HAL
open science

Metastability of Queuing Networks with Mobile Servers

F Baccelli, A Rybko, Senya Shlosman, A Vladimirov

To cite this version:

F Baccelli, A Rybko, Senya Shlosman, A Vladimirov. Metastability of Queuing Networks with Mobile Servers. Journal of Statistical Physics, 2018, 173 (3-4), pp.1227-1251. 10.1007/s10955-018-2023-z . hal-01648391

HAL Id: hal-01648391

https://hal.science/hal-01648391v1

Submitted on 25 Nov 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

arXiv:1704.02521v1 [math.PR] 8 Apr 2017

Metastability of Queuing Networks with Mobile Servers

F. Baccelli[2,5], A. Rybko[3], S. Shlosman[4,1,3] and A. Vladimirov[3]

[1]Aix Marseille Université, Université de Toulon, CNRS, CPT, UMR 7332, France

[2]UT Austin, Department of Mathematics, USA

[3]Institute of the Information Transmission Problems, RAS, Moscow, Russia

[4]Skolkovo Institute of Science and Technology, Moscow, Russia

[5]INRIA Paris, France

April 11, 2017

Abstract

We study symmetric queuing networks with moving servers and FIFO service discipline. The mean-field limit dynamics demonstrates unexpected behavior which we attribute to the metastability phenomenon. Large enough finite symmetric networks on regular graphs are proved to be transient for arbitrarily small inflow rates. However, the limiting non-linear Markov process possesses at least two stationary solutions. The proof of transience is based on martingale techniques.

This work was initiated in 2010 at the Newton Institute in Cambridge, UK, as part of the semester on Stochastic Networks. It then developed thanks to a CNRS France-Russia international grant (Mobilité, Dynamique et Théorie des Files d'Attente to 1,3 and 5). The core of the research was then carried out in France, Russia and the USA: in France through grants to 1, in the framework of the Labex Archimede (ANR-11-LABX-0033) and of the A*MIDEX project (ANR-11-IDEX-0001-02), funded by the "Investissements d'Avenir" French Government programme managed by the French National Research Agency (ANR); in Russia through a grant of the Russian Foundation for Sciences (project No. 14-50-00150) to 3; in the USA through a grant of the Simons Foundation (Grant No. 197982) to 2. All these supports are gratefully acknowledged.

MCS: 60G10, 60G44, 60K25, 60K35, 60J25, 37A60.

1 Introduction

In this paper we consider networks with moving servers. The setting is the following: the network is living on a finite or countable graph $G = (V, E)$, at every node $v \in V$ of which one server s is located at any time. For every server, there are two incoming flows of customers: the exogenous customers, who come from the outside, and the transit customers, who come from some other servers. Every customer c coming into the network (through some initial server $s(c)$) is assigned a destination $D(c) \in V$ according to some randomized rule. If a customer c is served by a server located at $v \in V$, then it jumps to a server at the node $v' \in V$, such that $\mathrm{dist}(v', D(c)) = \mathrm{dist}(v, D(c)) - 1$, thereby coming closer to its destination. If there are several such v', one is chosen uniformly. There the customer c waits in the FIFO queue until his service starts. If a customer c completes his service by the server located at v, and it so happens that $\mathrm{dist}(v, D(c))$ is 1 or 0, the customer is declared to have reached its destination and leaves the network.

The important feature of our model is that the servers of our network are themselves moving over the graph G. Namely, we suppose that any two servers s, s' located at adjacent nodes of G exchange their positions as the clock associated to the edge rings. The time intervals between the rings of each alarm clock are i.i.d. exponential with rate β. When this happens, each of the two servers brings all the customer waiting in its buffer or being served, to the new location. In particular, it can happen that after such a swap, the distance between the location of the customer c and its destination $D(c)$ increases (at most by one). We assume that the service times of all customers at all servers are i.i.d. exponential with rate 1.

The motivation for this model comes from opportunistic multihop routing in mobile wireless networks. Within this context, the servers represent mobile wireless devices. Each device moves randomly on the graph G which represents the phase space of device locations. The random swaps represent the random mobile motions on this phase space. Each node $v \in G$ of the phase space generates an exogenous traffic (packetized information) with rate λ_v corresponding to the exogenous customers alluded to above. Each such packet has some destination, which is some node of G. In opportunistic routing (see Volume 2 of [BB]), each wireless device adopts the following simple

routing policy: any given packet scheduled for wireless transmission is sent to the neighboring node which is the closest to the packet destination. The neighbor condition represents in a simplified way the wireless constraints. It implies a multihop route in general. This routing policy is the most natural one to use in view of the lack of knowledge of future random swaps. For details on this motivation, the reader may refer to the literature on mobile ad hoc networks and that on delay-tolerant networks. To the best of our knowledge, the present paper is the first mathematical attempt to analyze queuing within this mobile wireless framework.

In this paper we study the following types of graphs:

1. Finite graphs G: we consider the cyclic graph $C_K = \mathbb{Z}^1/K\mathbb{Z}^1$ and the toric graph $\mathcal{T}_{KL} = \mathbb{Z}^2/K\mathbb{Z}^1 \oplus L\mathbb{Z}^1$, as well as general connected g-regular graphs (i.e. graphs where every vertex has g adjacent edges).

2. Mean-field graphs G^N: for a graph $G = (V, E)$ we denote by G^N a graph with the node set $V \times \{1, 2, ..., N\}$; two nodes (v', n') and (v'', n'') are connected by an edge iff $(v', v'') \in E$.

3. Infinite graphs (like $\mathbb{Z}^1$ and $\mathbb{Z}^2$).

4. Limit graphs $G^\infty = \lim_{N\to\infty} G^N$. The resulting limiting networks can be analyzed using the theory of Non-Linear Markov Processes (NLMP).

The first two classes of graphs represent the type of wireless networks alluded to above. The interest in mean-field versions of the last two types is both of mathematical and practical nature. The mathematical interest of the mean-field version of a network is well documented. There are also practical motivations for analyzing such networks: their properties are crucial for understanding the long-time behavior of finite size networks.

The results we obtain depend on the graph and look somewhat surprising. First of all, we find that for finite regular graphs, the network is transient once the diameter of the graph is large enough. For example, consider the network on the graph C_K with Poisson inflows with rate $\lambda > 0$ at all nodes, exponential service times with rate 1, FIFO discipline and node swap rate $\beta > 0$. Then for all $K \geq K(\lambda, \beta)$, the queues at all servers tend to infinity as time grows. In words this means that the network is unstable for any λ, however small it is – once the network is large enough. The same holds, probably, for the network on $\mathbb{Z}^1$, but we do not prove it here.

The same picture takes place for graphs G^N with N finite. However, the limiting picture, for $N = \infty$, is different: the corresponding NLM processes on C_K and $\mathbb{Z}^1$ have stationary distributions, provided $0 < \lambda \leq \lambda_{cr}(K, \beta)$, with $0 < \lambda_{cr}(K, \beta) < \infty$ for all $K \leq \infty$. Moreover, for all $\lambda < \lambda_{cr}$ there are at least two different stationary distributions, see Sect. 3 for more details.

On the other hand, the general convergence result of [BRS] claims the convergence of the networks on G^N to the one on G^∞ as $N \to \infty$, which seem to contradict to the statements above. The explanation of this 'contradiction' is that the convergence in [BRS] holds only on finite time intervals $[0, T]$. That is, for any T there exists a value $N = N(T)$, such that the network on G^N is close to the limiting network on G^∞ for all $t \in [0, T]$, provided $N \geq N(T)$. Putting it differently, the G^N network behaves like the limiting G^∞ network – and might even look as a stationary process – for quite a long time, depending on N, but eventually it departs from such a regime and gets into the divergent one. Clearly, the picture we have is an instance of metastable behavior. We believe that more can be said about the metastable phase of our networks, including the formation of critical regions of servers with oversized queues, in the spirit of statistical mechanics, see e.g. [SS], but we will not elaborate here on that topic.

2 Finite networks

This section starts with a detailed description of the methodology for proving the instability of finite networks. This is done on the special case of cyclic networks. We then discuss the extension to the mean-field versions of cyclic networks and to toric networks. The general results, which bear on connected d-regular graphs, are considered last.

2.1 The cyclic network

We start with the cyclic graph $C_K = \mathbb{Z}^1/K\mathbb{Z}^1$. We use the notation $C_K = (V_K, E)$, where $V_K = \{1, \ldots, K\}$ and $E = \{(1,2), \ldots, (K-1, K), (K, 1)\}$. For future simplicity we take K to be odd.

We study a continuous-time Markov process on a countable state Q, related to the graph C_K. Namely,

$$Q = \{q^v : v \in V_K\} = (V_K^*)^{V_K},$$

where V_K^* is the set of all finite words in the alphabet V_K, including the empty word $\varnothing$.

The queue $q^v \in V_K^*$ at a server located at $v \in V_K$ consists of a finite (≥ 0) number of customers which are ordered by their arrival times (FIFO service discipline) and are marked by their destinations which are vertices of the graph C_K. Since the destination of the customer is its only relevant feature, in our notations we sometime will identify the customers with their destinations.

2.1.1 Dynamics

Let us introduce the continuous-time Markov process $\mathcal{M} = \mathcal{M}(t)$ with the state space Q. Let h_v be the length of the queue q^v at node v. We have $q^v = \left\{q_1^v, \ldots, q_{h_v}^v\right\}$ if $h_v > 0$ and $q^v = \varnothing$ if $h_v = 0$.

The following events may happen in the process $\mathcal{M}$.

An arrival event at node v changes the queue at this node. If the newly arrived customer has node w for its destination, then the queue changes from q^v to $q^v \oplus w$, that is, to $\left\{q_1^v, \ldots, q_{h_v}^v, w\right\}$ if $h_v > 0$ or from $\varnothing$ to $\{w\}$ if $h_v = 0$.

