have infinitely many queues. Indeed the main achievement in this paper is in analyzing this model despite the lack of independence across queues.

1.2 Interpretation of the Model

We first describe a continuous space-time network model which is an infinite space version of the model introduced in [18]. We will subsequently argue that our queuing model can be seen as a natural space-discretized version of this continuum space description. The network can be thought of as being composed of 'wireless links', where each link is a transmitter-receiver pair. The links are present in the wireless network which we assume to be the infinite plane $\mathbb{R}^d$. In this context, our network model is also known as the *dipole model* [1], which is the single-hop version of the Gupta-Kumar model [12]. The flow dynamics we envision is the following. New wireless links 'arrive' randomly in space at some random time. More precisely assume that the arrival process of links (transmitters) is a Poisson Point Process in space and time. Each arriving transmitter brings its own receiver at a fixed distance say T from itself, but at a random direction. The transmitter of each arriving link has a file it wants to transmit to its corresponding receiver. For simplicity, we assume that the length of this file is an independent random variable with unit mean exponential distribution. The transmitter starts transmitting the file the moment it arrives and the link exits the network after completion of the file transfer. Thus, the links themselves represent the flows which arrive at random locations in space. Moreover, each arriving transmitter always transmits at full power the moment it arrives irrespective of network conditions. However, the instantaneous rate of communication is limited by the instantaneous interference seen at the receiver, which is in turn dictated by the geometry of the currently active links. To be precise, assume that signal power decays with distance by a non-increasing function $l(\cdot) : \mathbb{R}_+ \to \mathbb{R}_+$, where $l(r)$ represents the received power at distance r, when the transmitted power is unity. Then, if the receiver in consideration is at location $y \in \mathbb{R}^d$ and ϕ denotes the set of location of all interfering transmitters, then the total instantaneous interference is given by $\sum_{x\in\phi} l(||y-x||)$, the sum of powers from all interfering transmitters. At the physical layer, we assume that the receivers treat interference as noise. Thus the instantaneous rate of communication can then be given by Shannon's formula for point-to-point channels, i.e., $\log(1+\mathbf{SINR})$, where **SINR** refers to the Signal-to-Interference plus Noise power ratio. If we thus assume unit transmit signal power and thermal noise power equal to a constant N_0, the instantaneous rate of communication is given by $\log\left(1+\frac{l(T)}{N_0+\sum_{x\in\phi} l(||y-x||)}\right)$. Notice that since $l(\cdot)$ is non-negative and non-increasing, the farther the interferers are from a given tagged receiver, the higher is the instantaneous rate. This captures the correlations between temporal dynamics and geometry. After completion of its file transfer, a link leaves the network and becomes inactive.

Our model, described in Section 1.1, can be seen as a space-discretized version of the above spatial birth-death model. The grid points in our model represent tiny 'chunks' of space of $\mathbb{R}^d$. The number of customers at some time t in queue $z \in \mathbb{Z}^d$ is the number of links at time t located in the tiny region of space corresponding to queue z. For simplicity, we assume that links are very tiny (mathematically a point) and hence both the transmitter and receiver of a link can be seen to be located at the same point. This is of-course only done to have a clear presentation and convey the main message and one can easily have non-zero link lengths as done in [18], with significantly heavier notation. This is the key difference between our model and those of for example [19], where a queue represents a link in the model of [19]. In contrast, in our model, a queue represents a region of space, and the queue length denotes the number of transmitting devices in that region of space. We assume that, the interference function $l(\cdot)$ is discretized as the sequence $\{a_i\}_{i \in \mathbb{Z}^d}$. Thus a_i represents the received power at queue 0 (or queue i) if unit power is transmitted at queue i (or at queue 0). However, we relax the condition that a_i be non-increasing in $||i||$ but impose the stricter condition that it is finitely supported. Thus the total interference at any link in queue 0 at time t is $a_0(x_0(t) - 1) + \sum_{i \in \mathbb{Z}^d \setminus \{0\}} a_i x_i(t)$. To derive the instantaneous rate, we assume a linear approximation of the Shannon rate and model the instantaneous rate of communication is given by **SINR**. Thus, the instantaneous rate of communication at any link in queue 0 at time t is then given by $\frac{1}{a_0(x_0(t)-1)+\sum_{i \in \mathbb{Z}^d \setminus \{0\}} a_i x_i(t) + N_0}$, where N_0 is the thermal noise power. The minus 1 for the a_0 term is to account for the fact that a link will not interfere with itself. For further ease of notation, we assume $N_0 = a_0$ and thus obtain that the instantaneous rate of communication for any link at queue 0 is given by $\frac{1}{\sum_{i \in \mathbb{Z}^d} a_i x_i(t)}$. Now as there are $x_0(t)$ links at queue 0 and they all have independent exponential unit mean file sizes, the total rate of departure is then given by $\frac{x_0(t)}{\sum_{j \in \mathbb{Z}^d} a_j x_j(t)}$.

Comments on the Model

From a practical point of view, our model can be criticized in several ways. First, one may object that it is missing many important elements of a real interference channel model such as fading, power control and link-lengths. It also lacks the representation of access protocols used in this context. These are all perfectly valid objections which may limit the direct applications of our results in practice. However, analyzing the simple access protocol of the present paper is challenging and is a first necessary step. We hope that this will pave way to analyze and design more practical protocols in the future. Our theoretical result requires the statistical assumptions of Poisson arrivals and exponential file sizes, which may not always be valid in practice. Nonetheless the stylized queuing model presented in this paper already captures the fundamental difficulty in assessing the coupling of geometry and dynamics in this context, which, to the best of our knowledge, has not been rigorously studied before. All the simplifications made in the model are hence aiming at finding the simplest yet non-simplistic mathematical model for a large class of problems that will be studied later. The other questionable aspect in our model is that we use a linear-approximation to Shannon's formula for channel capacity. Our justification for this is that we consider the low SINR regime where a linear approximation is accurate. The low SINR naturally presents itself in the 'heavy-traffic' regime when λ is close to λ_c. This is because the queue lengths will be 'large', thereby making the interference typically 'large'. Thus one expects the same stability region if one used a linear approximation as opposed to the exact formula. Furthermore, one could object that it is odd to represent such a network through an infinite number of queues since any network in practice, no matter how large, is still finite. We want to stress that our results and insights exactly hold for finite networks as well (see Section 5 and Theorem 11). As stated, the infinite network allows us to *model* a *large* wireless system and provides a good abstraction to study many important questions pertaining to large wireless systems. The infinite model exhibits nice phase-transitions which provide qualitative understanding, and it is computationally and analytically tractable due to the symmetries. Understanding whether a design is scalable is of paramount importance in large-scale deployments of networks, and the infinite network model is the right mathematical abstraction to consider this question.

1.3 Interpretation of our Results

The main technical result in this paper is that there is a critical arrival intensity $\lambda_c := \frac{1}{\sum_{j \in \mathbb{Z}^d} a_j}$ such that if $\lambda < \lambda_c$, then all queues are stable (see Theorem 6). Furthermore in Corollary 10, we show that if $\lambda < \lambda_c$, the minimal solution is the *unique stationary distribution* for the queue lengths that have finite mean. We also compute an *exact* expression for the mean queue length for this stationary solution and show that it is equal to $(\lambda^{-1} - \lambda_c^{-1})^{-1}$, a surprisingly simple formula. Due to symmetry, all queue lengths are identically distributed and hence have the same mean. We furthermore show that if we started our network empty and $\lambda < \lambda_c$, then the queue lengths converge weakly to the unique stationary distribution with finite mean. Since our network is not finite-dimensional, stability does not imply ergodicity. Thus, it is not apriori clear if there is an unique stationary solution or not and whether all initial conditions converge to a stationary solution. On the converse, we conjecture that if $\lambda > \lambda_c$, all queues are transient. We give evidence towards this by proving the conjecture for the important special case of $d = 1$ and $a_i = 1$ for all $|i| \leq 1$ and $a_i = 0$ for all $|i| > 1$.

To have an intuitive understanding for why λ_c is the critical arrival rate, consider a queue $i \in \mathbb{Z}^d$, which is a 'local maximum', i.e., has the maximum queue length in a neighborhood of a l_∞ ball of radius L, the support of the seqeunce $\{a_i\}_{i \in \mathbb{Z}^d}$. Then the instantaneous rate of departure is $\frac{x_i(t)}{\sum_{j \in \mathbb{Z}^d} a_j x_{i-j}(t)}$ which is at-least as big as $\frac{1}{\sum_{j \in \mathbb{Z}^d} a_j}$. Hence if $\lambda < \lambda_c$, the queues having a local maximum experiences a negative drift. This is the intuition explaining why $\frac{1}{\sum_{j \in \mathbb{Z}^d} a_j}$ is the critical value. Another interpretation of the phase-transition that it is the inverse of the integral (sum) of the path-loss function. More specifically, if there is a single customer in all queues, transmitting at unit power, then the rate seen by any the customer in any queue, which is $\frac{1}{\sum_{i \in \mathbb{Z}^d} a_i}$ is the critical arrival rate. This result is intuitive in the sense that the critical arrival rate is smaller in environments with low attenuation and large in environments with high attenuation. Our result in Theorem 6 makes this intuition precise and provides guidelines in deciding how to operate networks depending on how signal power attenuates with distance. Thanks to Little's law, the formula for the mean queue length in steady state also gives an expression for mean-delay of a typical arrival in any queue. An exact expression for delay is of great importance in network dimensioning. For instance, in overlaid D2D networks, the formula for mean delay can give help set a threshold to determine how much traffic a cellular base-station can off-load onto a D2D network, while still maintaining Quality of Service. In this example, the arrival rate of links will be dictated by the fraction of traffic a base-station off-loads. More generally, our results provides a benchmark by assessing the performance of a simple spectrum access protocol wherein an arriving transmitter always transmits at full power and the receiver treats all interference as noise. This now opens up a way to assess the gain in performance obtained from more sophisticated physical and MAC layer algorithms such as if the receiver employed interference cancellation or the links used admission control.

The key technical challenge in analyzing our model is the positive correlation between queue lengths which persist even in the model with infinitely many queues (See also Figure 1). A proof of stability of the infinite network cannot be done by standard Lyapunov arguments. Instead, we rely on a fluid like construction of space-truncated systems (see Section 5). We then take the asymptotic as the space becomes all of $\mathbb{Z}^d$ and prove one can exchange limits in space and time (see Section 6) by using monotonicity and coupling arguments. To obtain an exact formula for mean queue length, we use the general Rate Conservation principle [2] as carried out in Section 5.1. However we cannot yet answer intriguing questions about uniqueness of the stationary solutions in general. A deep open question is whether there exists a stationary solution for our dynamics in which queue lengths are almost-surely finite but of infinite mean. This is of interest both mathematically and practically, since such solutions would indicate to long range spatial clustering in the network, which may be an undesirable operating point.

2 Related Work

Our work has connections with three different branches of literature which we identify here.

2.1 Wireless Performance Analysis

There are now two parallel paradigms for performance analysis of wireless networks - a queuing theoretic approach and a stochastic geometric approach. We view our paper as a bridge between these two different modeling ideas. We provide a *tractable* queuing model while still retaining the important properties of the geometric representation of interference in wireless networks. The typical queuing models (for eg. [16],[19]) considers each queue to be a link. Thus naturally, the SINR seen at any queue is only a function of which other queues are busy. In contrast, in our model, a queue represents a region of space and queue length denotes the number of transmitting devices in that region of space. Hence in our model, the SINR at any link in a queue is a function of the entire queue lengths and not only on the busy/idle status of the queues. Moreover, in our model, each arriving link has a *single file* of a certain size and the link vanishes after completion of that single file. Thus the arrival process is that of links, each of which bring a single file in a random location of space. Such a model is justified in many emerging applications where the links are mobile and typically transmit short flows spaced out in time. From an algorithmic perspective, the notion of scheduling in our model will be very different from that considered for example in [19]. It is no longer a queue's service that we need to control, but rather a mechanism for each *individual customer* of a queue to make decisions about the rate of communication, or the degree of interference cancellation etc. Our model thus provides a testbed to understand network level performance gains seen when individual links are equipped with such distributed protocols making it more suitable for performance analysis as compared to the setting of [19]. Such user centric flow models were previously introduced in the context of cellular networks by [6],[7]. However, those models are intractable even for the smallest network comprising of just 2 base-stations and one needs to resort to bounds and approximations. In contrast, our model is meant to capture the ad-hoc network setting and we show is tractable even in the case of infinitely many queues.

From a methodological point of view, this work builds on the model in [18], which was also introduced as a way to bridge the stochastic geometry and queuing theoretic approaches for wireless networks. Nonetheless, we establish much stronger mathematical results and prove a stability phase-transition for the infinite network, while the techniques of [18] can only be applied to networks with finitely many queues. This requires fundamentally different tools, for establishing tightness of the marginal for queue length at steady-state in 'space truncated systems'. As already explained, we do so by applying the Rate Conservation principle to spatially finite restrictions of our model. We then show that the infinite network performance can be computed by taking the limit of spatial restrictions to the infinite grid. Our achievement in this regard is that we establish that one an 'exchange the limits in space and time' using very elementary monotonicity and coupling arguments. We note that the Rate Conservation method has been applied for infinite network performance assessment previously in [5] and [4]. However, the models in those papers are fundamentally very different from the model considered here as those models are intrinsically stable and are 'repulsive'. In contrast, our model is 'attractive', i.e. the neighboring queues are positively correlated and can be potentially unstable. Furthermore, the departure rate in our model is a non-linear function of the queue lengths, while that in [4] and [5] is a linear function of the state. Nevertheless, we show in this paper that a suitable application of the Rate Conservation argument is fruitful. We chose to present our analysis in discrete space and the language of queues rather than in continuum as done in [18] and [5] to illustrate the main ideas without technical overhead needed to discuss the continuum. We believe there is no fundamental difficulty now in porting the arguments in this paper to the continuum and we leave that as future work.

2.2 Large-Scale Queuing Systems

More broadly, we find it useful to connect our model and results to other *mean-field* type queuing networks that model large-scale infrastructure such as content-delivery (for eg. [21]), cloud-storage (for eg, [23]), computing clusters (for eg. [17]), and in wireless systems (for eg. [8]) to name a few. In these 'mean-field' type networks, one starts with a finite number n of queues, defines the dynamics and then considers the asymptotic as $n \to \infty$. This is the right asymptotic in the sense that it models applications accurately, and also is mathematically tractable. In most such models, the finite n case exhibits correlations among the queue lengths thereby making them difficult to analyze. However, in the limit as $n \to \infty$, one typically shows that there is 'propagation of chaos'. This then gives that the queue lengths become independent as $n \to \infty$. This independence can then be leveraged to write evolution equations for the limiting dynamics which can be analyzed. Our model differs fundamentally from the above models in many aspects. First, unlike the models described above, we can directly define the limiting infinite object, i.e., a model with infinitely many queues. Thus, we present our model with infinite number of queues, which represents the asymptotic of a large wireless network, much like the large-system asymptotic considered in the mean-field type models. Secondly and more crucially, our infinite model will *not* exhibit any independence properties, i.e., queue lengths will be positively correlated even in the infinite model, which forces us to use different techniques to study this model. Our main technical achievement in this context is to introduce coupling and Rate Conservation techniques to study this model of infinite queues without relying on any independence properties. The key structural difference between our model and the models where there is asymptotic independence is that the interactions in our setting are 'dense' and 'local', whereas in the mean-field models of Vvedenskaya et.al. [24] and others, the interactions are 'weak' and 'global'. By 'dense' we mean that the neighboring queues in our present context interact through the sequence $\{a_i\}_{i \in \mathbb{Z}^d}$, i.e., the interactions are strong and independent of the network size. However, two queues 'far away' in space do not interact with each other at all since $\{a_i\}_{i\in\mathbb{Z}^d}$ is finitely supported. Hence, the graph on which the queues are located plays an important role in our model. In contrast, in the mean-field models with n queues, every pair of queues 'interacts' with a probability proportional to $1/n$, i.e., there is no graph structure and each arrival interacts with every pair. Thus, we view this mean field model as based on 'weak' (interacts at rate $1/n$) and 'global' interactions (every pair interacts equally). The weak global interactions leads to asymptotic independence, while the localized nature of interactions in our model does not vanish even in the infinite system limit.