In this paper we consider the situation where each exterior customer acquires its destination at the moment of first arrival to the system, in a translation-invariant manner: the probability to get destination w while arriving to the network at node v depends only on $w - v \bmod K$. The case $w = v$ is not excluded. We thus have the rates $\lambda_{v,w}$, $v, w \in C_K$, where $\lambda_{v,w}$, depends only on $w - v \bmod K$, and the jump from q^v to $q^v \oplus w$, corresponding to the arrival to v of the exterior customer with final destination w happens with the rate $\lambda_{v,w}$. We introduce the rate λ_v of exterior customers to queue v as

$$\lambda_v = \sum_w \lambda_{v,w}. \tag{1}$$

According to our definitions, $\lambda_v = \lambda$ does not depend on v.

Each node is equipped with an independent Poisson clock with parameter 1 (the service rate). As it rings, the service of customer q_1^v is completed, provided $h_v > 0$; nothing happens if $h_v = 0$. In the former case the queue at node v changes from q^v to

$$q_-^v = \left\{q_2^v, \ldots, q_{h_v}^v\right\}$$

(we also define $\varnothing_- = \varnothing$) and immediately one of the two things happen: either the customer q_1^v leaves the network, or it jumps to one of the two

neighboring queues, $q^{v\pm 1}$. The customer q_1^v leaves the network only if its current position, v, is at distance ≤ 1 from its destination, i.e. iff $q_1^v = v-1$, v, or $v+1$. This is just one of many possible choices we make for simplicity. Otherwise it jumps to the neighboring vertices $w = v \pm 1$, which is the closest to its destination, i.e. to the one which satisfy: $\mathrm{dist}(w, q_1^v) = \mathrm{dist}(v, q_1^v) - 1$ (there is a unique such $w \in V_K$ since we assume K to be odd. The case of even K requires small changes).

The last type of events is the swap of two neighboring servers. Namely, there is an independent Poisson clock at each edge $uv \in E$ of C_K, with rate $\beta > 0$. As it rings, the queues at the vertices u and v swap their positions, that is,

$$q^v(t_+) = q^u(t), \qquad q^u(t_+) = q^v(t).$$

2.1.2 Submartingales

Here we introduce some martingale technique that will be used for the proof of transience of $\mathcal{M}$ for K large enough. To begin with, we label the K servers by the index $k = 1, ..., K$; this labelling will not change during the evolution. Together with the original continuous-time Markov process $\mathcal{M}(t)$ we will consider the embedded discrete time process $M(n)$, which is the value of $\mathcal{M}(t)$ immediately after the n-th event. The state of the process M consists of the states of all K servers and all their locations.

The general theorem below will be applied to the quantities X_n^k, which are, roughly speaking, the lengths of the queues at the servers k, $k = 1, ..., K$, of the process $M(\Lambda n)$. The integer parameter $\Lambda = \Lambda(K, \lambda, \beta)$ will be chosen large enough, so that, in particular, after time Λ, the locations of the servers are well mixed on the graph C_K, and the joint distribution of their location on V_K is close to uniform on the set of permutations on V_K. Below, we aim to prove that the conditional expectations of all the differences $X_{n+1}^k - X_n^k$ are positive. We start with the following theorem.

Theorem 1 *Let $\mathcal{F} = \mathcal{F}_n$, $n = 0, 1, \ldots,$ be a filtration and let X_n^k, $k = 1, \ldots, K$, be a finite family of non-negative integer-valued submartingales adapted to $\mathcal{F}$, such that for all $k = 1, \ldots, K$, and all $n = 0, 1, \ldots,$ the following assumptions hold:*

(1) For some $\rho > 0$ the inequality

$$\mathbb{E}_{\mathcal{F}_n}(X_{n+1}^k - X_n^k) \geq \rho \tag{2}$$

holds whenever $X_n^k > 0$.

(2) The increments are bounded by a constant R*:*

$$|X_{n+1}^k - X_n^k| \leq R \quad a.s. \tag{3}$$

Then there exists an initial state $(X_0^1, \ldots, X_0^K)$ *such that, with positive probability,* $X_n^k \to +\infty$ *as* $n \to +\infty$ *for all* $k = 1, \ldots, K$.

In order to prove the theorem we begin with an auxiliary lemma.

Lemma 2 *Let* $\mathcal{Y}^k = \{Y_n^k : n = 0, 1, \ldots,\ k = 1, \ldots, K,$ *be a finite family of submartingales adapted to the same filtration* $\mathcal{F}$ *and such that* $Y_n^k \in [0, 1]$ *for all* k, n*. Suppose also that for any* $\varepsilon > 0$ *there exists a* $\delta > 0$ *such that, for all* k *and* n*,*

$$\mathbb{E}_{\mathcal{F}_n}(Y_{n+1}^k - Y_n^k) > \delta \quad on \quad 0 < Y_n^k < 1 - \varepsilon.$$

Suppose that the initial vector $Y_0 \in A = [0, 1]^K$ *is deterministic and satisfies the condition*

$$\sum_{k=1}^{K} Y_0^k > K - 1. \tag{4}$$

Then, with positive probability, $Y_n^k \to 1$ *as* $n \to \infty$*, for all* $k = 1, \ldots, K$.

Proof. Since each submartingale Y_n^k is bounded, there is a limit $\lim_{n\to\infty} Y_n^k = Y^k$ almost surely for all k, see the Martingale Convergence Theorem in [D]. Let us first show that the limit vector $Y = (Y^k)$ has its support on the union of the 'maximal' vertex $(1, \ldots, 1)$ of the cube A and the 'lower boundary' B of A: $B = \{a : \min_{k=1,\ldots,K} a_k = 0\}$. Indeed, if Y had parts of its support on the complement C of this union in A, then there would exist a k and $0 < \alpha < \beta < 1$ such that, $P(Y^k \in (\alpha, \beta)) > 0$. This would imply that for some $\rho > 0$,

$$\begin{aligned}
\mathbb{E}\left[(Y_{n+1}^k - Y_n^k)\mathbb{I}_{\left\{Y_n^k \in (\alpha,\beta)\right\}}\right] &= \mathbb{E}\left[\mathbb{E}_{\mathcal{F}_n}\left[(Y_{n+1}^k - Y_n^k)\mathbb{I}_{\left\{Y_n^k \in (\alpha,\beta)\right\}}\right]\right] \\
&\geq \mathbb{E}\left[\mathbb{E}_{\mathcal{F}_n}\left[\rho\mathbb{I}_{\left\{Y_n^k \in (\alpha,\beta)\right\}}\right]\right] \\
&= \rho\mathbb{P}(Y_n^k \in (\alpha, \beta)) \to_{n\to\infty} a > 0.
\end{aligned}$$

But this contradicts the fact that

$$\lim_{n\to\infty} \mathbb{E}\left[(Y^k_{n+1} - Y^k_n)\mathbb{I}_{\{Y^k_n \in (\alpha,\beta)\}}\right] = 0,$$

which follows from the convergence a.s. of Y^k_n and the dominated convergence theorem.

To conclude the proof of the lemma, note that for any random vector $V = (V^k)$ with support in B,

$$\sum_{k=1}^{K} V^k \leq K - 1. \tag{5}$$

But, by the submartingale property, for all $n = 1, \ldots,$

$$\mathbb{E}\sum_{k=1}^{K} Y^k_n \geq \sum_{k=1}^{K} Y^k_0 > K - 1. \tag{6}$$

Inequalities (4)-(6) rule out the option that Y has its support in B. ■

Now, in order to derive Theorem 1 from Lemma 2, we make the following change of variables for submartingales X^k_n. For a positive parameter $\alpha < 1$ we define an 'irregular lattice' $h_i \in \mathbb{R}_+$, by

$$h_0 = 0, \quad h_{i+1} = h_i + \alpha^i, \quad i = 0, 1, \ldots.$$

We get $\lim_{i\to\infty} h_i = H = (1-\alpha)^{-1} < \infty$. Now, for each $k = 1, \ldots, K$, we define the process Y^k_n on the same filtration $\mathcal{F}$ by the relation

$$Y^k_n(\omega) = h_{X^k_n(\omega)}.$$

For $k = 1, \ldots, K$, the process Y^k_n takes its values on the 'lattice' $\{h_i\}$. It is easy to see that under the assumptions that $|X^k_{n+1} - X^k_n| \leq R$ and that $\mathbb{E}_{\mathcal{F}_n}(X^k_{n+1} - X^k_n) \geq \rho$, for each k, Y^k_n is still a submartingale, provided $1-\alpha$ is small enough. Then the hypothesis of Lemma 2 holds (up to a constant factor H) and Theorem 1 is proved.

2.1.3 Transience

Let us return to the process $\mathcal{M}(t)$. Suppose that the parameters $\lambda > 0$ and $\beta > 0$ are fixed. We remind the reader that our service rate is set to 1.

Theorem 3 *For each $\lambda > 0$ and $\beta > 0$, there exists $K^* \in \mathbb{Z}_+$ such that for any $K \geq K^*$ the process $\mathcal{M}$ is transient.*

The proof consists of the following steps: first of all, we construct a discrete time Markov chain $\mathcal{D}$ on the state space Q. To define this chain, we start with the embedded Markov chain $M(n)$, defined earlier, and then pass to the chain $M(\Lambda n)$, with the integer Λ to be specified later. To get the chain $\mathcal{D} = \{\mathcal{D}_n\}$, we modify the chain $M(\Lambda n)$ as follows: if for some n some of the K queues are less than or equal to Λ, we make all queues to be of length exactly Λ, and then freeze the process at this point forever. Otherwise we do no changes. We start the process $\mathcal{D}$ at some configuration Q_0 with all queues longer than Λ. We then prove the following statement: if Λ is large enough, the process $\mathcal{D}$ at any given server k is a submartingale satisfying the conditions of Theorem 1, with respect to the filtration defined by the discrete-time Markov chain $M(n)$ (the individual queue length processes are clearly adapted to this filtration). This will then complete the proof thanks to Theorem 1.

These steps are embodied in a series of lemmas.

Consider the following function $\pi(t)$ of the process $\mathcal{M}$ defined in Subsection 2.1.1. At each $t \geq 0$, let $\pi(t)$ be the current permutation of indices of the K servers with respect to the K nodes and let $v(i,t)$ denote the current position of server i.

Lemma 4 *As $t \to \infty$,*

1. *(1) the distribution of $\pi(t)$ converges to the uniform distribution on the set S_K of all permutations;*

2. *(2) the distribution of $v(i,t)$ converges to the uniform distribution on $\{1, \ldots, K\}$ as $t \to \infty$.*

Proof. Note that $\{\pi(t)\}$ is a continuous time Markov process on its own. To best represent this Markov process, let us introduce a graph structure on the permutation group S_K. Consider all the transpositions $\tau \in S_K$ corresponding to the exchanges of pairs of neighboring servers. We connect two permutations π', π'' by an edge iff $\pi' = \pi''\tau$ for some τ.

The resulting graph on S_K is connected – because G_K is connected and each of its nodes has the same degree. The process of migration of servers is a uniform random walk on this graph, that is, a reversible process. Hence,

as $t \to \infty$, the distribution of permutations converges to the uniform one, for all initial states. The assertions of the lemma follow. ■

Lemma 5 *For all initial states Q_0, the probability that a customer with position $H > 0$ in some queue leaves the network after being served at this queue tends to $3/K$ as $H \to \infty$, uniformly in Q_0.*

Proof. As the waiting time of the customer tends to infinity with $H \to \infty$, the distribution of its server on V_K tends to the uniform one on C_K (see Lemma 4). In order for the customer c to exit the network, the last server visited by c has to be located at this moment at one of the three nodes: $D(c)+1$, $D(c)$ or $D(c)-1$. ■

Hence, for all customers in the initial queues whose positions are at least H, the mean chance of exit approaches $3/K$ as $H \to \infty$ and the rate of this approach does not depend on the particularities of the initial state Q_0, but only on H.

The next remark is that if a customer with initial position H, with H large, is served and then jumps to a different server, then the index j of that server is distributed almost uniformly over the remaining $K-1$ indices. This fact follows from Lemma 5. Again, the rate of convergence is independent of Q_0 because the servers swap their positions independently of anything else. So we have established:

Lemma 6 *The probability that a customer with position H on server i jumps to server j at the completion of its service in this server tends to $1/(K-1)$ as $H \to \infty$, uniformly in $i \neq j$ and in the initial states $Q_0 \in Q$.*

We need a third combinatorial lemma. We start with some definitions before formulating and proving this lemme. Let $\{u, v\}$ be an ordered pair of elements of V_K. We define the map T from the set of all such pairs to $V_K \cup \{*\}$, by

$$T\{u,v\} = \begin{cases} w & \text{for } w \text{ defined by } |u-w| = 1,\ |v-w| = |u-v|-1, \\ & \text{provided } |u-v| > 1, \\ * & \text{otherwise.} \end{cases}$$

For K odd, the map T is well-defined. In case $T\{u,v\} = w$, we say that a customer transits through w on his way from u to v.

Let $D : V_K \to V_K$ be an arbitrary map. We want to compute the quantity

$$p_K = \frac{1}{K!} \sum_{\pi \in S_K, i \in V_K} \mathbb{I}_{\{T\{\pi(i), D(i)\} = \pi(j)\}}, \tag{7}$$

where S_K is the symmetric group, while j is an element of V_K. That is, p_K is the continuous time transit rate through server j in stationary regime – when the locations of servers are uniformly distributed on S_K – provided all servers have infinite queues. Of course, this rate does not depend on j.

Lemma 7

$$p_K = \frac{K-3}{K}.$$

Proof. Without loss of generality we can take $j = 1$. Instead of performing the summation in (7) over whole group S_K, we partition S_K into $(K-2)!$ subsets A_π, and perform the summation over each A_π separately. If the result will not depent on π, we are done. Here $\pi \in S_K$, and, needless to say, for π, π' different we have either $A_\pi = A_{\pi'}$ or $A_\pi \cap A_{\pi'} = \varnothing$.

Let us describe the elements of the partition $\{A_\pi\}$. So let π is given, and the string $i_1, i_2, i_3, ..., i_l, i_{l+1}, ..., i_K$ is the result of applying the permutation π to the string $1, 2, ..., K$. Then we include into A_π the permutation π, and also $K-1$ other permutations, which correspond to the cyclic permutations, e.g. we add to A_π the strings $i_K, i_1, i_2, i_3, ..., i_l, i_{l+1}, ..., i_{K-1}, i_K, i_1, i_2, i_3, ..., i_l, i_{l+1}, ...$, and so on. We call these transformations 'cyclic moves'. Now with each of K permutations already listed we include into A_π also $K-2$ other permutations, where the element i_1 does not move, and the rest of the elements is permuted cyclically, i.e., for example from $i_{K-1}, i_K, i_1, i_2, i_3, ..., i_l, i_{l+1}, ...$, we get $i_K, i_2, i_1, i_3, ..., i_l, i_{l+1}, ..., i_{K-1}$, $i_2, i_3, i_1, ..., i_l, i_{l+1}, ..., i_{K-1}, i_K$, and so on. We call these transformations 'restricted cyclic moves'. The main property of thus defined classes of configurations is the following: Let $a \neq b \in \{1, 2, ..., K\}$ be two arbitrary indices, and $l \in \{2, ..., K\}$ be an arbitrary index, different from 1. Then in every class A_π there exists exactly one permutation π', for which $i_1 = a$ and $i_l = b$.

Given π, take the customer $l \neq 1\,(= j)$, and its destination, $D(l)$. If we already know the position i_1 of customer 1 on the circle C_K, then in the class A_π there are exactly $K-1$ elements, each of them corresponds to a different position of the server l on C_K. If it so happens that $i_1 = D(l)$, then for no position of the server l the transit from l through i_1 happens. The same

also holds if $i_1 = \left(D(l) + \frac{K-1}{2}\right) \bmod K$ or $i_1 = \left(D(l) + \frac{K+1}{2}\right) \bmod K$. For all other $K-3$ values of i_1 the transit from l through i_1 happens precisely for one position of l (among $K-1$ possibilities). Totally, within A_π we have $(K-1)(K-3)$ transit events. Since $|A_\pi| = K(K-1)$, the lemma follows. ■

Proof of Theorem 3. We recall that M_n^k denotes the queue length of server k at the n-th transition of the chain $M(t)$ and that $\{\mathcal{F}_n\}$ denotes the natural filtration of the chain $\{M(n)\}$.

We now define the submartingales $\{X_n^k\}$ and show that they satisfy all the properties of Theorem 1.

Let Λ be a positive integer and for all k, let

$$X_n^k = M_{n\Lambda}^k - \Lambda, \quad n = 0, 1, \ldots$$

as long as the R.H.S. is positive, and 0 from the first n such that it is less than or equal to 0 (see above). Clearly, $X_n^k \geq 0$. We now show that if K and Λ are both suitably large, then, for all k, $\{X_n^k\}$ is a submartingale w.r.t. the filtration $\mathcal{G}_n = \mathcal{F}_{\Lambda n}$, and in addition, Properties (1) and (2) of Theorem 1 hold.

On $X_n^k = 0$, the submartingale property $\mathbb{E}_{\mathcal{G}_n} X_{n+1}^k \geq X_n^k$ is satisfied as $X_n^k = 0$ implies $X_{n+1}^k = 0$.

Relation (3) is evidently satisfied with $R = \Lambda$. Let us now check (2). Let us start the process M at a configuration where all the queue lengths are of the form $X_0^k + \Lambda$ with $X_0^k > 0$, $k = 1, ..., K$. We want to show that $\mathbb{E}_{\mathcal{F}_0}(M_\Lambda^k - M_0^k) \geq \rho$, for some $\rho > 0$, which will prove (2) and the submartingale property on $X_n^k > 0$. Let $H = H(K)$ be the time after which the distribution of the K servers is almost uniform on C_K, see Lemma 5. Before this moment, we do not know much about our network; we can nevertheless bound the lengths of the queues M_H^k from below by $M_H^k \geq M_0^k - H$. After time H, the probability that a customer leaving a server leaves the network is almost $1/K$, and the probabilities that it jumps to the left or the right are both close to $\frac{K-1}{2K}$. More precisely, by Lemma 7, if Λ is large enough, after time H, the rate of arrival to every server is approximately $\lambda + (K-3)/K$, which is higher than the exit rate, 1, provided K is large enough (namely, $K > K^* = 3/\lambda$). Hence the expected queue lengths in the process M grow linearly in time, at least after time H, which implies the existence of $\Lambda > 0$ such that $\mathbb{E}_{\mathcal{F}_0}\left(M_\Lambda^k\right) \geq M_0^k + \rho$. So, Theorem 1 applies. This completes the proof of Theorem 3. ■

2.2 Mean-field version of the cyclic network

In this subsection, we analyze the mean field graph C_K^N defined in the introduction, where at each vertex $v \in C_K$ we now have N servers. The dynamics of the system is a modification of the case $N = 1$ with the following characteristics: the

- exogenous customers arrive to each node (v, n) $v = 1, ..., K, n = 1, ..., N$ with the rate λ;
- the destination of an exogenous customer is a node w in C_K (and not a node in C_K^N);
- if a customer c completes its service at a node (v, n), then it leaves the network in case its destination $D(c)$ is v or $v \pm 1$; otherwise it transits to the node $(\tilde{v}, k)$, where $\tilde{v}$ is the node which is the closest to $D(c)$ among $v + 1$ and $v - 1$, while k is chosen uniformly from the N values $1, ..., N$;
- the two servers at locations (v, n) and $(v + 1, k)$ swap with rate $\frac{\beta}{N}$.

The results in this case, as well as the proofs, are similar to those of Subsection 2.1, except for the analogue of Lemma 7. Below we define the corresponding ensemble and we formulate the analogous statement.

Two nodes (v, k) and (v', k') of C_K^N are connected by an edge iff v and v' are connected by an edge in G. Let $\{(u, k), (v, l)\}$ be an ordered pair of nodes of C_K^N. We define the *random* map T from the set of all such pairs into the union $V_K \times N \cup \{*\}$, by

$$T\{(u,k),(v,l)\} = \begin{cases} (w,m) \text{ with probability } \frac{1}{N} & \begin{array}{c} \text{if } |u-v| > 1, \\ w \text{ satisfies } |u-w| = 1, \\ |v-w| = |u-v| - 1, \end{array} \\ * & \text{otherwise.} \end{cases}$$

For K odd, the map T is well-defined. In case $T\{(u, k), (v, l)\} = (w, m)$ we say that we have a transit of a customer through (w, m).