2.3 Interacting Particle System

Another viewpoint for our model is that of an 'Interacting Particle System' in the sense of [15], with an important distinction that the state-space for our particles is infinite, i.e., is $\mathbb{N}$, as opposed to the models studied in [15] wherein the particles have finite state-spaces. Here, we view each queue as a 'particle' and the state of a particle at some time is the number of customers in the queue. The particle states evolve as a Markov process with local dependency, i.e. the departure rate only depends on the 'local geometry' around the queue. This connection also implies that the question of uniqueness of stationary solutions could potentially have non-trivial phase transitions, similar to the Ising and voter models considered in [15]. Exploring the connections between our model and those studied in [15] will be interesting future work.

3 Mathematical Preliminaries

In this section, we setup the mathematical problem studied in this paper. We first give a precise definition of our queuing model and subsequently define the notion of stochastic stability of the infinite network (in Definition 4) we are interested in.

3.1 Framework

We first make our model described in Section 1.1 mathematically precise. We assume there exists an abstract probability space $(\Omega, \mathcal{F}, \mathbb{P})$ that contains the stationary and ergodic driving sequences $(\mathcal{A}_i, \mathcal{D}_i)_{\imath\in\mathbb{Z}^d}$. For each $i \in \mathbb{Z}^d$, $\mathcal{A}_i$ is a Poisson Point Process of intensity λ on $\mathbb{R}$, independent of everything else and $\mathcal{D}_i$ is a Poisson Point Process of intensity 1 on $\mathbb{R}\times[0,1]$, independent of everything else. Our stochastic process denoting the queue lengths $t \to \{x_i(t)\}_{i\in\mathbb{Z}^d}$ will be constructed as a factor of the process $(\mathcal{A}_i, \mathcal{D}_i)_{\imath\in\mathbb{Z}^d}$. The process $\mathcal{A}_i := \sum_{q\in\mathbb{Z}} \delta_{A_q^{(i)}}$ encodes the fact that, at times $\{A_q^{(i)}\}_{q\in\mathbb{Z}}$, there is an arrival of a customer in queue i. Thus the arrivals to queues form PPPS of intensity λ and are independent of everything else. The process $\mathcal{D}_i := \sum_{q\in\mathbb{Z}} \delta_{(D_q^{(i)}, U_q^{(i)})}$ encodes that there is a possible departure from queue i at time $D_q^{(i)}$, with an additional independent $U[0,1]$ random variable provided by $U_q^{(i)}$. A customer if any is removed from queue i at times $D_q^{(i)}$ if $U_q^{(i)} \le \frac{x_i(D_q^{(i)})}{\sum_{j\in\mathbb{Z}^d} a_{j-i}x_j(D_q^{(i)})}$. In other words, we remove a customer from queue i at time $D_q^{(i)}$ with probability $\frac{x_i(D_q^{(i)})}{\sum_{j\in\mathbb{Z}^d} a_{j-i}x_j(D_q^{(i)})}$, independently of everything else. Thus we see that the instantaneous rate of departure from any queue $i \in \mathbb{Z}^d$ and time $t \in \mathbb{R}$ is $\frac{x_i(t)}{\sum_{j\in\mathbb{Z}^d} a_j x_{i-j}(t)}$, independently of everything else. Observe that since $a_0 = 1$, if $x_i(t) > 0$, then necessarily, $\frac{x_i(t)}{\sum_{j \in \mathbb{Z}^d} a_j x_{i-j}(t)} \in [0,1]$.

We further assume that we are equipped with a family $(\theta_x)_{x \in \mathbb{R}}$ of measurable bijective functions from Ω to itself which denotes the 'shift-operator' by the vector x. More precisely $(\mathcal{A}_i, \mathcal{D}_i)_{\imath \in \mathbb{Z}^d} \circ \theta_x$ is the same driving sequence where each of the arrivals is shifted by the vector x in all queues, i.e., if $\mathcal{A}_i := \sum_{q \in \mathbb{Z}} \delta_{A_q^{(i)}}$ and $\mathcal{D}_i := \sum_{q \in \mathbb{Z}} \delta_{(D_q^{(i)}, U_q^{(i)})}$, then $\mathcal{A}_i \circ \theta_x := \sum_{q \in \mathbb{Z}} \delta_{A_q^{(i)} - x}$ and $\mathcal{D}_i \circ \theta_x := \sum_{q \in \mathbb{Z}} \delta_{(D_q^{(i)} - x, U_q^{(i)})}$, for all $i \in \mathbb{Z}^d$. We also assume that the system $(\mathbb{P}, (\theta_x)_{x \in \mathbb{R}})$ is ergodic, i.e. if for some event $A \in \mathcal{F}$ it holds that $\mathbb{P}[A \triangle A \circ \theta_x] = 0$ for all $x \in \mathbb{R}$, then it must be the case that $\mathbb{P}[A] \in \{0,1\}$. This is not a restriction since the canonical space of these sequences satisfies this property.

3.2 Construction of the Process

Before preceeding to analyze the above model, one needs to ensure that it is 'well-defined'. We mean that our model is well defined if given the initial network state $\{x_i(0)\}_{i \in \mathbb{Z}^d}$, any time $T \geq 0$ and any index $k \in \mathbb{Z}^d$, we are able to construct the queue length $x_k(T)$ unambiguously and exactly. In the case of finite networks (i.e., networks with finitely many queues), the simulation is trivial: almost-surely, one can order all possible events in the network with increasing time, and then update the network state sequentially using the evolution dynamics described above. Such a scheme works unambiguously since, almost-surely, all event times will be distinct and in any interval $[0, T]$, there will be finitely many events. However the main difficulty in the case of infinite networks is that there is no *first-event* in the network, and hence one cannot simulate by ordering all possible events. In other words, in any arbitrarily small interval of time, infinitely many events will occur almost-surely and hence we cannot construct by ordering all the events in the network. However we show in Theorem 31 below in Appendix A that in order to determine the value of any arbitrary queue $k \in \mathbb{Z}^d$ at any time $T \geq 0$, we can effectively restrict our attention to an almost-surely finite subset $X_{k,T} \subset \mathbb{Z}^d$ and determine $x_k(T)$ by restricting the dynamics to $X_{k,T}$ in the interval $[0, T]$. This is then easy to construct as it is a finite system, thereby we can determine $x_k(T)$ unambiguously. We defer the technical details of the construction to Appendix A.

3.3 Monotonicity

In this section, we establish an obvious but an extremely useful property of pathwise monotonicity satisfied by the dynamics. Note that our model is not monotone separable in the sense of [3] since the dynamics does not satisfy the external monotonicity condition. We state our lemmas and the key trick for the proof and defer the details to Appendix B.

Lemma 1. *If we have two initial conditions $\{x_i^{'}(0)\}_{i \in \mathbb{Z}^d}$ and $\{x_i(0)\}_{i \in \mathbb{Z}^d}$ such that for all $i \in \mathbb{Z}^d$, $x_i^{'}(0) \geq x_i(0)$, then for almost-every driving sequence $(\mathcal{A}_i, \mathcal{D}_i)_{i \in \mathbb{Z}^d}$, we have that $x_i^{'}(T) \geq x_i(T)$ for all $i \in \mathbb{Z}^d$ and all $T \geq 0$.*

The proof is by a path-wise coupling argument, where the two different initial conditions are driven by the same arrival and potential departures. The key trick is to notice the following. At arrival times, the ordering will trivially be maintained. Consider some queue i and time t where there is a potential departure. If $x_i^{'}(t) \geq x_i(t) + 1$, then, since at-most one departure occurs, the ordering will be maintained. But if $x_i^{'}(t) = x_i(t)$, then the rates $\frac{x_i^{'}(t)}{\sum_{j \in \mathbb{Z}^d} a_j x_{i-j}^{'}(t)} \leq \frac{x_i(t)}{\sum_{j \in \mathbb{Z}^d} a_j x_{i-j}(t)}$ and hence the ordering will again be maintained. This observation can be leveraged again to have the following form of monotonicity.

Lemma 2. *For all initial conditions $\{x_i(0)\}_{i\in\mathbb{Z}^d}$, for all $0 \le s \le t \le \infty$, all $X \subset \mathbb{Z}^d$, and all $T > 0$, $\{x_i(T)\}_{i\in\mathbb{Z}^d}$ is coordinate-wise larger in the true dynamics than in the dynamics constructed by setting $\mathcal{A}_j([s,t]) = 0$ for all $j \in X$.*

3.4 Stochastic Stability

We establish a $0-1$ law stating that either all queues are transient or all queues are recurrent (Lemma 3). Thus, we can then claim that the entire network is stable if and only if any (say queue indexed 0 without loss of generality) is stable (Definition 4).

To state the lemmas, we set some notation. Let $T \ge 0$ and $s > -T$ be arbitrary and finite. Denote by $\{x_{i;T}(s)\}_{i\in\mathbb{Z}^d}$ the value of the process seen at time s when started with the empty initial state at time $-T$, i.e., with the initial condition of $x_{i;T}(-T) = 0$ for all $i \in \mathbb{Z}^d$, and then running the construct procedure described in Theorem 31 from $-T$ till s. Lemma 1 implies that for every queue $i \in \mathbb{Z}^d$, and for $\mathbb{P}$ almost-every $\omega \in \Omega$, we have $T \to x_{i;T}(s)$ is non-decreasing for every fixed s. Thus, for every i, and every $s \in \mathbb{R}$, there exists an almost-sure limit $\lim_{T\to\infty} x_{i:T}(s) := x_{i;\infty}(s)$.

Lemma 3. *We have either $\mathbb{P}[\cap_{i\in\mathbb{Z}^d}\{x_{i;\infty}(0) = \infty\}] = 1$ or $\mathbb{P}[\cap_{i\in\mathbb{Z}^d}\{x_{i;\infty}(0) < \infty\}] = 1$.*

The proof follows from standard shift-invariance arguments which we present here for completeness.

Proof. It suffices to first show that for any fixed $i \in \mathbb{Z}^d$, we have $\mathbb{P}[x_{i;\infty}(0) < \infty] \in \{0,1\}$. Assume that we have established for some i (say 0 without loss of generality that) $\mathbb{P}[x_{0;\infty}(0) < \infty] \in \{0,1\}$. From the translation invariance of the dynamics, it follows that, for all $i \in \mathbb{Z}^d$, we have $\mathbb{P}[x_{i;\infty}(0) < \infty] = \mathbb{P}[x_{0;\infty}(0) < \infty]$. Thus, if $\mathbb{P}[x_{0;\infty}(0) < \infty] = 1$, then $\mathbb{P}[\cap_{i\in\mathbb{Z}^d} x_{i;\infty}(0) < \infty] = 1$. Similarly, if $\mathbb{P}[x_{0;\infty}(0) = \infty] = 1$, then $\mathbb{P}[\cap_{i\in\mathbb{Z}^d} x_{i;\infty}(0) = \infty] = 1$. Thus to prove the lemma, it suffices to prove that $\mathbb{P}[x_{0;\infty}(0) < \infty] \in \{0,1\}$.

The key observation is that the event $A := \{\omega \in \Omega : x_{0;\infty}(0) < \infty\}$ is invariant under shifts θ_x for all $x \in \mathbb{R}$. To show this, first notice that from elementary properties of PPP, we have that for every $i \in \mathbb{Z}^d$ and every compact set $B \subset \mathbb{R}$, $\mathcal{A}_i(B) < \infty$ a.s.. Now for any $x \ge 0$, we have $x_{0;\infty}(0) \circ \theta_x \le x_{0;\infty}(0) + A_0([0,x])$, which is finite almost-surely if $x_{0;\infty}(0) < \infty$ almost-surely. Similarly for every $x < 0$, $x_{0;\infty}(0) = x_{0;\infty}(0) \circ \theta_x + A_0([x,0])$, which again implies that $x_{0;\infty}(0) \circ \theta_x$ is almost-surely finite if $x_{0;\infty}(0) < \infty$. Hence ergodicity implies that $\mathbb{P}[A] \in \{0,1\}$ and thus the lemma is proved.

□

In light of this, the following definition of stability is natural.

Definition 4. *The system is **stable** if $x_{0;\infty}(0) < \infty$ almost-surely. Conversely, we say the system is **unstable** if $x_{0;\infty}(0) = \infty$ almost-surely.*

Observe that the definition of stability does not require $\mathbb{E}[x_{0;\infty}(0)]$ to be finite.

3.5 Stationary Solutions

The following definition formalizes the notion of stationary solution to our dynamics.

Definition 5. *A probability measure* $\boldsymbol{\pi}$ *on* $(\mathbb{Z}^d)^{\mathbb{N}}$ *is said to be invariant for the dynamics* $\{x_i(t)\}_{i\in\mathbb{Z}^d}$ *if, whenever* $\{x_i(0)\}_{i\in\mathbb{Z}^d}$ *is distributed according to* $\boldsymbol{\pi}$ *independently of everything else, then, for all* $t \geq 0$*, the random variables* $\{x_i(t)\}_{i\in\mathbb{Z}^d}$ *are also distributed as* $\boldsymbol{\pi}$*.*

Since our network is not a finite-dimensional Markov process, stability in the sense of Definition 4 does not imply ergodicity in the Markov chain sense. In particular, it does not imply that stationary distributions are unique, and starting from any initial condition on $\mathbb{N}^{\mathbb{Z}^d}$, the queue lengths converge in some sense to the minimal stationary distribution considered in Definition 4. Stability only implies the *existence* of a stationary solution namely the law of $\{x_{i\infty}(0)\}_{i\in\mathbb{Z}^d}$ is an invariant measure for the dynamics. However, uniqueness is not granted, and one of our main results bears on this.

4 Main Results

4.1 Existence of a Stationary Solution

The following theorem is the central result of our paper.

Theorem 6. *If* $\lambda < \frac{1}{\sum_{i\in\mathbb{Z}^d} a_i}$*, then the process* $\{x_i(\cdot)\}_{i\in\mathbb{Z}^d}$ *is stable and*

$$\mathbb{E}[x_{0;\infty}(0)] = \frac{\lambda}{1 - \lambda \sum_{i\in\mathbb{Z}^d} a_i}. \tag{1}$$

The proof of this theorem is split between Sections 5 and Section 6. In Section 5, we consider large space-truncated finite system versions of our model and prove the *exact same result* for this case. The finite network will be a finite-dimensional Markov chain and we analyze its positive recurrence by coupling and fluid-like arguments. We also employ a Rate Conservation argument to conclude that the marginal distribution of queue sizes as space is truncated forms a tight sequence. In Section 6, we then leverage monotonicity and tightness to argue that one can switch limits, i.e. consider the steady state distribution of the infinite network as a limit of the steady state distribution of space-truncated systems. As far as the converse goes, we make the following conjecture.

Conjecture 7. *If* $\lambda > \frac{1}{\sum_{j\in\mathbb{Z}^d} a_j}$*, then the system is unstable, i.e.* $x_\infty(0) = \infty$ *a.s.*

We are however unable to prove this conjecture in general. We however provide a proof in the case $d = 1$ and when the interaction sequence is such that $a_i = 0$ for all $i \notin \{-1, 0, 1\}$ and $a_i = 1$ if $i \in \{-1, 0, 1\}$. We have the following theorem.

Theorem 8. *In this special case, if* $\lambda > \frac{1}{3}$*, then the queuing system is unstable.*

Note that $\lambda_c = 1/3$ in this important special case, and hence our characterization is tight, at-least in this special case. We provide the proof in Appendix D. The key novelty there is in the construction of a 'triangular Lyapunov function' to establish transience.

4.2 Uniqueness of Stationary Solutions

We establish a weak form of uniqueness as a corollary to the following proposition.

Proposition 9. *Let $\{x_i(0)\}_{i\in\mathbb{Z}^d}$ be distributed according to any invariant measure $\boldsymbol{\pi}$ of the dynamics. Then,*

$$\mathbb{E}[x_0(0)] \in \left\{ \frac{\lambda}{1-\lambda\sum_{i\in\mathbb{Z}^d} a_i}, \infty \right\}. \tag{2}$$

The above proposition along with Theorem 6 yields the following statement on the uniqueness of invariant measures.