We define a subgroup $\tilde{S}_K \subset S_{KN}$ of permutations of the nodes of the graph $G \times N$ as the one generated by the transpositions $(u, k) \leftrightarrow (v, l)$, where $u \leftrightarrow v$ is a transposition from S_K, and k, l are arbitrary.

Let $D : V_K \times N \to V_K$ be an arbitrary map. The analogue of p_K in (7) is the quantity

$$p_{K,N} = \frac{1}{\left|\tilde{S}_K\right|} \sum_{\pi \in \tilde{S}_K, (u,k) \in V_K \times N} \sum_{(u,k)} \mathbb{P}\left(T\left\{\pi\left(u,k\right), D\left(u,k\right)\right\} = \pi\left(w,m\right)\right).$$

By arguments similar to those of the last subsection, it is easy to show that $p_{K,N} = \frac{K-3}{K}$. Hence, the following theorem holds:

Theorem 8 *For all $\lambda > 0$, $\beta > 0$ and $N \geq 1$, and for all $K > K^* = 3/\lambda$, the Markov process $\mathcal{M}_{K,N}$ is transient.*

2.3 The toric network

In this subsection, the graph is $\mathcal{T}_{KL} = (V_{KL}, E)$, the discrete torus of size $K \times L$. We assume K, L to be odd.

The dynamics of the network is a straightforward generalization of that of the C_K case. Again, the results and the proofs are similar, after we prove the analog of Lemma 7.

We can fix the labelling $(1,1), (1,2), ..., (K,L)$ on $\mathcal{T}_{KL}$; without loss of generality we can take $j = (1,1)$. But it is notationally more convenient to introduce other coordinates on $\mathcal{T}_{KL}$. Namely, we treat $\mathcal{T}_{KL}$ as a product, $\mathcal{T}_{KL} = \left\{-\frac{K-1}{2}, ..., \frac{K-1}{2}\right\} \times \left\{-\frac{L-1}{2}, ..., \frac{L-1}{2}\right\}$, and, for our tagged element j, we now take $j = (0,0)$.

For every ordered pair $\{u,v\}, u, v \in V_{KL}$ such that $|u-v| > 1$, we define the set on next hop nodes from u to v as

$$W(u,v) = \{w \in V_{KL} : |u-w| = 1, |v-w| = |u-v| - 1\}.$$

Clearly, $|W(u,v)|$ is either 1 or 2. We define the *random* map T from the set of all ordered pairs $\{u,v\}, u, v \in V_{KL}$ into the $V_{KL} \cup \{*\}$ by

$$T\{u,v\} = \begin{cases} w \text{ with probability } \frac{1}{|W(u,v)|} & \text{if } |u-v| > 1, \\ & w \in W(u,v), \\ * & \text{otherwise.} \end{cases}$$

In case $T\{u,v\} = w$ we say that we have a transit of a customer through w.

Let $D : V_{KL} \to V_{KL}$ be an arbitrary map. Let

$$p_{KL} = \frac{1}{|S_{KL}|} \sum_{\pi \in S_{KL}, u \in V_{KL}} \mathbb{P}\left(T\left\{\pi\left(u\right), D\left(u\right)\right\} = \pi\left(w\right)\right).$$

The proof of the following lemma is forwarded to the appendix. This lemma is also a corollary of Lemma 11 below.

Lemma 9 *For $K, L \geq 3$*

$$p_{KL} = \frac{KL - 5}{KL}.$$

Hence, the following analogue of Theorem 3 holds.

Theorem 10 *For each $\lambda > 0$, $\beta > 0$ and for each K and L such that $KL > K^* = 5/\lambda$, the process $\mathcal{M}_{K,L}$ is transient.*

2.4 Regular graphs

Let us recall that a graph $G = (V, E)$ is called g-regular if every vertex has g edges adjacent to it. In this section we show that the instability result established above actually holds for all connected and g-regular graphs G. Clearly, the graphs C_K and $\mathcal{T}_{KL}$ are simple instances of such graphs.

Let us define p_G analogously to the definition of p_{KL} in Lemma 9:

$$p_G = \frac{1}{K!} \sum_{\pi \in S_K, i \in V} \mathbb{P}\left\{T\left\{\pi\left(i\right), D\left(i\right)\right\} = \pi\left(j\right)\right\}, \tag{8}$$

where $K = |V|$.

If G is connected and g-regular, then the symmetric nearest neighbor random walk on it has the uniform measure as its unique stationary state. Hence, lemmas 4, 5 and 6 hold for G. So the only step needed is the following generalization of Lemma 7:

Lemma 11

$$p_G = \frac{|V| - (g+1)}{|V|}. \tag{9}$$

Proof. Let us label by $i = 1, 2, ..., |V|$ the nodes and the servers of our network. Suppose that server i is initially located at node i and at node $\pi_t(i)$ at time t, where $\pi_t \in S_{|V|}$ is a random permutation. Let i, j be two indices. We want to compute the stationary transit rate from server i to j, assuming server i has an infinite backlog of customers. For the transit event to happen, it is necessary that the two nodes $\pi_t(i)$ and $\pi_t(j)$ be neighbors. The fraction of time it is the case is equal to $\frac{g}{|V|-1}$ (in the stationary regime).

If it does happen, then there are two options. The first is that customer c, served at $\pi_t(i)$, leaves the network. This happens with probability $\frac{(g+1)}{|V|}$. The complementary event has probability $\frac{|V|-(g+1)}{|V|}$. In this case, customer c jumps to one of the g neighboring nodes of $\pi_t(i)$. Since server $\pi_t(j)$ can be at any of these g nodes with probability $\frac{1}{g}$, independently of the destination of c, the probability that c will land on $\pi_t(j)$ is $\frac{1}{g}$. Since there are $|V|-1$ servers different from server j, we have:

$$p_G = (|V|-1)\frac{g}{|V|-1}\frac{|V|-(g+1)}{|V|}\frac{1}{g} = \frac{|V|-(g+1)}{|V|}.$$

■

Hence, the following theorem holds:

Theorem 12 *For each $\lambda > 0$, $\beta > 0$, and for each $K > K^* = (g+1)/\lambda$, the Markov process $\mathcal{M}_G$ is transient.*

2.5 General graphs

Analogously, the following result holds for a connected graph G that is not regular. Denote by g the maximum degree of vertices $v \in V$ and let $K = |V|$.

Theorem 13 *For each $\beta > 0$, the process $\mathcal{M}_G$ as well as each mean-field processes $\mathcal{M}_{G^N}$, $N = 1, 2, \ldots,$ is transient whenever $K > K^* = (g+1)/\lambda$.*

The only difference in the proof is that the equality (9) is replaced by the estimate

$$p_G \geq \frac{|V|-(g+1)}{|V|}.$$

We omit the proof.

3 Mean-field infinite networks

This section is focused on the mean-field version of certain infinite networks. We first consider the networks $(\mathbb{Z}^1)^N$ as $N \to \infty$. This leads to a NLMP on $\mathbb{Z}^1$, which we study using the methodology introduced in [BRS] for a more general setting. This approach is then generalized to Cayley graphs of discrete groups.

3.1 Non-linear Markov processes on $\mathbb{Z}^1$

In this section we consider the limit of the network $(\mathbb{Z}^1)^N$ as $N \to \infty$, and more precisely the stationary distributions of this limiting network. We focus on translation-invariant distributions, where invariance is w.r.t. translations on $\mathbb{Z}$.

In this case, the NLMP is a dynamical system acting on measures μ on the state of a queue. The state of a queue q_v can be identified with the sequence of (signed) distances between the position of server v and the destinations $D(c_i)$ of its customers. So the state becomes a finite integer-valued sequence $\mathcal{N} \equiv \{n_1, ..., n_l; n_i \in \mathbb{Z}^1\}$, where $l \geq 0$ is the length of the queue q_v.

The rate of arrivals of transit customers, leading, say, from the state $[n_1, ..., n_{l-1}]$ to the state $[n_1, ..., , n_{l-1}, n_l]$, is then a function of the integer n_l only, that we will denote by ν_{n_l}. With the above notation, we thus have

$$\nu_k \equiv \nu_k(\mu) = \begin{cases} \sum_{\mathcal{N}} \mu(k+1, \mathcal{N}) & \text{if } k > 0, \\ \sum_{\mathcal{N}} \mu(k-1, \mathcal{N}) & \text{if } k < 0, \\ 0 & \text{if } k = 0. \end{cases} \tag{10}$$

In what follows we look only for states μ which have symmetric rates ν_k, namely such that

$$\nu_k = \nu_{-k}, \tag{11}$$

for all $k \in \mathbb{Z}$. We also assume (for the sake of simplicity) that the destination of a customer arriving at a node is this very same node.

The following result leverages the methodology developed in [BRS]. It provides a functional equation for fixed points of the NLMP. Each such fixed point is a stationary regime of the NLMP.

Since the evolution of the NLMP is described by an infinite-dimensional dynamical system in a space of probability measures (see Appendix), there might exist non-trivial invariant sets of this evolution in the space of probability measures (not just fixed points) Hence, other non-trivial stationary measures might exist. The existence of non-trivial attractors for the NLMP is discussed in [RSV]. In the present paper, we restrict ourselves to fixed points. The notation is that of the finite network case.

Theorem 14 *Under the foregoing assumptions, each fixed point μ of the*

NLMP satisfy the functional equation:

$$\begin{aligned}
&\mu\left(n_1,...,n_{l-1}\right)\left[\nu_{n_l}(\mu)+\lambda\delta\left(n_l,0\right)\right]-\mu\left(n_1,...,n_l\right)\left(\sum_{k\neq 0}\nu_k(\mu)+\lambda\right)\\
&+\sum_k \mu\left(k,n_1,...,n_l\right)-\mu\left(n_1,...,n_l\right)\mathbb{I}_{l\neq 0} \qquad (12)\\
&+\beta\left[\mu\left(n_1+1,...,n_l+1\right)+\mu\left(n_1-1,...,n_l-1\right)-2\mu\left(n_1,...,n_l\right)\right]=0\ .
\end{aligned}$$

The proof is forwarded to the Appendix. This equation has a simple interpretation. The term

$$\mu\left(n_1,...,n_{l-1}\right)\left[\nu_{n_l}(\mu)+\lambda\delta\left(n_l,0\right)\right]$$

is the arrival rate of the NLMP leading to state $[n_1,...,n_l]$. The term

$$\mu\left(n_1,...,n_l\right)\left(\sum_{k\neq 0}\nu_k(\mu)+\lambda+2\beta\right)$$

is the total rate out of $[n_1,...,n_l]$. The term

$$\sum_k \mu\left(k,n_1,...,n_l\right)-\mu\left(n_1,...,n_l\right)\mathbb{I}_{l\neq 0}$$

is the departure rate leading to state $[n_1,...,n_l]$. The term

$$+\beta\left[\mu\left(n_1+1,...,n_l+1\right)+\mu\left(n_1-1,...,n_l-1\right)\right]$$

is the swap rate leading to state $[n_1,...,n_l]$.

As we will see later, Equations $(10)-(12)$ can have several solutions, one solution or no solution, depending on the value of the parameter λ. If μ is a solution of Equations $(10)-(12)$ for some λ, then we denote by

$$\nu\left(\mu\right)=\sum_{k\neq 0}\nu_k\left(\mu\right) \qquad (13)$$

the rate, in state μ, of the transit customers to every node and by $\eta\left(\mu\right)$ the rate, in state μ, of the total flow to every node:

$$\eta\left(\mu\right)=\nu\left(\mu\right)+\lambda. \qquad (14)$$

Theorem 15 *For every positive $\eta < 1$ there exists a unique value $\lambda(\eta)$ of the exogenous flow rate λ and a unique measure μ_η on the set of queue states satisfying Equations* (10) – (12) *with $\lambda = \lambda(\eta)$, and such that $\eta(\mu_\eta) = \eta$.*

Proof. In the mean-field limit, the total inflow rate to each node $v \in \mathbb{Z}$ is a Poisson process with the rate η for all v. It is splitted according to the possible destinations, $v+h$, $h \in \mathbb{Z}$ of the arriving customers. The customers arriving to v with destination $v+h$ also form a Poisson process with rate ν_h, so that we have

$$\eta = \sum_{h \in \mathbb{Z}} \nu_h,$$

with $\nu_0 = \lambda$. All these arrival Poisson processes are independent.

Consider the random variable ξ_η, which is the total time the customer spends in any given server in the stationary regime. It has exponential distribution, which depends only on $\eta = \sum_{k \neq 0} \nu_k + \lambda$, which does not depend on the customer type. It is defined uniquely by its expectation, which is $\mathbb{E}(\xi_\eta) = (1-\eta)^{-1}$.

Consider now some tagged customer. Suppose it has type k when arriving to the tagged server. When it leaves the server, its type is changed to $k + \tau_\eta$, where τ_η is an integer valued random variable. This change happens due to the fact the server can move during the service time of the tagged customer, i.e. to β-terms in (12). By symmetry, $\mathbb{E}(\tau_\eta) = 0$. The distribution of τ_η is the following. Consider a random walker $W(t)$, living on $\mathbb{Z}^1$, which starts at 0 (i.e. $W(0) = 0$) and which makes ± 1 jumps with rates β. Then $\tau_\eta = W(\xi_\eta)$.

The above observations lead to the following characterization of the rates ν_k obtained from the stationary distribution of the following ergodic Markov process on $\mathbb{Z}^1$. Define the probability transition matrix $P_1 = \{\pi_{st}\}$ by $\pi_{st} = \Pr(\tau_\eta = s - t)$. Of course, this Markov chain on $\mathbb{Z}^1$ is not positive recurrent since its mean drift is zero. Let P_2 be a second Markov chain, with transition probabilities

$$\rho_{st} = \begin{cases} 1 & \text{for } t = s-1, \quad s \geq 2, \\ 1 & \text{for } t = 0, \ s = 1, 0, -1, \\ 1 & \text{for } t = s+1, \ s \leq -2, \\ 0 & \qquad \text{in other cases.} \end{cases}$$

The map P_2 is non-random map of $\mathbb{Z}_1$ into itself. Consider the composition Markov chain, with transition matrix $Q = P_1 P_2$. This chain, which will be referred to as the single particle process below, is positive recurrent (it has

a drift towards the origin), and it hence has a unique stationary distribution $q = \{q_k, k \in \mathbb{Z}^1\}$. We take

$$\nu_k = \eta q_k, \ k \neq 0; \ \lambda = \eta q_0. \tag{15}$$

Consider now the evolution of one single server queue with infinitely many types $k \in \mathbb{Z}$ of customers, arriving to the queue according to Poisson point processes with rates ν_k, $k \neq 0$, and with rate λ for $k = 0$, as defined by Equation (12).

Assume in addition that all customer types in the queue are incremented of one unit according to a global exponential clock with rate β, and decremented of one unit according to another independent global exponential clock, also with rate β. This queuing process is an irreducible Markov process, and since $\eta < 1$, it is ergodic, so that it has a unique stationary distribution. Denote by $\mu(n_1, ..., n_l)$ the stationary probability of state $n_1, ..., n_l$ for this queue. By definition, these probabilities satisfy (12).

In addition, it follows from the fact that $\{q_k\}$ is the steady state of the Markov chain Q that the rate ν_k (with $k > 0$, say) coincides with the probability to find the queue in the state with the first customer having the type $k + 1$. But this is exactly relation (10).

Note that the rates ν_k are defined in a unique way, once η is given, see above. As a result, the same uniqueness holds for the probabilities $\mu(n_1, ..., n_l)$. This proves the existence and the uniqueness statements of the theorem. ■

We now state some properties of the function $\lambda(\eta)$ as the parameter η varies in $(0, 1)$.

Proposition 16 *There is a $\lambda_+ > 0$ such that, for any positive $\lambda < \lambda_+$, there are at least two different values $\eta = \eta_-(\lambda)$ and $\eta = \eta_+(\lambda)$ satisfying the relation $\lambda(\eta) = \lambda$ and such that $\eta_-(\lambda) \to 0$ and $\eta_+(\lambda) \to 1$ as $\lambda \to 0$.*

Proof. Clearly, $\lambda(\eta) \to 0$ as $\eta \to 0$. We want to argue that $\lambda(\eta) \to 0$ also when $\eta \to 1$. Indeed, in this regime every customer spends more and more time waiting in the queue, so for every k the probability $\Pr(\xi_\eta \leq k) \to 0$ as $\eta \to 1$. Therefore the distribution of the random variable τ_η becomes more and more spread out: for every k, $\mathbb{P}\left(|\tau|_\eta \leq k\right) \to 0$ as $\eta \to 1$. Therefore the same property holds for the stationary distribution q, and the claim follows

from (15) and Proposition 16. In particular, this implies that the equation (in η):

$$\lambda(\eta) = a > 0$$

has at least two solutions for small a: η_- close to 0 and η_+ close to 1. This follows from the continuity of $\lambda(\eta)$. ■

3.1.1 Computational illustration

Consider the evolution of the distance between the tagged customer and its destination node in the above mean-field model. This is a continuous time Markov chain on the non-negative integers with the following transition rates: for $n > 1$,

$$q(n, n+1) = \beta, \quad q(n, n-1) = \beta + \gamma,$$

with $\gamma = 1 - \eta$. This is because the tagged customer spends on a given server an exponential time with parameter γ. Similarly,

$$q(1, 2) = \beta, \quad q(1, 0) = \beta, \quad q(1, *) = \gamma$$

and

$$q(0, 1) = 2\beta, \quad q(0, *) = \gamma,$$

where $*$ is absorbing. Let $T(n)$ be the mean time to absorption for a customer at distance n from its destination. The function $T(\cdot)$ satisfies the equations:

$$\begin{aligned}
T(n) &= \frac{1}{2\beta+\gamma} + \frac{\beta}{2\beta+\gamma}T(n+1) + \frac{\beta+\gamma}{2\beta+\gamma}T(n-1), \quad n > 1, \\
T(1) &= \frac{1}{2\beta+\gamma} + \frac{\beta}{2\beta+\gamma}T(2) + \frac{\beta}{2\beta+\gamma}T(0), \\
T(0) &= \frac{1}{2\beta+\gamma} + \frac{2\beta}{2\beta+\gamma}T(1).
\end{aligned}$$

There is exactly one solution of these equations which behaves asymptotically linearly as $n \to \infty$; all the other solutions behave exponentially. This linear function is

$$T(n) = \frac{1}{\gamma}\left(n + \frac{\beta}{\gamma}\frac{2\beta+\gamma}{3\beta+\gamma}\right),$$

for all $n \geq 1$. This in turn determines that

$$T(0) = \frac{1}{\gamma} + \frac{\beta^2}{\gamma^2}\frac{2}{3\beta+\gamma}.$$

Using the assumption that the destination of the tagged customer is the node where it arrives, we get that the mean number of servers visited by the tagged customer till absorption is

$$\mathbb{E}[N] = \gamma T(0) = 1 + \frac{2\beta^2}{\gamma(3\beta+\gamma)}. \tag{16}$$

The total rate in a queue hence satisfies the equation

$$\eta = 1 - \gamma = \lambda\left(1 + \frac{\beta^2}{\gamma}\frac{2}{3\beta+\gamma}\right).$$

When β is large, this boils down to the equation for $\nu = \eta - \lambda$

$$\nu = \frac{2\lambda\beta}{3\gamma} = \frac{2\lambda\beta}{3(1-\lambda-\nu)},$$

or equivalently to the equation

$$3\nu^2 - 3\nu(1-\lambda) + 2\lambda\beta = 0.$$

The discriminant is is positive if $\lambda < \frac{4}{3}\beta - 4$ and in this case, there are two roots

$$\nu^+ = \frac{1-\lambda+\sqrt{(1-\lambda)^2 - \frac{8}{3}\lambda\beta}}{2}$$
$$\nu^- = \frac{1-\lambda-\sqrt{(1-\lambda)^2 - \frac{8}{3}\lambda\beta}}{2}.$$

It is easy to see that under these conditions,

$$0 < \nu^- < \nu^+ < 1 - \lambda,$$

so that these are the two solutions given by the theory.