Corollary 10. *There is a unique invariant measure for the dynamics $\{x_i(t)\}_{i\in\mathbb{Z}^d}$ that has finite mean.*

We give the proof of this corollary in Section 7. The proof again relies on monotonicity and backward coupling arguments. In words, the only possible stationary regime for our dynamics with a finite mean queue length is the minimal one, namely the regime obtained by the backward construction which consists in looking at the state at time 0 when starting empty at time $-\infty$. There may be other stationary regimes, but necessarily with infinite mean. A deep open question is whether there exist stationary solutions that have infinite mean queue lengths.

5 Space Truncated Torus Systems

In this section, we study a finite version of the aforementioned infinite queuing network. For any $n \in \mathbb{Z}_+$, the n-truncated system is that obtained by restricting the dynamics to the set $B_n(0)$, the L_∞ ball of radius n centered at 0, with the edges 'wrapped around' to form a torus. More precisely, in this section, we consider a stochastic process $\mathbf{x}^{(n)}(t) := \{x_i^{(n)}(t)\}_{i\in B_n(0)}$, where for each $i \in B_n(0)$, $x_i^{(n)}(t)$ denotes the number of customers in queue i at time t in this n-truncated system. For notational ease, we refer to the set B_n to be $B_n(0)$, i.e. we drop the center as being the origin. The driving data for the n-truncated process is the same as in the infinite one, but restricted to the set B_n, i.e., $(\mathcal{A}_i, \mathcal{D}_i)_{i\in B_n}$. The departure dynamics is as before, but we treat the set B_n as a torus. More precisely, given any two $i, j \in B_n$, define $d_n(i,j) := (i-j) \mod n$, where the modulo operation is coordinate-wise. Thus, at any time t, and any $i \in B_n$, the rate at which a departure occurs from queue i at time t in the process $\{x_i^{(n)}(t)\}_{i\in B_n(0)}$ is $\frac{x_i^{(n)}(t)}{\sum_{j\in B_n} a_{d_n(i,j)} x_j^{(n)}(t)}$. Since n is finite, the stochastic process $\mathbf{x}^{(n)}(t)$ is a continuous time Markov process on a countable state-space, i.e., on $\mathbb{N}^{(2n+1)^d}$. Moreover, since the jumps are triggered by a finite number of Poisson processes, this chain has almost-surely no-explosions. The following theorem gives a sufficient condition for positive recurrence. The proof of this theorem is the main goal of the present section.

Theorem 11. *For all $n > L$, and for all $\lambda < \frac{1}{\sum_{j\in\mathbb{Z}^d} a_j}$, the Markov process $\{x_i^{(n)}(t)\}_{i\in B_n(0)}$ is positive recurrent. Let $\pi^{(n)}$ denote the stationary queue length distribution on $\mathbb{N}$ of any queue $i \in B_n(0)$ and let Z be distributed as $\pi^{(n)}$. Then there exists a $c > 0$ such that $\mathbb{E}[e^{cZ}] < \infty$.*

Remark 12. *The symmetry in the torus implies that the marginal stationary queue length distribution of any queue i, $\pi^{(n)}$, is the same for all i.*

Remark 13. *The existence of an exponential moment in particular yields that all power moments of $\pi^{(n)}$ are finite.*

Proof Sketch

We provide a sketch here and defer the details to the Appendix C. The proof is technical with a lot of details, and hence summarizing it here will be useful in understanding the key ideas involved. To prove the theorem, we will define a modified dynamics $\{\tilde{x}_i^{(n)}(t)\}_{t\geq 0, i\in B_n(0)}$ which is coupled with the evolution of $\{x_i^{(n)}(t)\}_{t\geq 0, i\in B_n(0)}$. We construct the modified dynamics such that it satisfies $\tilde{x}_i^{(n)}(t) \geq x_i^{(n)}(t)$ a.s. for all $i \in B_n(0)$ and $t \geq 0$. We do this by discretizing continuous time to discrete by choosing sufficiently small interval h, i.e. times $\cdots, -h, 0, h, 2h, \cdots$ will form time slot boundaries. We then restrict departures so that at-most one departure can occur in a time period. We also modify the arrivals so that in any time slot, the difference between the maximum number of arrivals and the minimum number of arrivals in a time slot is at-most a constant. From monotonicity, the dynamics with such modified arrivals and departures can be coupled to provide an upper bound to the true queue lengths. We describe in detail this construction in Appendix C.1. We further identify a large r, and *equalize* the queues after every r time-slots, i.e. at times $\cdots, -rh, 0, rh, 2rh, \cdots$, we add fictious customers so that all queues have the same number of customers. If the number of customers is smaller than a constant y_0, we further add more customers till every queue has at-least y_0 customers. Thus, at the end of every r time-slots, every queue has the same number of customers which is at-least y_0. From a coupling argument, we show that after the addition of the fictitious customers, the queue length follows the trajectory of an appropriately modified $GI/GI/1$ queue which is stable. Thus, we have dominated our process $\{x_i^{(n)}(t)\}_{i\in\mathbb{Z}^d}$ so that every one of them is dominated from above by a stable $GI/GI/1$ queue, and hence $\{x_i^{(n)}(t)\}_{i\in\mathbb{Z}^d}$ is positive recurrent.

5.1 Rate Conservation Argument

Here, we will rely on Theorem 11 to derive a closed form expression for the mean queue length. From Theorem 11, we know that if $\lambda < \frac{1}{\sum_{j\in\mathbb{Z}^d} a_j}$, then, there exists a unique probability measure $\boldsymbol{\pi}^{(n)}$ on $\mathbb{N}^{|B_n(0)|}$ corresponding to the stationary distribution of $\{x_i^{(n)}\}_{i\in B_n(0)}(t)$. From Remark 12, we know that the measure $\boldsymbol{\pi}^{(n)}$ has equal marginals which we denote by $\pi^{(n)}$. We further denote by $\mu^{(n)}$ the first moment of $\pi^{(n)}$. Remark 13 states that $\mu^{(n)} < \infty$ for all n. The main goal of this subsection is to derive an explicit formula for $\mu^{(n)}$ as stated in the following lemma.

Lemma 14. *For all $\lambda < \frac{1}{\sum_{j\in\mathbb{Z}^d} a_j}$,*

$$\mu^{(n)} = \frac{\lambda a_0}{1 - (\sum_{j\in\mathbb{Z}^d} a_j)\lambda}. \tag{3}$$

Remark 15. *Note that we assumed $a_0 = 1$ in the model. For completeness, we give the derivation for any general $a_0 > 0$.*

This lemma in particular yields that the mean number of customers in steady state of the space truncated torus is independent of n, provided n is large enough, i.e., $\sup_n \mu^{(n)} < \infty$.

Proof. We use the general Rate Conservation principle [2]. Consider the system in steady state, i.e., assume that $\{x_i^{(n)}(0)\}_{i\in B_n(0)}$ to be distributed according to $\boldsymbol{\pi}^{(n)}$. Consider the stochastic process $A_t := x_0^{(n)}(t)(\sum_{j\in\mathbb{Z}^d} a_j x_j^{(n)}(t))$, for $t \geq 0$. Since $\{x_i^{(n)}(0)\}_{i\in B_n(0)}$ is distributed as $\boldsymbol{\pi}^{(n)}$, the quantity $\mathbb{E}[A_t]$ must be a constant that does not depend on t. Indeed since $\{x_i^{(n)}(0)\}_{i\in B_n(0)}$ is distributed according to the stationary measure, the process A_t is also a stationary process and hence its expectation is the same for all time t. For the rest of this proof, let $\{x_i^{(n)}\}_{i\in B_n(0)} := \{x_i^{(n)}(0)\}_{i\in B_n(0)}$, i.e., we drop the time index 0 and the random variables $\{x_i^{(n)}\}_{i\in B_n(0)}$ is distributed according to the measure $\boldsymbol{\pi}^{(n)}$, independently of everything else. The following proposition is the essence of the Rate Conservation statement.

Proposition 16. *Let $\{x_i^{(n)}(0)\}_{i\in B_n(0)}$ be distributed according to $\boldsymbol{\pi}^{(n)}$, independently of everything else. Then for all $t \geq 0$,*

$$\lambda a_0 + 2\lambda \left(\sum_{j\in B_n} a_j\right)\mu^{(n)} = \mathbb{E}\left[R(0)\left(a_0(2x_0^{(n)}(0)-1) + \sum_{i\in B_n\setminus\{0\}} a_i x_i(0)\right) + \sum_{i\in B_n\setminus\{0\}} R(i) a_i x_0^{(n)}(0)\right], \tag{4}$$

where for any $i \in B_n$,

$$R(i) := \frac{x_i^{(n)}(0)}{\sum_{j\in B_n} a_{d_n(i,j)x_j^{(n)}(0)}}.$$

We postpone the rigorous proof and provide an interpretation here. In a small interval of time dt, there will be at-most one of either an arrival or a potential departure from queue 0 and its neighbors. With probability λdt, there will be an arrival in any queue 0 or its neighboring queues. On an arrival at queue 0, the increase in the quantity A_0 is $\mathbb{E}[(x_0^{(n)}+1)(a_0(x_0^{(n)}+1) + \sum_{j\in B_n\setminus\{0\}} a_j x_j^{(n)}) - x_0^{(n)}(\sum_{j\in B_n} a_j x_j^{(n)})]$, which is equal to $\mathbb{E}[a_0 + \sum_{j\in B_n} a_j x_j^{(n)}]$. Similarly, the average increase in A_0 due to an arrival in the neighboring queues is $\mathbb{E}[(x_0^{(n)})(a_i(x_i^{(n)}+1) + \sum_{j\in B_n\setminus\{i\}} a_j x_j^{(n)}) - x_0^{(n)}(\sum_{j\in B_n} a_j x_j^{(n)})]$, which is equal to $\mathbb{E}[a_i x_0^{(n)}]$. The chance that there are two or more arrivals is $\mathrm{O}(dt^2)$, which is small. Thus, the average increase due to arrivals is $\lambda dt \mathbb{E}[a_0(x_0^{(n)}+1) + \sum_{j\in B_n} a_j x_j^{(n)} + \sum_{j\in B_n\setminus\{0\}} a_j x_0^{(n)}] + \mathrm{O}(dt^2)$. After simplification, and using the fact that the variables $x_j^{(n)}$ all have the same mean, we get that the average increase in time dt is $\lambda dt \left(a_0 + 2\mu^{(n)}\left(\sum_{j\in B_n} a_j\right)\right) + \mathrm{O}(dt^2)$.

Likewise, with probability $R(i)dt$, there will be a departure from queue i. When a customer leaves from queue 0, the average decrease in A_0 is then $\mathbb{E}[(a_0((x_0^{(n)})^2 - a_0(x_0^{(n)}-1)^2 + \sum_{i\in B_n\setminus\{0\}} a_i x_i^{(n)})]$. Similarly, a departure from queue i results in an average decrease in A_0 of

$\mathbb{E}[a_i x_0^{(n)}]$. Thus, the total average decrease in A_0 due to departures is

$$dt\mathbb{E}\left[R(0)\left(a_0(2x_0^{(n)}-1)+\sum_{i\in B_n\setminus\{0\}} a_i x_i^{(n)}\right)+\sum_{i\in B_n\setminus\{0\}} R(i)a_i x_0^{(n)}\right]+\mathrm{O}\big(dt^2\big).$$

Since A_t is a stationary process, the Rate Conservation principle gives

$$\lambda\left(a_0+2\mu^{(n)}\left(\sum_{j\in\mathbb{Z}^d} a_j\right)\right)=\mathbb{E}\left[R(0)\left((a_0(2x_0^{(n)}-1)+\sum_{i\in B_n\setminus\{0\}} a_i x_i^{(n)}\right)+\sum_{i\in B_n\setminus\{0\}} R(i)a_i x_0^{(n)}\right].$$

Now, we use the following version of the Mass-Transport Theorem for unimodular random graphs:

Proposition 17. *The following formula holds.*

$$\mathbb{E}\left[\sum_{i\in B_n\setminus\{0\}} R(i)a_i x_0^{(n)}\right]=\mathbb{E}\left[\sum_{i\in B_n\setminus\{0\}} R(0)a_i x_i^{(n)}\right]. \tag{5}$$

Proof. The proof follows from the standard argument of Mass-Transport Theorem involving swapping double sums. Observe from the definition of the dynamics, the queue lengths $\{x_k^{(n)}\}_{k\in B_n}$ is translation invariant on the torus B_n. Hence, for all $j\in B_n$, $x_j^{(n)}\sum_{i\in B_n\setminus\{j\}} R(i)a_{i-j}$ are identically distributed, and in particular have the same means. The proposition is now proved thanks to the following calculations.

$$\begin{aligned}
\mathbb{E}\left[\sum_{i\in B_n\setminus\{0\}} R(i)a_i x_0^{(n)}\right] &= \frac{1}{|B_n|}\mathbb{E}\left[\sum_{j\in B_n} x_j^{(n)}\sum_{i\in B_n\setminus\{j\}} R(i)a_{i-j}\right],\\
&\stackrel{(a)}{=} \frac{1}{|B_n|}\mathbb{E}\left[\sum_{i\in B_n} R(i)\sum_{j\in B_n\setminus\{i\}} a_{i-j}x_j^{(n)}\right],\\
&\stackrel{(b)}{=} \frac{1}{|B_n|}\mathbb{E}\left[\sum_{i\in B_n} R(i)\sum_{j\in B_n\setminus\{i\}} a_{j-i}x_j^{(n)}\right],\\
&\stackrel{(c)}{=} \mathbb{E}\left[\sum_{i\in B_n\setminus\{0\}} R(0)a_i x_i^{(n)}\right].
\end{aligned}$$

Equality (a) follows by swapping the order of summations, which is licit since they each contain finitely many terms. Equality (b) follows since $a_k=a_{-k}$ for all $k\in\mathbb{Z}^d$. Equality (c) again follows from the fact that for all $i\in B_n$, $R(i)\sum_{j\in B_n\setminus\{i\}} a_{j-i}x_j^{(n)}$ are identically distributed. This is a consequence of the queue lengths $\{x_k^{(n)}\}_{k\in B_n}$ being translation invariant on the torus. □

Intuitively, Equation (5) can be interpreted by considering the finite graph with vertices on the torus B_n with a directed edge from i to j in B_n with weight $R(i)a_{d_n(i-j)}x_j$. This random graph, when rooted in 0, is unimodular and hence Mass-Transport holds. Since $a_i = a_{-i}$, we get that the average decrease is $\mathbb{E}[-a_0R(0) + 2R(0)\sum_{i\in B_n} a_i x_i^{(n)}]$. Now, $\mathbb{E}[R(0)] = \lambda$, and since, for all $i \in B_n$, $\mathbb{E}[x_i^{(n)}] = \mu^{(n)}$,

$$2\lambda(\sum_{i\in B_n} a_i)\mu^{(n)} + 2\lambda a_0 = \mathbb{E}[2R(0)(\sum_{i\in B_n} a_i x_i^{(n)})]. \tag{6}$$

But since $R(0)(\sum_{i\in B_n} a_i x_i^{(n)}) = x_0$, we get

$$\mu^{(n)} = \frac{\lambda a_0}{1 - (\sum_{i\in\mathbb{Z}^d} a_i)\lambda}.$$

□

In the following immediate corollary, $\pi^{(n)}$ denotes the probability measure on $\mathbb{N}$ corresponding to the marginal on queue 0 of the measure $\boldsymbol{\pi}^{(n)}$.

Corollary 18. *If $\lambda < \frac{1}{\sum_{j\in\mathbb{Z}^d} a_j}$, then the sequence of probability measures $\{\pi^{(n)}\}_{n>L}$ is tight.*

Proof. From Markov's inequality, we have

$$\mathbb{P}[X > Q] \le \frac{1}{Q}\left(\frac{\lambda a_0}{1 - (\sum_{i\in\mathbb{Z}^d} a_i)\lambda}\right),$$

where X is distributed according to $\pi^{(n)}$. Thus, for every $\epsilon > 0$, we can find Q large such that

$$\sup_{n>L} \mathbb{P}_{\pi^{(n)}}[X > Q] < \epsilon.$$

□

6 Stability of the Infinite Network- Proof of Theorem 6

The goal of this section is to conclude the proof of Theorem 6. We establish stability in Corollary 25 and the mean queue length in Corollary 27. The key idea is to use monotonicity and the backward coupling representation as carried out in Lemma 24.