This computational framework allows one to check the robustness of the proposed framework to the specific assumptions made for mathematical simplicity. One can for instance change the destination of a customer to be at distance d from the arrival node (rather than 0 here), or change the absorption rule to be only when a service completes at the destination (rather than the destination or one of the two neighbors here) and check by computations of the same type that one still finds quite similar phenomena.

3.2 Non-linear Markov processes on Cayley graphs

This subsection extends the previous results in two ways. First, $\mathbb{Z}^1$ is replaced by the Cayley graph of a countable group. Second, we relax the assumption that the destination of an exogenous customer arriving at some node is this same node.

3.2.1 Cayley networks

In this section we rewrite the equations which were studied in the last subsection for the case of $\mathbb{Z}^1$ and we prove a theorem about the structure of the stationary measures δ_μ of the NLMP associated with fixed points. Such measures μ will be called equilibria.

The underlying graph G is assumed to be the Cayley graph of a countable group (also denoted by G) with a finite generating set

$$F = \left\{g_1, ..., g_k, g_1^{-1}, ..., g_k^{-1}\right\}.$$

Typical examples are $\mathbb{Z}^d$ or $\mathbb{T}^d$. The main theorem will focus on the case of an infinite group.

We suppose that the destination assignment rule, the jump direction, the jump rates etc. are all G-invariant. The destination assignment rule is described by $\Lambda = \{\lambda_h, h \in G\}$, where λ_h denotes the rates of external inflows to node e of customers with address h. At all other nodes the external flows have the same structure. Let $\lambda = \sum_{h \in G} \lambda_h$. Denote by X_G the associated NLMP on G.

Consider a customer which finished its service at node v, and assume it has the neutral node $e \in G$ as its destination. Then it goes to the neighboring node $b(v)$ which is closest to e in graph distance on the Cayley graph. If there are several such nodes, say $b_1(v), \cdots, b_R(v)$, $R = R(v)$, then it chooses one of them with probability $\frac{1}{R(v)}$. We only look for equilibria μ which are G-invariant. This means that, under μ, the rate ν of the Poisson flow of transit customers is the same at every node v and that the part of this flow consisting of customers with destination vh has rate ν_h, which does not depend on v. This also means that, in state μ, the probability to have a queue of l customers with destinations $vh_1, ..., vh_l$ at node v depends only on the string $(h_1, ..., h_l)$ of elements of G. We denote this probability by $\mu(h_1, ..., h_l)$.

Denote

$$\eta_h = \nu_h + \lambda_h \quad \text{and} \quad \eta = \sum_{h \in G} \eta_h \tag{17}$$

The functional equations for the stationary measure now take the form

$$0 = -\mu(h_1, \ldots, h_l)\left[(1 - \delta_{l=0}) + \eta\right] + \mu(h_1, \ldots, h_{l-1})\eta_{h_l}$$

$$+ \sum_{h \in G} \mu(h, h_1, \ldots, h_l) + \sum_{g \in F} \beta(\mu(h_1 g, \ldots, h_l g) - \mu(h_1, \ldots, h_l)),\ l = 0, 1, 2, \ldots . \tag{18}$$

Our assumptions on the service discipline imply that the rates ν_h are determined by the measure μ through the following generalization of (10):

$$\nu_h = \sum_{i=1}^{s} \frac{1}{R(v_i)} \sum_{\mathcal{N}} \mu(v_i, \mathcal{N}), \tag{19}$$

where the inner summation is over all finite strings of nodes of G. As in Section 3.1, our goal is to find the rates ν_h, $h \in G$, satisfying (18) and (19).

3.2.2 Single particle process

As in the special case of $\mathbb{Z}^1$ above, it will be useful to follow the evolution of a single customer of the process X_G. In this subsection we describe the associated continuous time Markov random walk.

Since the inflow rates at each node $v \in V$ in the process X_G are Poisson, in equilibrium all queue lengths are distributed geometrically and customers have an exponentially distributed sojourn time T. If the total arrival rate is $\eta < 1$ and the service rate is 1, then T is exponentially distributed with mean H with $H = H(\eta) = 1/(1 - \eta)$.

We now describe the continuous-time Markov process $\mathcal{B}^\eta$ of a an exogeneous particle that arrives to $e \in G$, say, while the probability distribution d_h of its destination node h is given by

$$d_h = \frac{\lambda_h}{\lambda}. \tag{20}$$

The particle makes jumps of two kinds: random jumps due to the jumps of the server harboring it (collectively with all customers sitting there) with rate β, and directed individual jumps to a neighboring server that is closer to

the destination of the particle. The directed jumps happen at random times (service event times) with inter-service intervals distributed exponentially with mean H. If the particle is at distance 0 or 1 from its destination and its service event happens, the particle dies (reaches the absorbing state).

Summarizing, we have the following:

Claim 17 *For all jump rates β, probability distributions $\{d_h\}$ and total rate η, the address of the tagged customer is a continuous time Markov random walk $\mathcal{B}^\eta$ on G.*

For some infinite graphs G (say, for regular trees of degree 3 or more) the process $\mathcal{B}^\eta$ might be transient for all $\eta \in [0,1)$, but this does not happen for β small enough, and we only consider the latter case below. Then the expected number of directed jumps of the particle until absorption is finite for η small enough. Let us denote this expected number by $N(H(\eta))$. The function $N(\cdot)$ is a continuous increasing function, $N : \mathbb{R}_+ \to \mathbb{R}_+ \cup \infty$, such that $N(0) = \mathbb{E}_{\{d_h\}}(\max\{\mathrm{dist}(e,h), 1\})$ and $\lim_{\eta\to 1} N(H(\eta)) = \frac{|V|}{|F|}$.

Some properties of the process X_G can be retrieved from those of the process $\mathcal{B}^\eta$. In particular, one can find the value of the rate λ of the external inflows to each node v from the relation

$$\lambda = \lambda(\eta) = \frac{\eta}{N(H(\eta))}. \tag{21}$$

Indeed, η, which is the load factor per station, is equal to the product of the mean number of arrivals per unit time λ and the mean number of nodes visited by a typical customer.

The function $\lambda(\eta)$ thus defined is continuous on $[0,1]$ and takes values $\lambda(0) = 0$ and $\lambda(1) = |F|/|V|$. For infinite graphs G, we have $\lambda(1) = 0$. Note that in some cases $\lambda(\eta) \equiv 0$. This happens, for instance, if G is a tree of degree $g \geq 3$ and β is large enough.

3.2.3 Equilibria

Here is the generalization of Theorem 15, where G, the exogenous customers rates λ_h, $h \in G$, and the swap rate β are the basic data.

Theorem 18 *For every $0 < \eta < 1$, consider the processes $\mathcal{B}^\eta$ defined by the jump rate parameter β, the probability distribution $\left\{d_h = \frac{\lambda_h}{\lambda}\right\}$, and the rate*

η. Assume that the lifetime of the random walk $\mathcal{B}^\eta$ has finite mean value. For all $0 < \eta < 1$, there exist a unique λ and a unique solution $\mu_\eta(\cdot)$ of (18)-(19) such that the associated transit rates $\{\nu_h, h \in G\}$ satisfy the condition

$$\sum_{h \in G} \nu_h + \lambda_h = \eta.$$

Proof. The proof follows closely that of Theorem 15. The random walker $W(t)$ now lives on G. It starts at e and it makes jumps from each node v to one of the neighboring nodes $\left\{vg_1, ..., vg_k, vg_1^{-1}, ..., vg_k^{-1}\right\}$ with rates β. The matrix of transition probabilities P_1 is now defined via that random walk. Again the corresponding Markov chain on G is not positive recurrent. Let P_2 be the Markov chain with transition probabilities

$$\rho_{st} = \begin{cases} \frac{1}{R(s)} & \text{for } t = b_j(s) \\ 0 & \text{in other cases.} \end{cases}$$

Then we consider the composition Markov chain with transition matrix $Q = P_1 P_2$. The rest of the constructions proceeds in the same way. ■

3.3 Extension to non-exponential service times

Let us describe a simple extension of the results of this section to the case where the customer service times are i.i.d. with unit mean and bounded second moment, again at the nodes of Cayley graph. Let us denote by ξ the random service time of a single customer and by F_ξ its distribution function.

The description of the corresponding NLMP, which will be denoted by $X_{G,\xi}$, requires of course the extra variables – namely, the amounts of service already received by the customers becomes relevant, see [BRS].

For all G-invariant stationary measures of $X_{G,\xi}$, each server receives a total inflow which is a homogeneous Poisson point process of rate η, and we can again decompose η as $\eta = \sum_{h \in G} \eta_h$, as described in the previous subsection, with η_h being the total arrival rate of customers with destination h at server e. Each customer stays at each node for a stationary sojourn time which is that of a $M/GI/1$ queue with parameters η and ξ. Additionally, all customers in this queue change their locations simultaneously with rate β as their server swaps its position with the adjacent server.

Consider the stochastic process $\mathfrak{B}^{\eta,\xi}$ of the random walk of a single particle over the Cayley graph of G till its exit. This process is analogous to the

process $\mathfrak{B}^\eta$ in Section 3.2.2. The only difference between these two processes is that the exponentially distributed random variable T (stationary sojourn time of the system $M/M/1$) is replaced by $T_{\eta,\xi}$, the stationary sojourn time in the queue $M/GI/1$ with i.i.d. service time with distribution F_ξ given by the Pollaczek-Khinchine formula.