In order to prove this, we need some additional notation. For any $T > 0$ and $n \in \mathbb{N}$ such that $n > L$, and any $i \in \mathbb{Z}^d$, we define the random variables $x_{i;T}^{\infty}(0), y_{i;T}^{(n)}(0)$ and $z_{i:T}^{(n)}(0)$. These variables represent the number of customers in queue i at time 0 in three different dynamics which will be coupled and driven by the same arrival-departure process $(\mathcal{A}_i, \mathcal{D}_i)_{i\in\mathbb{Z}^d}$. In all of them, the subscript i refers to queue i and T refers to the fact that the system started empty at time $-T$. We now describe the three different dynamics in question:

1. $x_{i;T}(0)$ denotes the number of customers in queue i at time 0 in the true infinite dynamics as defined in Section 1.1.

2. $y^{(n)}_{i;T}(0)$ denotes the number of customers in queue i at time 0 for the dynamics restricted to the set $B_n(0)$ *viewed as a torus.* Hence $\{y^{(n)}_{i;T}(0)\}_{i\in B_n(0)}$ is the queue length of the process studied in Section 5.

3. $z^{(n)}_{i;T}(0)$ denotes the number of customers at time 0 for the dynamics restricted set B_n, not seen as a torus. In this dynamics, the 'edge effects' of the finite set $B_n(0)$ are taken into account. Thus for all $i \in B_n(0)^{\mathsf{c}}$, we have $z^{(n)}_{i;T}(0) = 0$, by definition.

The following two statements follow from monotonicity.

Proposition 19. *For all $T > 0$, all $n > L$, and all $i \in \mathbb{Z}^d$, we have $x_{i;T}(0) \geq z^{(n)}_{i;T}(0)$ and $y^{(n)}_{i;T}(0) \geq z^{(n)}_{i;T}(0)$ almost-surely.*

Moreover, due to monotonicity, we have:

Proposition 20. *Almost-surely, the following limits exist:*

$$x_{i;\infty}(0) := \lim_{T\to\infty} x_{i;T}(0),$$
$$y^{(n)}_{i;\infty}(0) := \lim_{T\to\infty} y^{(n)}_{i;T}(0),$$
$$z^{(n)}_{i;\infty}(0) := \lim_{T\to\infty} z^{(n)}_{i;T}(0).$$

Note that the distribution of the random variable $y^{(n)}_{0;\infty}$ is the marginal on queue 0 of the probability measure $\boldsymbol{\pi}^{(n)}$, whose existence was proved in Theorem 11 We also established in Corollary 18 that the sequence of probability measures $\{\pi^{(n)}\}_{n\in\mathbb{N}}$ is tight. Moreover, in view of Lemma 3, it suffices to show that queue 0 is stable to conclude that the entire network is stable. Hence for notational brevity, we will omit the queue and time index by adopting the following simplified notation: $x_T := x_{0;T}(0)$, $y^{(n)}_T := y^{(n)}_{i;T}(0)$, $z^{(n)}_T := z^{(n)}_{i;T}(0)$, where $T \in [0,\infty]$. The main argument for establishing stability is contained in the following propositions.

Proposition 21. *Almost-surely, for every $T \geq 0$, we have $\lim_{n\to\infty} z^{(n)}_T = x_T$.*

Proof. From Corollary 38, for every finite T, there exists a random subset $X \subset \mathbb{Z}^d$ which is almost-surely finite and such that the value of x_T can be obtained by restricting the dynamics to the set X in the time interval $[-T, 0]$. Let N be any integer such that X is properly contained in B_n. Then, for all $n \geq N$, $x_T = z^{(n)}_T$. □

Lemma 22. *The sequence $z^{(n)}_\infty$ is non-decreasing in n and almost-surely converges to a finite integer valued random variable denoted by $z^{(\infty)}_\infty$.*

Proof. Note that for all finite T, $z^{(n)}_T$ is non-decreasing in n. Thus for any $n > m$, we have $z^{(n)}_T \geq z^{(m)}_T$, for all T. Now, taking a limit in T on both sides, which we know exist, we see that $z^{(n)}_\infty \geq z^{(m)}_\infty$. This establishes the fact that $z^{(n)}_\infty$ is an non-decreasing sequence and hence the limit $\lim_{n\to\infty} z^{(n)}_\infty := z^{(\infty)}_\infty$ exists.

We now show the finiteness of $z_\infty^{(\infty)}$. Note that for all n and T, $z_T^{(n)} \leq y_T^{(n)}$. Now, taking a limit in T, we see that $z_\infty^{(n)} \leq y_\infty^{(n)}$. The distribution of the random variable $y_\infty^{(n)}$ is the probability measure $\pi^{(n)}$ on $\mathbb{N}$. From Corollary 18, the sequence $\{\pi_n\}$ is tight. Let $\tilde{\pi}^{(n)}$, $n \in \mathbb{N}$, denote the distribution of $z^{(n)}$. Thus the sequence $\{\tilde{\pi}^{(n)}\}_{n\in\mathbb{N}}$ is tight as well. Moreover due to monotonicity, $z_\infty^{(n)}$ converges almost-surely. But since the sequence $\{\tilde{\pi}^{(n)}\}_{n\in\mathbb{N}}$ is tight, we have that $z_\infty^{(n)}(0)$ almost-surely converges to a finite limit. ☐

Lemma 23. *There exists a random $N \in \mathbb{N}$, such that for all $n \geq N$, there exists a random $T_n \in \mathbb{R}^+$, such that for all $t \geq T_n$, $z_\infty^{(\infty)} = z_t^{(n)}$.*

Proof. From the previous lemma, $z_\infty^{(n)}$ converges almost-surely to a finite limit as $n \to \infty$. Since the random variables $\{z_\infty^{(n)}\}_{n\in\mathbb{N}}$ are integer valued, there exists a random N such that $z_\infty^{(\infty)} = z_\infty^{(n)}$, $\forall n \geq N$.

Now, since, for each T and n, $z_T^{(n)}$ is integer valued, the existence of an almost-surely finite limit $\lim_{T\to\infty} z_T^{(n)}$ implies that there exists a T_n, almost-surely finite and such that $z_t^{(n)} = z_\infty^{(n)}$ for all $t \geq T_n$.

Now combining the two, for every $n \geq N$, we can find a T_n such that $z_t^{(n)} = z_\infty^{(n)}$ for all $t \geq T_n$. Since N is such that for all $n \geq N$, $z_\infty^{(n)} = z_\infty^{(\infty)}$, the lemma is proved. ☐

Lemma 24. *Let T_N be the random variable defined in Lemma 23. For all $t \geq T_N$, we have $x_t = z_\infty^{(\infty)}$.*

Proof. Let $m \geq N$ and $t \geq T_N$ be arbitrary. Observe that $\lim_{T\to\infty} z_T^{(m)} = z_\infty^{(m)} = z_\infty^{(\infty)}$, where the second equality follows from the fact that $m \geq N$. From Lemma 23, there exists an almost-surely finite T_m such that for all $t \geq T_m$, we have $z_t^{(m)} = z_\infty^{(m)} = z_\infty^{(\infty)}$. Let $t' \geq \max(t, T_m)$. Since $t' \geq T_m$, we have $z_{t'}^{(m)} = z_\infty^{(\infty)}$. Basic monotonicity gives us the following two inequalities:

$$z_t^{(m)} \geq z_t^{(n)} = z_\infty^{(\infty)},$$
$$z_t^{(m)} \leq z_{t'}^{(m)} = z_\infty^{(\infty)}.$$

The first inequality follows from monotonicity in space and the second from monotonicity in time. Thus, $z_t^{(m)} = z_\infty^{(\infty)}$. But since $m \geq N$ was arbitrary, it must be the case that $x_t = \lim_{m\to\infty} z_t^{(m)} = z_\infty^{(\infty)}$, where the first equality follows from Proposition 21. Thus we have established that for all $t \geq T_N$, we have $x_t = z_\infty^{(\infty)}$ and, in particular, $x_\infty = \lim_{t\to\infty} x_t = z_\infty^{(\infty)}$ is an almost-surely finite random variable.

☐

Corollary 25. *If $\lambda < \frac{1}{\sum_{i\in\mathbb{Z}^d} a_i}$, then the following interchange of limits holds true:*

$$\lim_{t\to\infty} \lim_{n\to\infty} z_t^{(n)} = \lim_{n\to\infty} \lim_{t\to\infty} z_t^{(n)} = x_\infty = z_\infty^{(\infty)} < \infty \ a.s. \tag{7}$$

Corollary 26. *If $\lambda < \frac{1}{\sum_{j\in\mathbb{Z}^d} a_j}$, then*
$\mathbb{E}[x_\infty] \leq \frac{\lambda a_0}{1-\lambda(\sum_{j\in\mathbb{Z}^d} a_j)} < \infty.$

Proof. From Corollary 25, $x_\infty = \lim_{n\to\infty} z_\infty^{(n)}$. Moreover since $z_\infty^{(n)}$ is non-decreasing in n, it follows from the monotone convergence theorem that $\mathbb{E}[x_\infty] = \lim_{n\to\infty} \mathbb{E}[z_\infty^{(n)}]$. But since $z_\infty^{(n)} \le y_\infty^{(n)}$ and since Lemma 14 gives that $\sup_{n\ge L} \mathbb{E}[y_\infty^{(n)}] = \frac{\lambda a_0}{1-\lambda(\sum_{j\in\mathbb{Z}^d} a_j)}$, we get that $\mathbb{E}[x_\infty] \le \frac{\lambda a_0}{1-\lambda(\sum_{j\in\mathbb{Z}^d} a_j)} < \infty$. □

Now to finish the proof of Theorem 6, we need to conclude about the mean queue length value. To do this, we prove Proposition 9, which will conclude the proof of Theorem 6.

Proof. Proof of Proposition 9
The proof is a repetition of the proof of Lemma 14, except that $\{x_i^{(n)}\}_{i\in B_n(0)}$ is replaced by $\{x_{i;\infty}^{\infty}(0)\}_{i\in\mathbb{Z}^d}$. Proposition 17 holds since $\{x_i(0)\}_{i\in\mathbb{Z}^d}$ is stationary and translation-invariant and hence Mass-Transport holds. □

The proof of Theorem 6 is now complete thanks to the following corollary.

Corollary 27. *If* $\lambda < \frac{1}{\sum_{i\in\mathbb{Z}^d} a_i}$, *then* $\mathbb{E}[x_{0;\infty}(0)] = \frac{\lambda a_0}{1-\lambda\sum_{i\in\mathbb{Z}^d} a_i}$.

7 Uniqueness of Stationary Regime - Proof of Corollary 10

Proof. Let $\boldsymbol{\pi}^{'}$ be an invariant measure on $(\mathbb{Z}^d)^{\mathbb{N}}$, different from the distribution $\boldsymbol{\pi}$ corresponding to $\{x_{i;\infty}(0)\}_{i\in\mathbb{Z}^d}$. We show by elementary coupling and monotonicity arguments that $\boldsymbol{\pi} = \boldsymbol{\pi}^{'}$. Let $T > 0$ be arbitrary. We couple the evolutions of the two systems $\{y_{i;T}(\cdot)\}_{i\in\mathbb{Z}^d}$ and $\{x_{i;T}(\cdot)\}_{i\in\mathbb{Z}^d}$ as follows: Let $\{q_i\}_{i\in\mathbb{Z}^d}$ be distributed according to $\boldsymbol{\pi}^{'}$, independently of everything else. Let $\{y_{i;T}(-T)\}_{i\in\mathbb{Z}^d}$ be such that $y_{i;T}(-T) = q_i$, for all $i \in \mathbb{Z}^d$ and $\{x_{i;T}(-T)\}_{i\in\mathbb{Z}^d}$ be empty, i.e., for all $i \in \mathbb{Z}^d$, we have $x_{i;T}(-T) = 0$. Thus, for all $i \in \mathbb{Z}^d$, $x_{i;T}(-T) \le y_{i;T}(-T)$. Monotonicity in Lemma 1 implies that, for all $i \in \mathbb{Z}^d$, we have $x_{i;T}(0) \le y_{i;T}(0)$. By the definition of invariance, $\{y_{i;T}(0)\}_{i\in\mathbb{Z}^d}$ is distributed as $\boldsymbol{\pi}^{'}$ with $\mathbb{E}[y_{0;T}(0)]$ given in Equation (2). From Proposition 24, we know that as $T \to \infty$, $x_{0;T}(0)$ converges almost-surely to a random variable which has a finite first moment. Thus from the dominated convergence theorem, we have that $\lim_{T\to\infty} \mathbb{E}[x_{0;T}(0)] = \mathbb{E}[x_{0;\infty}(0)]$, which is also the same as given in Equation (2). Thus $\boldsymbol{\pi}^{'}$ coordinate-wise dominates $\boldsymbol{\pi}$. But they have the same first moment. This implies that the two probability measures are the same. □

8 Discussion and Conclusion

We conclude by highlighting the main take-aways from this paper. From a networking perspective, our model opens up new methods to design and compute performance of various physical and MAC layer protocols. The model we propose, accounts for traffic dynamics like queuing based models while also representing the geometry and interference at the physical layer in a simplified way. For example, our framework can potentially allow to quantify the gains in performance that can be obtained at the network level, by implementing interference mitigation techniques at the physical layer.

From a mathematical perspective, we demonstrate an instance of a large interacting queuing network, that does not have any asymptotic independence properties, and yet is tractable. We do this by using Rate Conservation and coupling arguments, which can be of potential interest to study other large interacting queuing networks. Our methods however leave open several intriguing questions - like the existence of other stationary solutions, the correlation decay of queues and

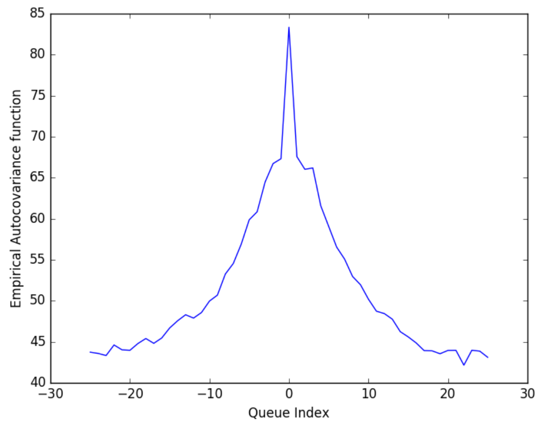

Figure 1: A plot of the empirical covariance function of queue lengths in steady state. We consider $d = 1$ and 51 queue placed on a ring. The arrival $\lambda = 0.1419$ while $\lambda_c = 1/7$ and the interaction function is $a_i = 1$ if $|i| \leq 3$ and 0 otherwise.

convergence to a stationary solution when the initial network is not all empty. The correlation across queues is particularly interesting as shown in Figure 1. In this figure, we are empirically estimating the function $i \to \mathbb{E}[(x_0(t) - \mu)(x_i(t) - \mu)]$, where μ is the mean queue length given in the formula in Theorem 6. However, we cannot simulate an infinite system and hence consider a finite system of 51 queues placed on a ring (i.e., one dimensional torus). We use the interaction function $a_i = 1$ if $|i| \leq 3$ and 0 otherwise. The critical arrival rate is 0.14285 and we used a $\lambda = 0.1419$ to simulate. The mean queue length in this example is 21.18. We estimate the function $\mathbb{E}[(x_0(t) - \mu)(x_i(t) - \mu)]$ by collecting many independent samples approximating the steady state queue lengths $\{x_i^{(25)}\}_{i \in [-25,25]}$. For each collected sample, we evaluate an empirical covariance function by setting the value at $i \in [-25, 25]$ to be $(x_i^{(25)} - \mu)(x_0^{(25)} - \mu)$, where μ is the mean queue length equal to 21.18 in this example. We plot after averaging over many such functions computed on independent queue-length samples. From the plot, the strong positive correlations are very evident, as the function plotted is always large and positive. The figure also supports our intuition that the correlations must decay with distance as expected. Furthermore, it is not hard to prove that this correlation is not an artifact of a finite network and will exist in the limit, unlike in mean-field models. It will be of both mathematical and practical importance to understand how correlations decay with distance. Is it polynomial or exponential ? Is there a phase-transition when the correlation decay switches from one to the other ? Similar to the results on Glauber dynamics of Ising model, does the correlation decay property say something about uniqueness of stationary measures ? These are very interesting questions even from a practical viewpoint, as slow decay of correlations indicate long range spatial clustering, which may be an undesirable operating 'phase' of the network. Exploring this connection with statistical physics models is an exciting line of future work.

Acknowledgements

A.Sankararaman and F.Baccelli were supported by the grant of the Simons Foundations (#197982 to The University of Texas at Austin). S.Foss was partially supported by RSF research grant No. 17-11-01173. S.Foss thanks the hospitality of F.Baccelli for inviting and hosting him at The University of Texas at Austin, when parts of this work were carried out.

References

[1] François Baccelli and Bartlomiej Blaszczyszyn. *Stochastic geometry and wireless networks: Theory*, volume 1. Now Publishers Inc, 2009.

[2] Francois Baccelli and Pierre Brémaud. *Elements of queueing theory: Palm Martingale calculus and stochastic recurrences*, volume 26. Springer Science & Business Media, 2013.

[3] François Baccelli and Serguei Foss. On the saturation rule for the stability of queues. *Journal of Applied Probability*, 32(2):494–507, 1995.

[4] François Baccelli, Fabien Mathieu, and Ilkka Norros. Mutual service processes in euclidean spaces: existence and ergodicity. *Queueing Systems*, 86(1-2):95–140, 2017.

[5] François Baccelli, Fabien Mathieu, Ilkka Norros, and Rémi Varloot. Can p2p networks be super-scalable? In *INFOCOM, 2013 Proceedings IEEE*, pages 1753–1761. IEEE, 2013.

[6] Thomas Bonald, Sem Borst, Nidhi Hegde, and Alexandre Proutiére. Wireless data performance in multi-cell scenarios. *ACM SIGMETRICS Performance Evaluation Review*, 32(1):378–380, 2004.

[7] Thomas Bonald, Sem Borst, Nidhi Hegde, and Alexandre Proutière. Wireless data performance in multi-cell scenarios. *ACM SIGMETRICS Performance Evaluation Review*, 32(1):378–380, 2004.

[8] Charles Bordenave, David McDonald, and Alexandre Proutiere. Asymptotic stability region of slotted aloha. *IEEE Transactions on Information Theory*, 58(9):5841–5855, 2012.

[9] Jacob Cohen and Onno Boxma. *Boundary value problems in queueing system analysis*, volume 79. Elsevier, 2000.

[10] Guy Fayolle and Roudolf Iasnogorodski. Two coupled processors: the reduction to a riemann-hilbert problem. *Probability Theory and Related Fields*, 47(3):325–351, 1979.

[11] Sergei Foss and Denis Denisov. On transience conditions for markov chains. *Siberian Mathematical Journal*, 42(2):364–371, 2001.

[12] Piyush Gupta and Panganmala R Kumar. The capacity of wireless networks. *IEEE Transactions on information theory*, 46(2):388–404, 2000.

[13] Martin Haenggi, Jeffrey G Andrews, François Baccelli, Olivier Dousse, and Massimo Franceschetti. Stochastic geometry and random graphs for the analysis and design of wireless networks. *IEEE Journal on Selected Areas in Communications*, 27(7):1029–1046, 2009.

[14] Frank P Kelly. *Reversibility and stochastic networks.* Cambridge University Press, 2011.

[15] Thomas Liggett. *Interacting particle systems*, volume 276. Springer Science & Business Media, 2012.

[16] Ciamac Moallemi and Devavrat Shah. On the flow-level dynamics of a packet-switched network. In *ACM SIGMETRICS Performance Evaluation Review*, volume 38, pages 83–94. ACM, 2010.

[17] Mariana Olvera-Cravioto and Octavio Ruiz-Lacedelli. Parallel queues with synchronization. *arXiv preprint arXiv:1501.00186*, 2014.

[18] Abishek Sankararaman and François Baccelli. Spatial birth–death wireless networks. *IEEE Transactions on Information Theory*, 63(6):3964–3982, 2017.

[19] Devavrat Shah, NC David, and John N Tsitsiklis. Hardness of low delay network scheduling. *IEEE Transactions on Information Theory*, 57(12):7810–7817, 2011.

[20] Devavrat Shah, Jinwoo Shin, et al. Randomized scheduling algorithm for queueing networks. *The Annals of Applied Probability*, 22(1):128–171, 2012.

[21] Virag Shah and Gustavo de Veciana. Performance evaluation and asymptotics for content delivery networks. In *INFOCOM, 2014 Proceedings IEEE*, pages 2607–2615. IEEE, 2014.

[22] Rayadurgam Srikant and Lei Ying. *Communication networks: an optimization, control, and stochastic networks perspective.* Cambridge University Press, 2013.

[23] Alexander L Stolyar. An infinite server system with customer-to-server packing constraints. In *Communication, Control, and Computing (Allerton), 2012 50th Annual Allerton Conference on*, pages 1713–1720. IEEE, 2012.

[24] Nikita Vvedenskaya, Roland Dobrushin, and Fridrikh Karpelevich. Queueing system with selection of the shortest of two queues: An asymptotic approach. *Problemy Peredachi Informatsii*, 32(1):20–34, 1996.

APPENDIX

A Construction of the Process

To show that the dynamics is well defined, it suffices to establish that the value of the process at some finite time $T < \infty$ can be expressed as a deterministic function of an *arbitrary* initial state $\{x_i(s)\}_{i\in\mathbb{Z}^d}$ for any $T > s > -\infty$ and the driving data $(\mathcal{A}_i, \mathcal{D}_i)_{i\in\mathbb{Z}^d}$. Roughly speaking, the queues evolve by adding a customer to queue i at times $A_q^{(i)}$ and removing a customer from a queue i at times $D_q^{(i)}$ if $U_q^{(i)} \leq \frac{x_i(D_q^{(i)})}{\sum_{j\in\mathbb{Z}^d} a_{j-i} x_j(D_q^{(i)})}$. In other words, we remove a customer from queue i at time $D_q^{(i)}$ with probability $\frac{x_i(D_q^{(i)})}{\sum_{j\in\mathbb{Z}^d} a_{j-i} x_j(D_q^{(i)})}$ independently of everything else. If we had a finite collection of queues, then the above verbose description would be a sufficient description of the dynamics as seen in Algorithm 1 described in the sequel. However, the main effort in this section is to show that the dynamics described above in words can in fact be constructed when there are infinitely many queues. To show this we will need a few definitions.

Definition 28. *For any $X \subset \mathbb{Z}^d$ and any $s \leq t \in \mathbb{R}$, we say that* an arrival occurs in X in the interval $[s,t]$ *if* $\sum_{i \in X} \mathcal{A}_i([s,t]) \geq 1$.

Definition 29. *For any $X \subset \mathbb{Z}^d$ and any $s \leq t \in \mathbb{R}$, we say that* a potential departure occurs in X in the interval $[s,t]$ *if* $\sum_{i \in X} \mathcal{D}_i([s,t]) \geq 1$.

Definition 30. *For any $X \subset \mathbb{Z}^d$ and any $s \leq t \in \mathbb{R}$, we say that* an event occurs in X in the interval $[s,t]$ *if* $\sum_{i \in X} \mathcal{A}_i([s,t]) + \mathcal{D}_i([s,t]) \geq 1$.

A.1 Construction for Finite Sets of Queues

In this section, we first consider the simpler problem of constructing the dynamics if the set of queues were a finite set $X \subset \mathbb{Z}^d$ instead of being the entire grid. In Algorithm 1, we show that given a finite set of queues $X \subset \mathbb{Z}^d$, an initial condition on these queues $\{x_i(s)\}_{i \in X}$, we can determine exactly the state of these queues at some time T in the future based on the driving data $(\mathcal{A}_i, \mathcal{D}_i)_{i \in X}$.

Algorithm 1 Finite Space Finite Time Construction

1: **procedure** F-CONST$(X, s, \{x_i(s)\}_{i \in X}, (\mathcal{A}_i, \mathcal{D}_i)_{i \in X}, T)$
2:     Time-Set $\leftarrow \{s \leq t_1 < t_2 \cdots t_n \leq T\}$.   ▷ The set of times when there is an event in the set X
3:     **for** $1 \leq k \leq n$ **do**
4:         **if** There is an arrival at queue i at time t_k **then**
5:             $x_i(t_k^+) \leftarrow x_i(t_k) + 1$
6:         **else if** There is a departure at queue i at time t_k **then**
7:             $x_i(t_k^+) \leftarrow x_i(t_k) - 1$ with probability $\frac{x_i(t_k)}{\sum_{j \in X} a_{i-j} x_j(t_k)}$ independent of everything else.
8:         **end if**
9:     **end for**
10: **return** $\{x_i(T)\}_{i \in X}$
11: **end procedure**

The algorithm above takes as input a finite set of queues on $\mathbb{Z}^d$, an initial configuration $\{x_i(s)\}_{i \in X}$, and the driving data $(\mathcal{A}_i, \mathcal{D}_i)_{i \in X}$, and it returns the state of all queues at time T. The key reason for the above simple algorithm is line 2, where we could order all the events in set X in increasing order of time. Since X is finite and $T - s$ is finite, we have that n, the total number of events, is finite almost surely. Moreover, from elementary properties of the Poisson process, all the $\{t_i\}_{i=1}^n$ are almost-surely distinct, i.e., no two events occur simultaneously. Thus, we can sequentially order all the potential events and consider them in order to determine exactly what happens at each of the event times given the trajectory up until that time.

A.2 Construction on the Infinite Domain

Theorem 31. *For every λ, every $\{a_i\}_{i \in \mathbb{Z}^d}$ having finite support, every finite $s \leq T$, for all $i \in \mathbb{Z}^d$ with any arbitrary initial conditions $\{x_j(s)\}_{j \in \mathbb{Z}^d}$, we can construct the state of every queue i at time T as a deterministic function of the driving data $(\mathcal{A}_i, \mathcal{D}_i)_{i \in \mathbb{Z}^d}$ and the initial condition $\{x_j(s)\}_{j \in \mathbb{Z}^d}$ by applying Algorithm* 1 *a finite number of times almost-surely.*

This theorem establishes the fact that the dynamics is a well defined function of the driving process. Since, our dynamics is translation invariant, it suffices to show the above theorem for queue 0.

Before we establish the proof of this theorem, we will need several definitions.

Definition 32. *A subset $S \subseteq \mathbb{Z}^d$ is said to be connected if for all $x, y \in S$, there exists $k \geq 1$ and $x_0 := x, x_1, \cdots, x_k := y$ such that $x_i \in S$ for all $i \in [0, k]$, and $||x_i - x_{i-1}||_\infty = 1$ for all $i \in [1, k]$.*

Definition 33. *For each $x \in \mathbb{Z}^d$ and each $L \in \mathbb{N}$, denote by $B_\infty(x, L)$ to be the l_∞ ball of side-length $2\lceil \frac{L}{2} \rceil + 1$ centered around x. Given a set $X \subset \mathbb{Z}^d$, denote by its* L-Thickening *to be the set $\tilde{X}_L := \cup_{z \in X} B_\infty(z, L)$.*

The following is a simple and well-known result in Boolean model percolation where the size of a connected component can be upper bounded by the total progeny of a certain branching process. We provide a short proof here for completeness.

Lemma 34. *For every $d \geq 1$ and every $L \in \mathbb{N}$ that is finite, there exists a $p > 0$, such that if each $z \in \mathbb{Z}^d$ is declared* open *with probability p independent of everything else, we have almost-surely, every connected subset of the random subset $\cup_{z \in \mathbb{Z}^d} \mathbf{1}(z \text{ is open }) B_\infty(z, L)$ to be finite.*

Proof. Denote by the term 'cluster of j' for some $j \in \mathbb{Z}^d$ to be the connected component containing j in the random set $\cup_{z \in \mathbb{Z}^d} \mathbf{1}(z \text{ is open }) B_\infty(z, L)$. To show the lemma, it is sufficient to show that for any arbitrary $j \in \mathbb{Z}^d$, the cluster of j is finite almost-surely. This, along with countable additivity will establish that the cluster of all $j \in \mathbb{Z}^d$ is finite almost-surely, which will prove the lemma.

Let $j \in \mathbb{Z}^d$ be arbitrary. Consider a p such that $p(L+1)^d < 1$. We can upper bound the cardinality of the cluster of j by the total progeny size in a certain sub-critical branching process. We sequentially construct the cluster of j through the following dynamic procedure. This is a classical method to provide an upper bound on the size of the connected component of a vertex in a random graph.

Loosely speaking, we are performing a breadth-first exploration of the cluster of j. It is easy to see that $\mathcal{Y}$ returned by the algorithm is the cluster of j. Moreover, it is easy to see that in each run of the while-loop, the size of $\mathcal{Y}$ increases by at-most $(L+1)^d$ with probability p or remains the same with probability $(1-p)$. Thus, $|\mathcal{Y}|$ is stochastically dominated by the progeny size in a branching process where each point produces either $(L+1)^d$ offsprings with probability p or no progeny with probability $1-p$. Since, we have that $p(L+1)^d < 1$, we know that $|\mathcal{Y}|$ is finite almost-surely and thus, by countable additivity, all clusters of $\mathbb{Z}^d$ are finite almost-surely.

□

We use the last lemma to give a construction of our process. Given λ and $\{a_i\}_{i \in \mathbb{Z}^d}$, choose $L := \sup\{||i||_\infty : i \in \mathbb{Z}_d, a_i > 0\}$. Choose time $\hat{t} > 0$ such that $\exp(-(\lambda+1)\hat{t})) \geq 1 - p$, where p is from the lemma above. Now, we will do our construction in time steps of $\hat{t}$ units.

We show that we can decide on the state of queue 0 at time T in an almost-surely finite number of steps. This will then conclude that we can do so for every queue, since the model is translation invariant. Divide the time interval $[0, T]$ into intervals $[0, \hat{t}], (\hat{t}, 2\hat{t}], \cdots$, i.e., the interval $[0, T]$ is partitioned into finitely many blocks (i.e. $\lceil T/\hat{t} \rceil$) with each block being of at-most $\hat{t}$. Denote by $\hat{\kappa} := T/\hat{t}$ and by $\kappa := \lceil T/\hat{t} \rceil$, the number of time blocks.

Algorithm 2 Identification of the $*$-Connected Component of j in $\cup_{z\in\mathbb{Z}^d}\mathbf{1}(z \text{ is open })B_\infty(z, L)$

procedure CONSTRUCT-CLUSTER(j)
  $\mathcal{Q} \leftarrow \{j\}$
  $\mathcal{U} \leftarrow \mathbb{Z}^d \setminus \{j\}$
  $\mathcal{Y} \leftarrow \emptyset$ ▷ We will given an upper bound to this set's cardinality
  while $\mathcal{Q} \neq \emptyset$ **do**
    $c \leftarrow \text{POP}(\mathcal{Q})$ ▷ Pop an arbitrary element from the set $\mathcal{Q}$
    Set c to be open with probability p
    $\mathcal{Q} \leftarrow \mathcal{Q} \setminus \{c\}$
    if c is declared open **then**
      $\mathcal{Q} \leftarrow \mathcal{Q} \cup (B_\infty(c, L) \cap \mathcal{U})$
      $\mathcal{Y} \leftarrow \mathcal{Y} \cup B_\infty(c, L)$
      $\mathcal{U} \leftarrow \mathcal{U} \setminus B_\infty(c, L)$
    end if
    $\mathcal{Y} \leftarrow \mathcal{Y} \cup \{c\}$
  end whilereturn $\mathcal{Y}$
end procedure

Definition 35. *Given any* $0 \leq s < t$ *and any* $j \in \mathbb{Z}^d$*, we say* j *is* open in the time interval $[s, t]$ *if* $\mathcal{A}_j([s, t]) + \mathcal{D}_j([s, t]) \geq 1$*, i.e. if there is either an arrival or a possible departure from queue* j *in the time interval* $[s, t]$*.*

To proceed with the construction, we set some further notation. For any $r \in [1, \kappa]$, denote by $\mathcal{O}^{(r)}$ the set of sites of $\mathbb{Z}^d$ open in the time interval $[(r-1)\hat{t}, \min(r\hat{t}, T)]$. Let $\tilde{\mathcal{O}}_L^{(r)}$ be its L-Thickening. For any $j \in \mathbb{Z}^d$, denote by $\mathcal{C}_r(j)$ the connected subset of $\tilde{\mathcal{O}}_L$ containing j.