As in Section 3.2.2, let us require that the triple (η, ξ, β) ensures finiteness of the mean lifetime of a customer, denoted by $E(\eta, \xi, \beta)$. For fixed ξ and β, denote by $\bar{\eta}$ the right endpoint of the maximal interval of values η where the function $E(\eta, \xi, \beta)$ is finite. Clearly, $\bar{\eta}$ depends continuously on β and, within the interval $[0, \bar{\eta})$, the function $E(\eta, \xi, \beta)$ is continuous in η and β.

Theorem 19 *Consider the NLMP $X_{G,\xi}$. Assume that $E(\eta, \xi, \beta)$ is finite. Let us fix a probability distribution $d_h = \lambda_h/\lambda$, $h \in G$, where $\lambda = \sum_h \lambda_h$. Then there is a unique value of λ such that the process $X_{G,\xi}$ has a unique G-invariant stationary distribution.*

Proof. Let us fix an arbitrary triple (u, v, w) of points of G. For the single particle process $\mathfrak{B}^{\eta,\xi}$, let us denote by $k(u, v, w)$ the mean number of direct jumps to node v (only the jumps of the particle are counted, not the jumps of servers) of the particle that is located at node u and has the address w. One can easily see that η can be decomposed as follows:

$$\eta = \sum_{u,w} \lambda_{v^{-1}w} k(u, v, w). \tag{22}$$

Thanks to G-invariance, η does not depend on v. The right-hand side of this equation is finite as shown when rewriting $k(u, v, w)$ as $k\left(e, vu^{-1}, wu^{-1}\right)$, thanks to G-invariance. So the right-hand side is finite as for $\mathfrak{B}^{\eta,\xi}$ (the single particle random walk) the mean life time of the particle is finite.

Accordingly, for the G-invariant stationary distribution $X_{G,\xi}$ of the NLMP, the rate of the homogeneous Poisson inflow to node e of particles with address h can be written as

$$\eta_h = \lambda_h k(e, e, h) + \sum_{u \neq e} \lambda_{u^{-1}h} k(u, e, h). \tag{23}$$

Also,

$$\nu_h = \sum_{u \neq e} \lambda_{u^{-1}h} k(u, e, h).$$

Since for a given distribution $\{d_h\}$the sum

$$\lambda_h k(e,e,h) + \sum_{u \neq e} \lambda_{u^{-1}h} k(u,e,h)$$

is a linear function of λ, for a given η, Equation (23) has a unique solution $\{\lambda_h, \nu_h, h \in G\}$.

Hence we have a stationary process of arrival to node e (or any other node) of independent Poisson flows of different types k. The total rate of this inflow is $\eta < 1$ and the total queue at a given server is stationary. The customers in this queue change their types to k' after an exponentially distributed time, in the same manner as in the case of exponentially distributed variable ξ.

The process which describes the queue is ergodic since it has renewal times when the queue becomes empty. The product of stationary measures on the queues over all the nodes is the required G-invariant measure for the NLMP $X_{G,\xi}$. ■

Theorem 20 *If $\lambda > 0$ is small enough, the NLMP on a translation invariant infinite graph has at least two translation invariant equilibria.*

Proof. It follows from the continuity of the function $E(\eta, \xi, \beta)$ in η and β and from the relations (23), (22), that λ is a continuous function of η. Moreover, $\lambda \to 0$ as $\eta \to 0$. Analogously, $\lambda \to 0$ as $\eta \to \bar{\eta}$. ■

4 Comparing $\mathbb{Z}^d_K$ and $\mathbb{Z}^d$

In the simplest case $d = 1$ we have the following property.

Theorem 21 *The circle $\mathbb{Z}^1_K$ is faster than the line $\mathbb{Z}^1$ in terms of absorption times.*

Proof. First, we consider $\mathbb{Z}^1_+$ instead of $\mathbb{Z}^1$. Then the circle $\mathbb{Z}^1_K$ can be associated with an interval $[0, K/2]$ within Z^1_+. The two processes are the same on this interval apart from the endpoint $K/2$.

One may say that the process on $\mathbb{Z}^1_+$ is coupled with a special process on $[0, K/2]$ as follows. As two processes are at $K/2$ and the "line" process goes further to $K/2 + 1$, the "circle" process waits at $K/2$ till the "line" process returns to $K/2$. Hence, the absorption time may only increase. ■

In the general case $d \geq 1$ we can only say that in some sense the equilibria on $\mathbb{Z}_K^d$ are close to the equilibrium on $\mathbb{Z}^d$ for K large enough. Namely, the following assertion holds.

Theorem 22 *For each* $\eta \in [0,1]$,

$$\lim_{K\to\infty} \lambda_K(\eta) = \lambda(\eta).$$

We omit the technical proof.

5 Conclusions

Let us conclude with a few observations on the physical meaning of our models and their connections to earlier models of the networking literature.

5.1 Relationship between the models

Let us first discuss the relationships between the replica 1 model, the replica N model and the replica ∞ model.

The replica-1 model and the replica-N model describe two different physical systems as illustrated by the following wireless communication network setting: The replica-N model features a system with N frequency bands and where each device has N radios, one per frequency band, that it can use simultaneously, for both transmission and reception. It then implements a set of N virtual FIFO queues of packets, one per band. At any time, a device transmits the packet head of the queue to the appropriate neighboring device (the one closest to packet destination) on that band. Upon reception, the packet is loaded in one of the N queues of the receiver chosen at random. In this sense, the replica-N model is *implementable*, and the 1-replica method is just the special case with one frequency band.

Let us now explain in what sense the replica-∞ model (which is computational but non implementable) tells us something on the replica-N model (which is implementable but not computational), in spite of the fact that the latter is always unstable whereas the former admits stationary regimes: the replica-N network has a metastable state, which lives for longer and longer times as N grows, and which is well described by the minimal solution of the replica-∞ model.

5.2 Relationship with the Gupta and Kumar models

Our setting is close to that of Gupta and Kumar [GK]. The latter also involves a network with N nodes using hop by hop relaying through nearest neighbors with traffic from every node to a node chosen at random among all others. The initial model has no node mobility. The main finding of [GK] is that the arrival rate that can be sustained on any such node is order $1/\sqrt{N}$ when nodes are located in the Euclidean plane. Indeed, since nearest neighbors are at distance $1/\sqrt{N}$, if packet destination is at distance order 1, then every packet has to travel order $\sqrt{N}$ hops, so that every node has to relay a total flow of order $\lambda\sqrt{N}$ (as seen by analyzing the load brought per queue). Hence the capacity (maximum value of λ for which the system is stable) is order $1/\sqrt{N}$.

The case with mobility was then considered by [GT] and [EMPS]. The main finding is that the above scaling of the capacity is in fact *improved* by node mobility. At first glance, our result suggests that mobility does worse than absence of mobility (for instance, in the setting of the present paper, where packet destination at distance order 1 for graph distance, the static network has a stability region order 1, whereas the mobile one has a stability region that tends to 0 when N tends to infinity).

Let us explain why there is no contradiction in fact: our model can be extended to the setting where the initial distance to destination is order $\sqrt{N}$ for graph distance (rather than 0 or order 1), which will then be comparable in terms of load to that of the Gupta and Kumar setting in the Euclidean plane. In the absence of mobility, our model has a stability region of order $1/\sqrt{N}$, as in Gupta and Kumar, whereas in the presence of mobility, as shown by our analysis, a customer brings a load to order N queues, which is much worse than in the case without mobility. The reason lies in the use of the FIFO discipline as a constraint. If the scheduling in queues were more adaptive than FIFO, we could do exactly as [GT], namely keep the packets in their arrival station and wait for this station to be close to a packet destination to schedule the latter. Then the load per queue is order 1. Of course, delay is terrible (return to 0 of a random walk) as in [GT]. In conclusion, there is no contradiction with this literature.

These connections lead to the following observations:

1. Our model offers a new microscopic view of this class of problems which complements both the Gupta and Kumar [GK] and the Grossglauser and Tse scaling [GT] laws. This microscopic view, describes what hap-

pens in a typical queue, and opens a new quantitative line of thoughts (through mean fields) for this class of problems.

2. Our model does not sacrifice delays (like Grossglauser and Tse) and finds one possible compromise between delay and rate in line with the ideas discussed by [EMPS]. A study of adaptive scheduling (e.g. priority to customers close to destination) might lead to innovative solutions to this class of questions.

6 Appendix

6.1 Proof of Lemma 9

Again we partition the group S_{KL} into cosets. Each coset A_π has now $K(K-1)L(L-1)$ elements. With every permutation π of the KL points of the discrete torus $\mathcal{T}_{KL}$ we first include in A_π all its '2D cyclic moves', i.e. permutation π followed by an arbitrary shift of $\mathcal{T}_{KL}$; there are KL of them. Also, with each permutation $\tilde{\pi}$ from A_π we include in A_π all $(K-1)(L-1)$ permutations which are obtained from $\tilde{\pi}$ by performing a pair of 'independent restricted cyclic moves': one of them cyclically permutes all the sites of the meridian of the point $\pi(0,0)$, and another cyclically permutes all the sites belonging to the parallel of the point $\pi(0,0)$. There are $(K-1)(L-1)$ such independent restricted cyclic moves. All other $(K-1)(L-1)$ points of the torus $\mathcal{T}_{KL}$, as well as the point $\pi(0,0)$, stay fixed during these independent restricted cyclic moves. The idea behind this definition is to ensure the following property: suppose we know for the permutation π that a point $\pi(k',l')$ belongs to the parallel of the point $\pi(0,0)$, while the point $\pi(k'',l'')$ belongs to the meridian of $\pi(0,0)$. Then for any three different points $a, b', b'' \in \mathcal{T}_{KL}$, such that b' belongs to the parallel of a, and b'' belongs to the meridian of a, there exists exactly one permutation $\bar{\pi} \in A_\pi$, such that $\bar{\pi}(0,0) = a$, $\bar{\pi}(k',l') = b'$ and $\bar{\pi}(k'',l'') = b''$.

Now we fix one class A_π, and compute the number of transits to node $\bar{\pi}(0,0)$ for all $\bar{\pi} \in A_\pi$. Without loss of generality and in order to simplify the notations we consider only the case when π is the identity $e \in S_{KL}$. We denote the permutations from A_e by the letter $\varkappa$.

Clearly, the transits from node $\varkappa(k,l)$ to $\varkappa(0,0)$ can happen only if either k or l are 0. So let us fix some integer $k \in \left\{-\frac{K-1}{2}, ..., \frac{K-1}{2}\right\}$, $k \neq 0$, and let us count the number of possible transits from $\varkappa(k,0)$ to $\varkappa(0,0)$ while $\varkappa$

runs over A_e. Without loss of generality we can assume that the destination $D(k,0) = (0,0)$.

As we said already, as $\varkappa$ runs over A_e, the node $\varkappa(0,0)$ can be anywhere on the torus $\mathcal{T}_{KL}$. The node $\varkappa(k,0)$ can be then anywhere on the parallel of $\varkappa(0,0)$. If $\varkappa(0,0) = (0,0)\,(= D(k,0))$, then no transit to $\varkappa(0,0)$ can happen, independently of the location of $\varkappa(k,0)$. The same is true when $\varkappa(0,0) = (x,y)$ with $x = \pm\frac{K-1}{2}$, $-\frac{L-1}{2} \leq y \leq \frac{L-1}{2}$.

For every of the $(K-3)$ locations $(x,0) \in \mathcal{T}_{KL}$ of the node $\varkappa(0,0)$ – namely, for $x = -\frac{K-1}{2}+1, ..., -1, +1, +2, ...\frac{K-1}{2}-1$ we have one transit per location (or, more precisely, $L-1$ transits per location, due to the restricted cyclic moves along the meridian).

For every of the location $(0,y)$ of the node $\varkappa(0,0)$ – namely, for $y = -\frac{L-1}{2}, ..., -1, +1, +2, ..., \frac{L-1}{2}$ we have $2 \times \frac{1}{2} = 1$ transits per location (or, again more precisely, $L-1$ transits per location), since there can be two transit events, each with probability $\frac{1}{2}$.

For any other remaining location of the node $\varkappa(0,0)$ – and there are $(K-3)(L-1)$ of them, we get $\frac{1}{2}$ of transit per location (more precisely, $\frac{L-1}{2}$ transits per location).

Summarizing, we have totally $\left[(K-3)+(L-1)+\frac{1}{2}(K-3)(L-1)\right](L-1)$ transits from the node $\varkappa(k,0)$ to the node $\varkappa(0,0)$, as $\varkappa$ runs over A_e. And there are $(K-1)$ such nodes.

All in all, we have

$$\left[(K-3)+(L-1)+\frac{1}{2}(K-3)(L-1)\right](K-1)(L-1)$$
$$+\left[(L-3)+(K-1)+\frac{1}{2}(L-3)(K-1)\right](L-1)(K-1)$$

transits, so the probability in question is given by

$$\frac{(K-3)+(L-1)+\frac{1}{2}(K-3)(L-1)+(L-3)+(K-1)+\frac{1}{2}(L-3)(K-1)}{KL}$$
$$= \frac{3(K-3)+3(L-3)+(K-3)(L-3)+4}{KL}$$
$$= \frac{3K+3L+(K-3)(L-3)-14}{KL}$$
$$= \frac{KL-5}{KL}.$$

6.2 Proof of Theorem 14

Using the results (and notation) of [BRS], we get that the NLMP is the evolution of the measure $\otimes\mu_v$ on the states (queues q_v) of the servers at the nodes $v \in \mathbb{Z}^1$, given by the equations

$$\frac{d}{dt}\mu_v\left(q_v,t\right) = \mathcal{A}+\mathcal{B}+\mathcal{C}+\mathcal{D}+\mathcal{E} \tag{24}$$

with

$$\mathcal{A} = -\frac{d}{dr_{i^*(q_v)}\left(q_v\right)}\mu_v\left(q_v,t\right) \tag{25}$$

the derivative along the direction $r\left(q_v\right)$ (in our case, since we assume exponential service times with rate 1, we have $\frac{d}{dr_{i^*(q_v)}(q_v)}\mu_v\left(q_v,t\right) = \mu_v\left(q_v,t\right)$),

$$\mathcal{B} = \delta\left(0,\tau\left(e\left(q_v\right)\right)\right)\mu_v\left(q_v\ominus e\left(q_v\right),t\right)\left[\sigma_{tr}\left(q_v\ominus e\left(q_v\right),q_v\right)+\sigma_e\left(q_v\ominus e\left(q_v\right),q_v\right)\right] \tag{26}$$

where q_v is created from $q_v \ominus e\left(q_v\right)$ by the arrival of $e\left(q_v\right)$ from v', and $\delta\left(0,\tau\left(e\left(q_v\right)\right)\right)$ takes into account the fact that if the last customer $e\left(q_v\right)$ has already received some amount of service, then he cannot arrive from the outside;

$$\mathcal{C} = -\mu_v\left(q_v,t\right)\sum_{q_v'}\left[\sigma_{tr}\left(q_v,q_v'\right)+\sigma_e\left(q_v,q_v'\right)\right], \tag{27}$$

which corresponds to changes in queue q_v due to customers arriving from other servers and from the outside (in the notations of (1) , $\sigma_e\left(q_v,q^v\oplus w\right) = \lambda_{v,w}$);

$$\mathcal{D} = \int_{q_v':q_v'\ominus C(q_v')=q_v} d\mu_v\left(q_v',t\right)\sigma_f\left(q_v',q_v'\ominus C\left(q_v'\right)\right)-\mu_v\left(q_v,t\right)\sigma_f\left(q_v,q_v\ominus C\left(q_v\right)\right), \tag{28}$$

where the first term describes the situation where the queue q_v arises after a customer was served in a queue q_v' (longer by one customer), and $q_v'\ominus C\left(q_v'\right) = q_v$, while the second term describes the completion of service of a customer in q_v;

$$\mathcal{E} = \sum_{v'\text{n.n.}v}\beta_{vv'}\left[\mu_{v'}\left(q_v,t\right)-\mu_v\left(q_v,t\right)\right], \tag{29}$$

where the β-s are the rates of exchange of the servers.

For the convenience of the reader we repeat the equation $(24-29)$ once more:

$$\begin{aligned}
&\frac{d}{dt}\mu_v(q_v,t) = -\frac{d}{dr_{i^*(q_v)}(q_v)}\mu_v(q_v,t) \\
&+\delta(0,\tau(e(q_v)))\,\mu_v(q_v\ominus e(q_v))\left[\sigma_{tr}(q_v\ominus e(q_v),q_v)+\sigma_e(q_v\ominus e(q_v),q_v)\right] \\
&-\mu_v(q_v,t)\sum_{q_v'}\left[\sigma_{tr}(q_v,q_v')+\sigma_e(q_v,q_v')\right]+\int_{q_v':q_v'\ominus C(q_v')=q_v} d\mu_v(q_v')\,\sigma_f(q_v',q_v'\ominus C(q_v')) \\
&-\mu_v(q_v)\,\sigma_f(q_v,q_v\ominus C(q_v))+\sum_{v'\text{n.n.}v}\beta_{vv'}\left[\mu_{v'}(q_v)-\mu_v(q_v)\right]\ .
\end{aligned} \tag{30}$$

Compared to the setting of [BRS], we have the following simplifications:

1. The graph G is the lattice $\mathbb{Z}^1$;
2. All customers have the same class;
3. The service time distribution η is exponential, with the mean value 1;
4. The service discipline is FIFO;
5. The exogenous customer c arriving to node v has for destination the same node v; inflow rates at all nodes are equal to λ;
6. The two servers at v, v', which are neighbors in $\mathbb{Z}^1$ exchange their positions with the same rate $\beta\equiv\beta_{vv'}$;

The equation for the fixed point then becomes:

$$\begin{aligned}
0 &= \mu_v(q_v\ominus e(q_v))\left[\sigma_{tr}(q_v\ominus e(q_v),q_v)+\sigma_e(q_v\ominus e(q_v),q_v)\right] \\
&-\mu_v(q_v)\sum_{q_v'}\left[\sigma_{tr}(q_v,q_v')+\lambda\right]+\sum_{q_v':q_v'\ominus C(q_v')=q_v}\mu_v(q_v') \\
&-\mu_v(q_v)\,\mathbb{I}_{q_v\neq\varnothing}+\sum_{v'=v\pm1}\beta\left[\mu_{v'}(q_v)-\mu_v(q_v)\right]\ .
\end{aligned}$$

The proof is concluded when using the fact that queue q_v can in this setting be identified with the sequence of destinations $D(c_i)$ of its customers.

References

[BB] F. Baccelli and B. Błaszczyszyn. *Stochastic Geometry and Wireless Networks*. NOW Publishers, 2009.

[BRS] F. Baccelli, A. Rybko, and S. Shlosman. Queuing Networks with Varying Topology – A Mean-Field Approach. *Information Transmission Problems*, 52(2):178-199, 2016.

[D] R. Durrett. *Probability: Theory and Examples.* Cambridge series on statistical and probabilistic mathematics. Cambridge University Press, 2010.

[EMPS] A. El Gamal, J. Mammen, B. Prabhakar and D. Shah, Throughput-Delay Trade-off in Wireless Networks, Proceedings of IEEE INFOCOM, Hong Kong, 2004.

[GT] M. Grossglauser and D. Tse, Mobility Increases the Capacity of Ad Hoc Wireless Networks, *IEEE Transactions on Networking*, Volume 10, Number 4, pp 477–486, 2002.

[GK] P. Gupta and P. R. Kumar, The Capacity of wireless networks, *IEEE Transactions on Information Theory*, Volume 46, Number 2, pp 388–404, 2000.

[RSV] A. Rybko, S. Shlosman, and A. Vladimirov. Spontaneous resonances and the coherent states of the queuing networks. *J. Stat. Phys.*, 134(1):67–104, 2009.

[SS] R. H. Schonmann and S. Shlosman. *Wulff Droplets and the Metastable Relaxation of the Kinetic Ising Models,* Comm. Math. Phys., 194, 389-462, 1998.