We define $\mathcal{L}_\kappa \subseteq \mathcal{L}_{\kappa-1} \subseteq \cdots \subseteq \mathcal{L}_1 = \mathcal{L}_0 \subset \mathbb{Z}^d$, to be the collection of connected subsets of $\mathbb{Z}_d$ that contain the origin in a recursive fashion as follows:

$$\mathcal{L}_\kappa := \mathcal{C}_\kappa(0)$$
$$\mathcal{L}_{i-1} := \cup_{j\in\mathcal{L}_i}\mathcal{C}_{i-1}(j) \;\forall i \in \{\kappa, \cdots, 2\}$$
$$\mathcal{L}_0 := \mathcal{L}_1$$

It is easy to check that we have $\cup_{i=1}^{\kappa}\mathcal{L}_i$ is finite almost-surely, since, almost-surely, for all $j \in \mathbb{Z}^d$ and all $i \in \{1, \cdots, \kappa\}$, $\mathcal{C}_i(j)$ is finite.

The following fact is now an immediate consequence of the definitions.

Proposition 36. *For all* $i \in \{1, \cdots, \kappa\}$ *and all* $j \in \mathcal{L}_i$ *and* $j^{'} \in \mathcal{L}_i^c$ *such that both* j *and* $j^{'}$ *are open in the time interval* $[(i-1)\hat{t}, i\hat{t}]$*, we have* $d_\infty(j, j^{'}) > L$*.*

Proof. Observe that, for any $r \in \{1, \cdots, \kappa\}$, if we have $\mathcal{C}_r(j) \neq \mathcal{C}_r(j^{'})$ and j and $j^{'}$ are open in the time interval $[(r-1)\hat{t}, \min(r\hat{t}, T)]$, then $d_\infty(j, j^{'}) > L$. This can be seen through contradiction as follows. Assume that j and $j^{'}$ are open in the time interval $[(r-1)\hat{t}, \min(r\hat{t}, T)]$ and $d_\infty(j, j^{'}) \leq L-1$. This implies that there exists a connected path from j to $j^{'}$ in the L-thickening of the set of open sites in the time interval $[(r-1)\hat{t}, \min(r\hat{t}, T)]$. This contradicts the fact that $\mathcal{C}_r(j) \neq \mathcal{C}_r(j^{'})$. Since the set $\mathcal{L}_i$ is the union of a $\mathcal{C}_i(j)$ for some set of j, the result follows. □

The following proposition establishes that we can construct the state of queue 0 at time T.

Proposition 37. *For all $i \in \{0, 1, \cdots, \kappa\}$, given the state of all queues in $\mathcal{L}_i$ at time $i\hat{t}$, the state of each queue in $\mathcal{L}_{i+1}$ at time $(i+1)\hat{t}$ is obtained by running Algorithm 1 on the input data $X = \mathcal{L}_{i+1}$, $s = i\hat{t}$, $T = (i+1)\hat{t}$ and the driving data.*

Proof. For any i, denote by $\tilde{\mathcal{L}_i} \subset \mathcal{L}_i$ the set of queues that are active in the time interval $[(i-1)\hat{t}, i\hat{t}]$. We know from Proposition 36 that any $j \in \mathcal{L}_i^c$ that is active in the time interval $[(i-1)\hat{t}, i\hat{t}]$ is such that $d_\infty(j, \tilde{\mathcal{L}_i}) > L$. In words, the queues outside $\mathcal{L}_i$ do not interact with the active queues in $\mathcal{L}_i$ during the time interval $[(i-1)\hat{t}, i\hat{t}]$. Thus, to know the state of queues in $\mathcal{L}_i$ in the time interval $[(i-1)\hat{t}, i\hat{t}]$, it suffices to look at the evolution of the dynamics inside the set $\mathcal{L}_i$ ignoring the evolutions outside this set. Thus, the statement of the proposition follows. □

As a corollary, for any finite T, and any initial state $\{x_i(0)\}_{i\in\mathbb{Z}^d}$, we can determine $x_0(T)$ by only looking at finitely many events of the driving data $(\mathcal{A}_i, \mathcal{D}_i)_{i\in\mathbb{Z}^d}$. Since the system is translation invariant, we can do this for all $j \in \mathbb{Z}^d$.

As a result of the analysis, we present the following corollary, which will be useful later on.

Corollary 38. *Given any $i \in \mathbb{Z}^d$, any $s \leq T$ and any initial condition $\{x_j(s)\}_{j\in\mathbb{Z}^d}$, there exists a random set $X_{i;s,T} \subset \mathbb{Z}^d$ which is a deterministic function of the driving data $(\mathcal{A}_i, \mathcal{D}_i)_{i\in\mathbb{Z}^d}$, such that the value of $x_i(T)$ obtained by the above construct algorithm is the same as that obtained by running Algorithm 1 on $X_{i;s,T}$.*

Proof. Setting $X_{i;s,T}$ to be equal to the set $\mathcal{L}_1$ concludes the proof. □

A.3 Specialization to One Dimensional Systems

The construction for one dimensions is far simple since for any finite L, all values of $p < 1$ satisfy Lemma 34. Thus given any T, any initial configuration $\{x_i(0)\}_{i\in\mathbb{Z}}$, given any $j \in \mathbb{Z}$ and the driving data $(\mathcal{A}_i, \mathcal{D}_i)_{i\in\mathbb{Z}}$ there almost-surely exists two finite coordinates j_l and j_r such that $j_l \leq j \leq j_r$ such that there is no event in the time interval $[0, T]$ in the set of queues $\{j_l, \cdots j_l - L\}$ and in the set of queues $\{j_r, \cdots, j_r + L\}$. Thus, the state $x_j(T)$ can be ascertained by running Algorithm 1 on the set $X := \{j_l, \cdots, j_r\}$.

B Monotonicity Proofs

B.1 Proof of Lemma 1

Proof. Pick $\hat{t}$ as described in the construction. We will show that $x_0^{'}(\hat{t}) \geq x_0(\hat{t})$. Since, the dynamics is translation invariant, this will then establish that $\{x_i^{'}(\hat{t})\}_{i\in\mathbb{Z}^d}$ coordinate-wise dominates $\{x_i(\hat{t})\}_{i\in\mathbb{Z}^d}$. Since T was finite, we can iterate the above argument in blocks of $\hat{t}$ steps and conclude the proof.

Denote by $\mathcal{O}$ the set of sites of $\mathbb{Z}^d$ open during the time interval $[0, \hat{t}]$ and by $\tilde{\mathcal{O}}_L$ its L-thickening. For any $j \in \mathbb{Z}^d$, denote by $\mathcal{C}(j)$ the connected subset of $\tilde{\mathcal{O}}_L$ containing j. From the definition of $\hat{t}$, we know that, for all $j \in \mathbb{Z}^d$, $\mathcal{C}(j)$ is finite almost-surely. Thus, we can order the events in $\mathcal{C}(0)$ during the time interval as $\mathcal{E}_1, \cdots \mathcal{E}_n$ which occur at times $0 \leq T_1 < T_2 \cdots < T_n \leq \hat{t}$. From elementary properties n is finite and $T_i < T_{i+1}$ almost surely.

Now, we show by induction that after the operations at all times $\{T_i\}_{i=1}^n$, the ordering $x_j^{'}(T_i) \geq x_j(T_i)$ is maintained for all $j \in \mathcal{C}(0)$. We know that at time 0 the inequality is true. Consider the first event. If it is an arrival, then the inequality holds true after the arrival since the arrivals occur in both systems. If the event $\mathcal{E}_1$ is a departure from a queue $j \in \mathcal{C}(0)$, then one of the two cases are possible. Either $x_j^{'}(T_1^-) \geq x_j(T_1^{-1}) + 1$, in which case the ordering $x_j^{'}(T_1) \geq x_j(T_1)$ is trivially true since we have at-most one departure per event. Or, we have equality, i.e. $x_j^{'}(T_1^-) = x_j(T_1^-)$, in which case we have the inequality of death probability $\frac{x_j^{'}(T_1^-)}{\sum_{k \in \mathbb{Z}^d} a_{k-j} x_k^{'}(T_1^-)} \leq \frac{x_j(T_1^-)}{\sum_{k \in \mathbb{Z}^d} a_{k-j} x_k(T_1^-)}$. We have this inequality since at time T_1^-, for all $k \in \mathcal{C}(0)$, we have $x_k^{'}(T_1^-) \geq x_k(T_1^-)$. Since, the death probability is ordered and the two systems are driven by the same data, if $x_j^{'}(T_1) = x_j^{'}(T_1^-) - 1$, then it must be the case that $x_j(T_1) = x_j(T_1) - 1$. Thus, we have that at time T_1, $x_j^{'}(T_1) \geq x_j(T_1)$ for all $j \in \mathcal{C}(0)$. Now, iterating the above arguments over the finitely many events, we see have the inequality $x_j^{'}(\hat{t}) \geq x_j(\hat{t})$ for all $j \in \mathcal{C}(0)$.

□

B.2 Proof of Lemma 2

Proof. We define two systems $\{x_j^{'}(u)\}_{j \in \mathbb{Z}^d}$ and $\{x_j(u)\}_{j \in \mathbb{Z}^d}$ such that at time s, we have for all $j \in \mathbb{Z}^d$, $x_j^{'}(s) = x_j(s)$ which are obtained by running the construction on the driving data $(\mathcal{A}_i, \mathcal{D}_i)_{i \in \mathbb{Z}^d}$ from time 0 until time s, where the initial conditions is arbitrary but the same for both systems. We compute the state of the queues $\{x_j^{'}(u)\}_{j \in \mathbb{Z}^d}$ for $u \geq s$ without the arrivals stopped in set X during the time interval $[s, t]$ and evolve the system $\{x_j(u)\}_{j \in \mathbb{Z}^d}$ with the arrivals stopped, i.e., setting $\mathcal{A}_i([s, t]) = 0$ for all $i \in X$.

Notice that at time s, we have $x_k^{'}(s) \geq x_k(s)$, for all $k \in \mathbb{Z}^d$. In fact, we have equality, but we represent it as an inequality to set up an induction argument. We first show that at time $\hat{t} + s$, we have the inequality $x_k^{'}(\hat{t} + s) \geq x_k(\hat{t} + s)$, for all $k \in \mathbb{Z}^d$. Now since T is finite, we can iterate the above argument in blocks of time steps $\hat{t}$ to conclude the lemma. To prove coordinate-wise domination at time s implies coordinate-wise domination at time $\hat{t} + s$, it suffices to shat that for any $j \in \mathbb{Z}^d$, $x_k^{'}(\hat{t} + s) \geq x_k(\hat{t} + s)$, for all $k \in \mathcal{C}(j)$. Note that in this proof, $\mathcal{C}(j)$ is the connected component containing j of the L-thickening of the set of sites open in the time-interval $[s, \hat{t} + s]$.

As above let $j \in \mathbb{Z}^d$ be arbitrary. Denote by $\mathcal{C}(j)$ the cluster of sites that contain j and are open in the time interval $[s, s + \hat{t}]$. As seen before, this cluster is almost-surely finite. Thus, there is a first event at time $T_1 \geq s$ and a last event at time $T_n \leq \hat{t} + s$ in the set X in the time interval $[s, \hat{t} + s]$. We show that the desired inequality holds through induction on the events, i.e., we show that for all i, $\{x_k^{'}(T_i)\}_{k \in \mathcal{C}(j)} \geq \{x_k(T_i)\}_{k \in \mathcal{C}(j)}$ holds coordinate-wise.

If the event $\mathcal{E}_1$ is an arrival in any queue of $\mathcal{C}(j)$, then the inequality is preserved trivially. If the event $\mathcal{E}_1$ is a departure event from queue $k \in \mathcal{C}(j)$, then there are two cases. Either $x_k^{'}(T_1^-) \geq x_k(T_1^-) + 1$ or $x_k^{'}(T_1^-) = x_k(T_1^-)$. Since, there is at-most one departure per event, the inequality trivially holds if $x_k^{'}(T_1^-) \geq x_k(T_1^-) + 1$. If on the other hand $x_k^{'}(T_1^-) = x_k(T_1^-)$, then the death probabilities are ordered, i.e., we have $\frac{x_k^{'}(T_1^-)}{\sum_{k \in \mathbb{Z}^d} a_{k-j} x_k^{'}(T_1^-)} \leq \frac{x_k(T_1^-)}{\sum_{k \in \mathbb{Z}^d} a_{k-j} x_k(T_1^-)}$. This follows from the fact that time T_1^-, for all $k \in \mathcal{C}(j)$, we have $x_k^{'}(T_1^-) \geq x_k(T_1^-)$. Thus, if there is a death in the system without stopping the arrivals, i.e., if $x_k^{'}(T_1) = x_k^{'}(T_1^-) - 1$, then we will have $x_k(T_1) = x_k(T_1^-) - 1$. Hence, the inequality is preserved after the first event. Thus, iterating over the finitely many events, we have our desired inequality.

□

C Proof of Theorem 11

To prove the theorem, we will define a modified dynamics $\{\tilde{x}_i^{(n)}(t)\}_{t\geq 0, i\in B_n(0)}$ which is coupled with the evolution of $\{x_i^{(n)}(t)\}_{t\geq 0, i\in B_n(0)}$. For notational brevity, in this section, we will drop the superscript n since all systems of interest are on a fixed torus $B_n(0)$. In particular, we write $\{x_i(t)\}_{i\in B_n(0)} := \{x_i^{(n)}(t)\}_{i\in B_n(0)}$ and $\{\tilde{x}_i(t)\}_{i\in B_n(0)} := \{\tilde{x}_i^{(n)}(t)\}_{i\in B_n(0)}$, i.e., we drop the superscript n in the true and modified dynamics.

We construct the modified dynamics such that it satisfies $\tilde{x}_i(t) \geq x_i(t)$ a.s. for all $i \in B_n(0)$ and $t \geq 0$. We will then conclude that this modified dynamics is positive recurrent, which, by standard coupling arguments, will imply that $\{x_i^{(n)}(t)\}_{i\in B_n(0)}$ is positive recurrent. In order to construct the process $\{\tilde{x}_i(t)\}_{i\in B_n(0)}$, we will need some constants that depend on n, which we omit for notational brevity. Given $(\sum_{i\in\mathbb{Z}^d} a_i)^{-1} - \lambda := 2\epsilon > 0$, we choose h small enough so that $\mathbb{P}[J \leq h] - \lambda h \geq \epsilon$, where J is an exponential random variable of mean $\sum_{i\in\mathbb{Z}^d} a_i$. We discretize time into 'slots' of duration h, with the slot boundaries at times $-2h, -h, 0, h, 2h, \cdots$. From the arrival process $(\mathcal{A}_i, \mathcal{D}_i)_{i\in B_n(0)}$, we construct a (deterministically) modified process $(\hat{\mathcal{A}}_i, \hat{\mathcal{D}}_i)_{i\in B_n(0)}$ such that $\mathcal{A}_i \subseteq \hat{\mathcal{A}}_i$ and $\hat{\mathcal{D}}_i \subseteq \mathcal{D}_i$ for all $i \in B_n(0)$. The modification $\{\hat{\mathcal{A}}_i\}_{i\in B_n(0)}$ form the arrival process to $\{\tilde{x}_i(t)\}_{i\in B_n(0)}$ and $\{\hat{\mathcal{D}}_i\}_{i\in B_n(0)}$ form the potential departure process to $\{\tilde{x}_i(t)\}_{i\in B_n(0)}$, with an additional re-normalization step which we will describe in subsection C.2.

C.1 Modification of the Arrival and Departure Events

We implement the modification to the arrival process as follows. In any time slot $[mh, (m+1)h)$, $m \in \mathbb{Z}$, if $\max_{i\in B_n(0)}$
$\mathcal{A}_i([mh, (m+1)h)) < K$, then the process $\{\hat{\mathcal{A}}_i\}_{i\in B_n(0)}$ is identical to $\{\mathcal{A}_i\}_{i\in B_n(0)}$ in the interval $[mh, (m+1)h)$. On the other hand, if $\max_{i\in B_n(0)} \mathcal{A}_i([mh, (m+1)h)) \geq K$, then there must exist a $p \in [mh, (m+1)h)$ and an $l \in B_n(0)$ such that $\mathcal{A}_l([mh, p]) = K$ and $\mathcal{A}_{l'}([mh, p]) < K$ for all $l' \in B_n(0) \setminus \{l\}$ almost surely. In other words, there will be a unique first time when a certain queue will receive more than K customers in that time slot. In this case, the process $\{\hat{\mathcal{A}}_i\}_{i\in B_n(0)}$ is identical to $\{\mathcal{A}_i\}_{i\in B_n(0)}$ in the interval $[mh, p]$. In the interval $(p, (m+1)h)$, the process $\hat{\mathcal{A}}_j, j \in B_n(0)$ is the superposition of all the processes $\{\mathcal{A}_i\}_{i\in B_n(0)}$ in the time interval $(p, (m+1)h)$. In other words, we keep the arrival the same until one of the queues receives at-least K customers in that particular slot, and, subsequently, we replicate any arrival to any queue in that slot to all other queues. This construction ensures that, for any slot $[mh, (m+1)h)$, $\max_{i\in B_n(0)} \hat{\mathcal{A}}_i([mh, (m+1)h)) - \min_{i\in B_n(0)} \hat{\mathcal{A}}_i([mh, (m+1)h)) \leq K$, i.e., the total possible discrepancy in any time slot and any queue is at-most K. We will chose K sufficiently large so that for any queue $i \in B_n(0)$ and any time slot m, the expected number $\mathbb{E}[\hat{\mathcal{A}}_i([mh, (m+1)h))] = h\lambda + \epsilon/10$.

We construct the departure process $\{\hat{\mathcal{D}}_i\}_{i\in B_n(0)}$ as follows, using a parameter δ. We choose δ sufficiently small so that $\mathbb{P}[\hat{J} \leq h] - \lambda h \geq 9\epsilon/10$, where $\hat{J}$ is an exponential random variable with mean $(\sum_{i\in\mathbb{Z}^d} a_i(1+\delta))$. To construct the process $\{\hat{\mathcal{D}}_i\}_{i\in B_n(0)}$, we do a selection by marks of the process $\{\mathcal{D}_i\}_{i\in B_n(0)}$ and retain those atoms whose marks are less than or equal to $(\sum_{i\in\mathbb{Z}^d} a_i(1+\delta))^{-1}$ and delete the other atoms. We then perform a further deletion of points by retaining in each time slot and each queue, at-most one (the first one if there are many) atom whose mark is less than or equal to $(\sum_{i\in\mathbb{Z}^d} a_i(1+\delta))^{-1}$ and delete the rest to obtain the process $\{\hat{\mathcal{D}}_i\}_{i\in B_n(0)}$. Thus the process $\{\hat{\mathcal{D}}_i\}_{i\in B_n(0)}$ is a subset of the process $\{\mathcal{D}_i\}_{i\in B_n(0)}$ and is such that, in any queue and any time-slot, $\hat{\mathcal{D}}_i$ has at-most one atom and the mark of this atom (if any) is less than or equal to $\sum_{i\in\mathbb{Z}^d} a_i(1+\delta))^{-1}$. Since $\{\mathcal{D}_i\}_{i\in B_n(0)}$ is a collection of independent Poisson point processes, the construction ensures that $\{\hat{\mathcal{D}}_i([mh,(m+1)h)\}_{i\in B_n(0),m\in\mathbb{Z}}$ is an i.i.d. collection of $\{0,1\}$ valued random variables. Moreover, it is immediate to verify that the probability $\hat{\mathcal{D}}_0([0,h))=1$ is equal to the probability that an exponential random variable of mean $\sum_{i\in\mathbb{Z}^d} a_i(1+\delta)$ is less than or equal to h.

C.2 Modified Dynamics

Before defining the dynamics of the process $\{\tilde{x}_i(t)\}_{i\in B_n(0)}$, we need to define two large integer constants r_0 and y_0. We first pick $r_1\in\mathbb{N}$ such that, for all $r\geq r_1$, we have

$$\frac{1}{r}\mathbb{E}\max_{j\in B_n(0)}\sum_{i=1}^{r}\hat{\mathcal{A}}_j([(i-1)h,ih))\leq\lambda h+\epsilon/5. \tag{8}$$

Note that such a choice of r_1 is possible since we know that the following limit

$$\begin{aligned}\lim_{r\to\infty}\frac{1}{r}\max_{j\in B_n(0)}\sum_{i=1}^{r}\hat{\mathcal{A}}_j([(i-1)h,ih)) &= \mathbb{E}[\hat{\mathcal{A}}_0([0,h))] \\ &= \lambda h+\epsilon/10,\end{aligned}$$

exists almost surely and in L^1. The almost-sure limit is a consequence of the Strong Law of Large Numbers and the L^1 limit follows since $\left\{\frac{1}{r}\sum_{i=1}^{r}\hat{\mathcal{A}}_0([rh,(r+1)h))\right\}_{j\in B_n(0),r\in\mathbb{N}}$ are uniformly integrable. Uniform integrability follows as the second moment of $\frac{1}{r}\sum_{i=1}^{r}\hat{\mathcal{A}}_0([rh,(r+1)h))$ is uniformly bounded in r and j. Similarly we pick $r_0\geq r_1$ such that, for all $r\geq r_0$, we have

$$\frac{1}{r}\mathbb{E}\min_{j\in B_n(0)}\sum_{i=0}^{r}\hat{\mathcal{D}}_j([(i-1)h,ih))\geq\lambda h+4\epsilon/5. \tag{9}$$

Such a choice for r_0 is possible since the following limit

$$\lim_{r\to\infty}\frac{1}{r}\min_{j\in B_n(0)}\sum_{i=0}^{r}\hat{\mathcal{D}}_j([(i-1)h,ih))=\lambda h+\epsilon/10,$$

exists a.s. Moreover since for all $r\in\mathbb{N}$ and $j\in B_n(0)$, $\frac{1}{r}\sum_{i=0}^{r}\hat{\mathcal{D}}_j([(i-1)h,ih))$ is bounded above and below by 1 and 0 respectively, the limit also exist in L^1 and hence we can choose such a r_0.

Now, we will pick a positive integer y_0 large enough so that

$$1\leq\frac{y_0+(K+1)r_0}{y_0-r_0}\leq(1+\delta). \tag{10}$$

The modified process $\{\tilde{x}_i(t)\}_{i\in B_n(0)}$ evolves with the same update rules as $\{x_i(t)\}_{i\in B_n(0)}$, except that it uses the modified arrival and departure process $(\hat{\mathcal{A}}_i,\hat{\mathcal{D}}_i)_{i\in B_n(0)}$ along with a key *renormalization* step. We will use the constants r_0,y_0 to renormalize so that, for all $m\in\mathbb{N}$, we have $\max_{i\in B_n(0)}$

$\tilde{x}_i(mr_0h) = \min_{i\in B_n(0)} \tilde{x}_i(mr_0h) \geq y_0$. In the rest of this section, we will assume (without loss of generality) that we start the modified system at time 0 with all queues having the same number $y \geq y_0$ of customers, i.e., for all $i \in B_n(0)$, we assume $\tilde{x}_i(0) = y \geq x_i(0)$. The modified system $\{\tilde{x}_i(t)\}_{i\in B_n(0)}$ is driven by the modified arrival-departure process $(\hat{\mathcal{A}}_i, \hat{\mathcal{D}}_i)_{i\in B_n(0)}$, with the same rules of evolution as those of $\{x_i(t)\}_{i\in B_n(0)}$, but with a re-normalization step. More precisely ,at times mr_0h, for $m \in \mathbb{N}$ (i.e., at the end of every r_0 time slots), we add customers and re-normalize as follows:

$$\tilde{x}_i(mr_0h) = \max(y_0, \max_{j\in B_n} \tilde{x}_j(mr_0h^-)), \tag{11}$$

where mr_0h^- is the time instant just before mr_0h. In words, we add more customers so that all queues in the modified dynamics have the same number of customers and this number is at least y_0. Thanks to the monotonicity lemmas 1 and 2, we still have that, for all $t \geq 0$ and all $j \in B_n(0)$, $\tilde{x}_j(t) \geq x_j(t)$ almost surely.

To conclude about the positive recurrence of $\{\tilde{x}_i(t)\}_{i\in B_n(0)}$, we consider the stochastic process $\{Y_m\}_{m\in\mathbb{N}}$ such that $Y_m := \tilde{x}_0(mr_0)$, i.e., the modified dynamics sampled after every r_0h units of time. Note that due to the re-normalization, it does not matter which queue we sample, since all of them will have the same number of customers. It is straightforward to observe that the process $(Y_m)_{m\in\mathbb{N}}$ is a Markov process on the set of integers $\{y_0, \cdots\}$ with respect to the filtration $\{\mathcal{F}_m\}_{m\in\mathbb{N}}$, where $\mathcal{F}_m$ is the sigma-algebra generated by $(\mathcal{A}_i, \mathcal{D}_i)_{i\in B_n(0)}$ up to time mr_0h. This follows from the fact that the process $(\mathcal{A}_i, \mathcal{D}_i)_{i\in B_n(0)}$ is Poisson and has independent increments, and the modification to obtain $(\hat{\mathcal{A}}_i, \hat{\mathcal{D}}_i)_{i\in B_n(0)}$ is *local*, i.e., only depends on the realization of $(\mathcal{A}_i, \mathcal{D}_i)_{i\in B_n(0)}$ in that particular slot. The following simple lemma shows that $(Y_m)_{m\in\mathbb{N}}$ essentially behaves as a $GI/GI/1$ queue.

Lemma 39. *The process $(Y_m)_{m\in\mathbb{N}}$ satisfies the recursion*

$$Y_{m+1} = \max\left(Y_m + \max_{j\in B_n(0)} \chi^{(n)}_{j,m}, y_0\right),$$

where

$$\chi^{(n)}_{j,m} = \hat{\mathcal{A}}_j[mr_0h, (m+1)r_0h) - \hat{\mathcal{D}}_j[mr_0h, (m+1)r_0h).$$

Proof. First notice that the total number of cumulative arrivals in queue $j \in B_n(0)$ from time $[mr_0h, (m+1)r_0h]$ is precisely $\sum_{i=1}^{r_0} \hat{\mathcal{A}}_j([(i-1+mr_0)h, (i+mr_0)h))$. To prove the lemma, we will show that for all queues $j \in B_n(0)$ and all slots $l \in \mathbb{N}$, exactly one departure occurs from queue j in that time slot *if and only if* $\hat{\mathcal{D}}_j([lh, (l+1)h]) > 0$. If we establish this fact, then, at time $(m+1)r_0h^-$, the number of customers in any queue $j \in B_n(0)$ will precisely be equal to $Y_m + \sum_{i=1}^{r_0}(\hat{\mathcal{A}}_j([(i-1+mr_0)h, (i+mr_0)h)) - \hat{\mathcal{D}}_j([(i-1+mr_0)h, (i+mr_0)h))$. The lemma will then follow from the definition of the re-normalization step in Equation 11.

The claim above is established by a sequence of elementary observations. It follows from the construction of $\{\hat{\mathcal{A}}_i\}_{i\in B_n(0)}$ that for all slots $l \in \mathbb{N}$, $\max_{i\in B_n(0)} \hat{\mathcal{A}}_i([lh, (l+1)h]) - \min_{i\in B_n(0)} \hat{\mathcal{A}}_i([lh, (l+1)h]) \leq K$. Since we have at most one departure event in the process $\{\hat{\mathcal{D}}_i\}_{i\in B_n(0)}$ per time slot, per queue, we have $\tilde{x}_i(t) \geq y_0 - r_0$, for all $t \geq 0$ and all $i \in B_n(0)$. There is only a loss of at-most r_0, since we re-normalize to at-least y_0 customers per queue after r_0 slots. These facts yield that, for every instant $t \geq 0$, we have $\max_{i \in B_n(0)} \tilde{x}_i(t) - \min_{i \in B_n(0)} \tilde{x}_i(t) \leq (K+1)r_0$, since we equalize all queues after r_0 slots. Thus the rate of departure in the modified dynamics at any instant $t \geq 0$ and any queue $i \in B_n(0)$ is at least $\frac{\tilde{x}_i(t)}{(\tilde{x}_i(t)+(K+1)r_0)\sum_{i \in \mathbb{Z}^d} a_i}$. Now since $\tilde{x}_i(t) \geq y_0 - r_0$, we have from Equation (10) that the departure rate at any time $t \geq 0$ and any queue $i \in B_n(0)$ is at least $(\sum_{i \in \mathbb{Z}^d} a_i(1+\delta))^{-1}$. By construction of the process $\{\hat{\mathcal{D}}_i\}_{i \in B_n(0)}$, we have at-most one departure event per slot per queue, and its mark is no larger than $(\sum_{i \in \mathbb{Z}^d} a_i(1+\delta))^{-1}$. This concludes the claim that in each slot $m \in \mathbb{N}$ and queue $i \in B_n(0)$, there will be a departure from that queue during the interval $[mh, (m+1)h)$ if and only if $\hat{\mathcal{D}}_i([mh, (m+1)h)) > 0$. □

C.3 Proof of Theorem 11

Proof. Lemma 39 gives that $(Y_m)_{m \in \mathbb{N}}$ is of the form of $Y_{m+1} = \max(Y_m + \xi_m, y_0)$ where $\{\xi_m\}_{m \in \mathbb{N}}$ is an i.i.d. sequence of random variables. Moreover we have

$$\begin{aligned}\mathbb{E}[\xi_0] &= \mathbb{E}\left[\sum_{i=1}^{r}\left(\hat{\mathcal{A}}_j([(i-1)h, ih)) - \hat{\mathcal{D}}_j([(i-1)h, ih)\right)\right] \\ &\leq -r_0\frac{3\epsilon}{5},\end{aligned} \tag{12}$$

where the second equality follows from Equations (8) and (9). Moreover $\{\xi_m\}_{m \in \mathbb{N}}$ has exponential moments, i.e., for all $c > 0$, $\mathbb{E}[e^{c\xi_0}] < \infty$, since Poisson random variables have exponential moments. Thus standard results from $GI/GI/1$ queues immediately imply that the Markov chain $(Y_m)_{m \in \mathbb{N}}$ is ergodic and its stationary distribution ζ on the set of integers $\{y_0, y_0+1, \cdots\}$ is such that there exists $c^{'} > 0$ such that $\mathbb{E}[e^{c^{'}Z}] < \infty$, where $Z \sim \zeta$.

Now assume we start the true system $\{x_i^{(n)}(t)\}_{i \in B_n(0)}$ with some arbitrary initial condition $\mathbf{a} \in \mathbb{N}^{|B_n(0)|}$. Then we start the modified system $\{\tilde{x}_i(t)\}_{i \in B_n(0)}$ with all queues having the same number of customers equal to $\max(y_0, ||\mathbf{a}||_\infty)$. The monotonicity properties from Lemma 1 and 2 imply that, for all $j \in B_n(0)$ and all $t \geq 0$, we have $x_j^{(n)}(t) \leq \tilde{x}_j(t)$. Furthermore, we have $Y_m := \tilde{x}_0(mr_0h)$, $m \in \mathbb{Z}$ as an ergodic Markov chain with stationary distribution ζ on the integers $\{y_0, y_0+1, \cdots\}$. From standard results on the $GI/GI/1$ queue, the sequence of random variables Y_m converges in total variation to the distribution ζ as m goes to infinity. Now since, for all $t \geq mr_0h$ and all $j \in B_n(0)$, we have the trivial inequality

$$\tilde{x}_j(t) \leq Y_m + \sum_{i \in B_n(0)} (\hat{\mathcal{A}}_i([mr_0h, t]) + \mathcal{D}_i([mr_0h, t])),$$

where Y_m is independent of
$\sum_{i \in \mathbb{Z}^d}(\hat{\mathcal{A}}_i([mr_0h, t]) + \mathcal{D}_i([mr_0h, t]))$. Hence as $t \to \infty$, for all $j \in B_n(0)$, $x_j(t)$ is stochastically dominated by a random variable that converges in total variation to a non-degenerate distribution on $\{y_0, y_0+1, \cdots\}$, which is the sum of independent random variables distributed as ζ and $\sum_{i \in B_n(0)}(\hat{\mathcal{A}}_i([0, r_0h]) + \mathcal{D}_i([0, r_0h]))$. Since $\{x_i(t)\}_{i \in B_n(0)}$ is a Markov process on a finite-dimensional space, the stochastic domination by a stationary process implies that $\{x_i(t)\}_{i \in B_n(0)}$ is ergodic. Moreover since Y_m has exponential moments, it follows that the stationary distribution of $\{x_i(t)\}_{i \in B_n(0)}$ will also has exponential moments.

□

D Instability - Proof of Theorem 8

Proof. From monotonicity, it suffices to establish that the restriction of the queuing dynamics to the set $[-n, n]$ is transient for some sufficiently large n. We will consider a slightly modified process $\{\tilde{x}_i(t)\}_{i\in[-n,n]}$, such that $\tilde{x}_i(t) \leq x_i(t)$ for all $i \in [-n, n]$ and all $t \geq 0$. This modified dynamics will evolve according to the same rules as our original dynamics, except that the arrival process is different along with a key-renormalization step. We will argue that, a.s., for all $i \in [-n, n]$, $\tilde{x}_i(t)$ converges to infinity, thereby implying transience for our original infinite system. To describe the modified dynamics, we will need several constants which we now describe.

Given $\lambda > \frac{1}{3}$, we choose a $\delta > 0$ such that $\lambda > \frac{1}{3} + \delta$. We pick $h > 0$ and $\epsilon > 0$ small enough for having $\mathbb{P}[\text{Exp}(\lambda) \leq h] - (1/3 + \delta)h \geq \epsilon > 0$. Here $\text{Exp}(\lambda)$ refers to an exponential random variable with mean λ^{-1}. Such a h and ϵ exist since $\lambda > 1/3 + \delta$. We pick $n \in \mathbb{N}$ sufficiently large so that $\frac{1}{3-2.01/n} < \frac{1}{3} + \delta$. From the arrival process $\mathcal{A}_i$, we construct a sub-process $\hat{\mathcal{A}}_i$ such that at-most one arrival (the first if there is more than one) occurs in each queue and each time-slot. We choose $r \in \mathbb{N}$ with $r \geq 4n/\epsilon$ sufficiently large so that

$$\frac{1}{r}\mathbb{E}[\min_{i\in[-n,n]} \hat{\mathcal{A}}_i[0, rh]] \geq (\frac{1}{3} + \delta)h + \frac{2}{3}\epsilon \tag{13}$$

$$\frac{1}{r}\mathbb{E}[\max_{i\in[-n,n]} \mathcal{D}_i[[0, rh] \times [0, 1/3 + \delta]]] \leq (\frac{1}{3} + \delta)h + \frac{1}{3}\epsilon \tag{14}$$

Such a choice for r exists since $\frac{1}{r}\min_{i\in[-n,n]} \hat{\mathcal{A}}_i[0, rh]$ converges a.s. as r goes to infinity to a value greater than or equal to $(1/3+\delta)h+\epsilon$. Moreover, the convergence is also in L^1 since the family $\{\frac{1}{r}\min_{i\in[-n,n]} \hat{\mathcal{A}}_i[0, rh]\}_{r\geq 1}$ is bounded from above and below by 1 and 0, respectively. Similarly, the strong law of large numbers entails that $\frac{1}{r}\max_{i\in[-n,n]} \mathcal{D}_i[0, rh]$ converges a.s. to $(1/3 + \delta)h$. This convergence also occurs in L^1 since the family $\{\frac{1}{r}\mathcal{D}_i[0, rh]\}_{i\in[-n,n],r\geq 1}$ is uniformly integrable (UI). Since n is fixed, this immediately implies that $\{\frac{1}{r}\min_{i\in[-n,n]} \mathcal{D}_i[0, rh]\}_{i\in[-n,n],r\geq 1}$ is UI. The UI of $\{\frac{1}{r}\mathcal{D}_i[0, rh]\}_{i\in[-n,n],r\geq 1}$ follows since the second moment $\sup_{r\geq 1,i\in[-n,n]} \mathbb{E}\left[\left(\frac{1}{r}\mathcal{D}_i[0, rh]\right)^2\right] < \infty$ is uniformly bounded in r and i.

The dynamics of our lower-bound is identical to that of the original system, except that it is driven by the modified arrival process $\hat{\mathcal{A}}$ along with a modification at the end of every r time slots. After every r time slots, we reduce queue lengths to the largest possible *envelope function*, which serves as a Lyapunov function. Denote by $L_n(\cdot) : \mathbb{N} \to \mathbb{N}^{(2n+1)} := [x, 2x, \cdots, (n-1)x, nx, (n-1)x, \cdots, x]$, the linear triangle function. At the end of every r slots, we will reduce the queue lengths to $L_n(x)$ for the largest possible x. In other words, if the queue length at the end of r slots is given by the vector $Q := [q_{-n}, \cdots, q_0, \cdots, q_n]$, then we reduce the queue lengths to $L_n(x)$ for the largest possible $x \in \mathbb{N}$ such that no coordinate of $Q - L_n(x)$ is strictly negative. From the monotonicity of the dynamics, for all $i \in [-n, n]$ and $t \geq 0$, we have $\tilde{x}_i(t) \leq x_i(t)$.

For $m \in \mathbb{Z}$, let $Y_m := \tilde{x}_n(mrh)$, i.e., Y_m is the length of queue n *just after* the reduction of customers to the level function $L_n(\cdot)$. Since the dynamics is driven by Poisson point processes, it is easy to verify that Y_m is a Markov process with respect to the filtration $\mathcal{F}^Y_m := \sigma(\cdots, Y_0, \cdots Y_m)$.

The following is a key structural lemma which will enable us to conclude about the proof of Theorem 8.

Lemma 40. *There exists a $x_0 \in \mathbb{N}$ and $\gamma > 0$ such that for all $x \geq x_0$, we have $\mathbb{E}[Y_1 - Y_0 | Y_0 = x] \geq \gamma$.*

Proof. To prove this lemma, we need to define a constant C which satisfies $r(2n+1)\mathbb{P}[\text{Poi}((\frac{1}{3} + \delta)rh) \geq Crh\log(x)] \leq x^{-2}$ for all sufficiently large x. We call a collection of r time-slots 'x-good' if in any block of r slots $[prh, (p+1)rh]$, for any $p \in \mathbb{Z}$, we have $\max_{i \in [-n,n]} \mathcal{D}_i[[prh, (p+1)rh] \times [0, 1/3+\delta]] \leq Crh\log(x)$. In other words a collection of r time-slots is 'x-good' if the total number of potential departures in every queue that has its mark less than or equal to $1/3+\delta$ does not exceed $Crh\log(x)$. We can then use this definition to break up the drift into two terms as follows -

$$\mathbb{E}[Y_1 - Y_0 | Y_0 = x] = \mathbb{E}[(Y_1 - Y_0)\mathbf{1}_{[0,rh] \text{ is } x\text{-Good}} | Y_0 = x] + \mathbb{E}[(Y_1 - Y_0)(1 - \mathbf{1}_{[0,rh] \text{ is } x\text{-Good}}) | Y_0 = x]$$

We can bound the drift trivially as follows -

$$\mathbb{E}[Y_1 - Y_0 | Y_0 = x] \geq \mathbb{E}[(Y_1 - Y_0)\mathbf{1}_{[0,rh] \text{ is } x\text{-Good}} | Y_0 = x] - x\mathbb{P}[[0, rh] \text{ is not } x\text{-Good}]$$

Now using the definition of x-Good, we get

$$\mathbb{E}[Y_1 - Y_0 | Y_0 = x] \geq \mathbb{E}[(Y_1 - Y_0)\mathbf{1}_{[0,rh] \text{ is } x\text{-Good}} | Y_0 = x] - \frac{1}{x}. \tag{15}$$

Now let x be sufficiently large so that

$$\frac{1}{3 - (\frac{2}{n} + \frac{3Crh\log(x)}{x} + \frac{3r}{x})} < \frac{1}{3} + \delta \tag{16}$$

Such a choice of x is possible since n is so large that $\frac{1}{3-2.01/n} < \frac{1}{3} + \delta$. The crucial fact is that if x satisfies Equation (16), then, on the event that a slot is x-Good, the number of departures from any queue in the collection of r-slots is at-most the number of potential departures that have a mark less than or equal to $1/3+\delta$. Conditioned on the collection of r time-slots being x-Good, the maximum departure probability in any queue in that time period is

$$\frac{nx + r}{2(n-1)x - 2Crh\log(x) + nx - Crh\log(x)}, \tag{17}$$

which is less than or equal to $1/3+\delta$, since x satisfies Equation (16).

For notational convenience, for any $p \in \mathbb{Z}$ and $x \in \mathbb{N}$, denote by E_x^p the event that the time interval $[prh, (p+1)rh]$ is x-Good.

If the time interval $[0, rh]$ is x-Good, then the drift $Y_1 - Y_0$ is at-least the minimum difference in the total number of arrivals and departures with marks at-most $1/3+\delta$, with a factor of $1/n$ to account for the re-normalization. In other words, we have

$$\mathbb{E}[(Y_1 - Y_0)E_x^0|Y_0 = x] \geq \mathbb{E}\left[\left\lfloor \frac{1}{n} \min_{i\in[-n,n]} \left(\hat{\mathcal{A}}_i[0,rh] - (\mathcal{D}_i[[0,rh]\times[0,1/3+\delta]])E_x^0\right)\right\rfloor \middle| Y_0 = x\right],$$

$$\stackrel{(a)}{\geq} \frac{1}{n}\left(r\left((\frac{1}{3}+\delta)h + \frac{2\epsilon}{3}\right) - \mathbb{E}[\max_{i\in[-n,n]} \mathcal{D}_i[[0,rh]\times[0,1/3+\delta]]E_x^0|Y_0 = x]\right) - 1, \tag{18}$$

where the inequality (a) follows from Equation (13) and the fact that for all $x \in \mathbb{R}$, $\lfloor x \rfloor \geq x-1$. Now since $\lim_{x\to\infty} \mathbf{1}_{[0,rh] \text{ is } x\text{-Good}} = 1$ almost-surely, we have from dominated convergence and Equation (14) that

$$\mathbb{E}[\max_{i\in[-n,n]} \mathcal{D}_i[[0,rh]\times[0,1/3+\delta]]E_x^0|Y_0 = x] \leq r\left((\frac{1}{3}+\delta)h + \frac{\epsilon}{3} + o(x)\right). \tag{19}$$

Combining Equations (15,18) and (19), we get that for all x that satisfy Equation (16), we have

$$\mathbb{E}[Y_1 - Y_0|Y_0 = x] \geq \frac{r}{n}\left(\frac{\epsilon}{3} - o(x)\right) - 1.$$

Since $r \geq 4n/\epsilon$, by choosing x sufficiently large such that the $o(x)$ term is less than $\epsilon/12$, the drift will be strictly positive.

□

Proof. (Proof of Theorem 8)

It suffices to establish that the Markov chain $(Y_m)_{m\in\mathbb{N}}$ is transient. This is apparent since Y_m satisfies all the conditions of the main result in [11], which we verify here for completeness. Denote by $\Delta_m := Y_{m+1} - Y_m$. If we establish the following three conditions on the one-step drift Δ_m, then transience is concluded in view of Lemma 40 and the main result in [11].

1. For each $m \in \mathbb{N}$, $\Delta_m \leq r$ a.s.

2. For all $x, t \in \mathbb{N}$, $\mathbb{P}[|\Delta_1| \geq t|Y_0 = x] \leq \mathbb{P}[\text{Poi}((2n+1)r(\lambda+1)rh) \geq t]$. In particular, for all $s \in \mathbb{R}$ and all $x \in \mathbb{N}$, $\sup_{x\in\mathbb{N}} \mathbb{E}[e^{s|\Delta_1|}|Y_0 = x] < \infty$.

3. For all $y \in \{0, 1, \cdots, x_0 - 1\}$ where x_0 is defined in Lemma 40, $\mathbb{P}[\exists m \geq 1 : Y_m \geq x_0|Y_0 = y] = 1$.

To finish the proof, we resort to the technique of [11] which we reproduce for our scenario for completeness.

Lemma 41. *There exists a $s > 0$ such that $\tilde{Y}_m = e^{-s(Y_m - x_0)^+}$ is a positive super-martingale.*

Proof. It suffices to establish that $\mathbb{E}[Y_1 - Y_0|Y_0 = y] \leq 0$, for all $y \in \mathbb{N}$. If $y \leq x_0$, the inequality is immediate. Let $y > x_0$ and denote by $a := y - x_0 > 0$. In this case one can decompose the martingale difference as

$$\mathbb{E}[Y_1 - Y_0|Y_0 = y] = \mathbb{E}[e^{-s(a+\Delta_y)^+} - e^{-sa}]. \tag{20}$$

But since for all $s>0$ and all y, $\mathbb{E}[e^{s|\Delta_y|}]<\infty$, the function $s\to\mathbb{E}[e^{-s(a+\Delta_y)^+}-e^{-sa}]$ is bounded and continuous for each fixed a. Moreover the derivate at $s=0$ satisfies

$$\begin{aligned}\frac{d}{ds}\mathbb{E}[e^{-s(a+\Delta_y)^+}-e^{-sa}]|_{s=0} &= -\mathbb{E}[\Delta_y \mathbf{1}_{\Delta_y\geq -a}]-a\mathbb{P}[\Delta_y\leq -a],\\ &\leq -\mathbb{E}[\Delta_y],\\ &\leq -\gamma.\end{aligned}$$

Thus there exists a $s>0$, such that for all $y\geq x_0$, we have $\mathbb{E}[e^{-s(a+\Delta_y)^+}-e^{-sa}]\leq 0$, which concludes the lemma. □

Thus, by familiar martingale convergence theorems, $\tilde{Y}_m$ and hence Y_m has an almost-sure limit as m goes to infinity. To conclude that $Y_m\to\infty$ almost-surely, it suffices to establish that $\mathbb{P}[\limsup_{m\to\infty}Y_m=\infty|Y_0=0]=1$. But this is in a sense immediate from the description of the dynamics. We know that for all $y\in\mathbb{N}$, $\mathbb{P}[Y_1-Y_0=r|Y_0=y]\geq\left(\left(\frac{1}{3}+\delta\right)h+\epsilon\right)^r e^{-rh}:=\eta>0$. Now let $b\in\mathbb{N}$ be arbitrary and let $\mathcal{E}_k$ be the event that $Y_{k\lceil\frac{b}{r}\rceil}\geq b$. From the monotonicity of the dynamics, for all $y\in\mathbb{N}$, we have that $\mathbb{P}[\mathcal{E}_k|Y_{(k-1)\lceil\frac{b}{r}\rceil}=y]\geq\eta>0$. Thus $\mathbb{P}[\mathcal{E}_k^{\mathsf{c}}|\mathcal{E}_1^{\mathsf{c}},\cdots,\mathcal{E}_{k-1}^{\mathsf{c}}]\leq 1-\eta^{'}<1$. It follows that $\mathbb{P}[\cap_{k\geq 1}\mathcal{E}_k^{\mathsf{c}}]=0$. This implies that the random variable $\tau_1:=\min\{m\geq 0: Y_m\geq b|Y_0=0\}$ is a.s. finite. Now, from the strong Markov property, the sequence of random times $\tau_{k+1}:=\min\{m\geq\tau_k+1: Y_m\geq b\}$ are almost-surely finite. This implies that $\mathbb{P}[Y_m\geq b$ infinitely often $|Y_0=y]=1$, for all $y\in\mathbb{N}$. Since b was arbitrary, this implies that $\mathbb{P}[\limsup_{m\to\infty}Y_m=\infty|Y_0=0]=1$. □

□