



## Spatial queues with nearest neighbour shifts<sup>☆</sup>

B.R. Vinay Kumar<sup>a,\*</sup>, Lasse Leskelä<sup>b</sup>

<sup>a</sup> Eindhoven University of Technology, P.O. Box 513, 5600 MB, Eindhoven, The Netherlands

<sup>b</sup> Department of Mathematics and Systems Analysis, Aalto University, Otakaari 1, 02150, Espoo, Finland

### ARTICLE INFO

#### Keywords:

Queues  
Electric vehicles  
Nearest neighbour  
Spatial queues  
k-NN graphs  
Spatial point process

### ABSTRACT

This work studies queues in a Euclidean space. Consider  $N$  servers that are distributed uniformly in  $[0, 1]^d$ . Customers arrive at the servers according to independent stationary processes. Upon arrival, they probabilistically decide whether to join the queue they arrived at, or shift to one of the nearest neighbours. Such shifting strategies affect the load on the servers, and may cause some of the servers to become overloaded. We derive a law of large numbers and a central limit theorem for the fraction of overloaded servers in the system as the total number of servers  $N \rightarrow \infty$ . Additionally, in the one-dimensional case ( $d = 1$ ), we evaluate the expected fraction of overloaded servers for any finite  $N$ . Numerical experiments are provided to support our theoretical results. Typical applications of the results include electric vehicles queueing at charging stations, and queues in airports or supermarkets.

### 1. Introduction

Traditional queueing systems involve multiple queues interacting with each other. The analysis of the queueing mechanism and related load balancing questions are tractable owing to the product-form stationary distribution. Most earlier works on queues distributed spatially involve customers arriving at random locations in a Euclidean space and a server travelling to serve them. However, most queueing networks in the present day involve servers that are distributed in space and arrivals deciding among multiple servers. This work addresses the problem of load distribution in such networks where the servers are spatially distributed and the arrivals decide to join a server based on proximity.

As a motivating example, consider the rapidly growing electric vehicles (EV) industry. With increased adaptation of EVs, the charging infrastructure is also being scaled up. However, physical and financial constraints put a cap on the number of charging stations that can be deployed. In such a scenario, it is natural to expect EV users to adopt strategies in order to minimize their waiting times at charging stations. For example, a user on arriving at a charging station and finding it to be occupied, might decide to travel to a farther station in the hope that it would be empty. The user strategy affects the load that is perceived at the servers in the long run. Some of the servers get overloaded which might degrade the performance of the system on the whole. An understanding of the fraction of overloaded servers helps in optimal resource allocation preventing such degradation. Additionally, monitoring the

<sup>☆</sup> A part of this work was accepted to the conference International Teletraffic Congress (ITC 35) held between 3–5 October 2023 in Turin, Italy.

\* Corresponding author.

E-mail addresses: [v.k.bindiganavile.ramadas@tue.nl](mailto:v.k.bindiganavile.ramadas@tue.nl) (B.R. Vinay Kumar), [lasse.leskela@aalto.fi](mailto:lasse.leskela@aalto.fi) (L. Leskelä).

<sup>1</sup> This work was partly done when VK was a post-doctoral researcher at INRIA, Sophia Antipolis, France and when he was visiting Aalto University, Finland in August 2023. VK was supported by the French government through the RISE Academy of UCA<sup>JEDI</sup> Investments in Future project managed by the National Research Agency (ANR), France with reference number ANR-15-IDEX-0001, and the Aalto Science Institute (ASCI) visiting researcher fellowship. VK has also received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement, Grant Agreement No 101034253.

<https://doi.org/10.1016/j.peva.2024.102448>

Available online 16 October 2024

0166-5316/© 2024 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

fraction of overloaded servers can be used to incentivize customers to change their strategies thus increasing the durability of the entire system. Similar considerations arise in other practical networks involving queues, like supermarkets, airports etc.

Motivated by such applications, in this work we consider a set of  $N$  servers that are deployed in a Euclidean space with queues at each of them modelled as stationary arrival processes. In the context of EVs, the servers are the charging stations, and a stationary arrival process models the EV users arriving at a charging station. EV charging stations have been modelled as a Poisson point process in two dimensional space in several previous works (see, for e.g., [1–3]). However, the problem that authors address in these works pertain to optimal placement of charging stations in the underlying space.

Another line of work that considers queues on spaces are polling systems. One of the earlier works in this domain [4] considers multiple queues in a convex space with a single server moving across to serve them. Several later works, such as [5,6], study vehicle routes and delay in such systems. In all of these (and some related) works there is a single server and there is no interaction between different queues.

In contrast, in the present work we consider queues where customers from one queue probabilistically move to a queue which is located close to it in the underlying space. Customers changing queues has been referred to as *jockeying* in the queueing theory terminology. There has been considerable work on jockeying in queues in the past [7–9]. The focus in these works is predominantly to analyse the steady-state distribution or find expected line-lengths or delays. However, they primarily consider two server systems with no spatial component in the problem.

In [10], the authors consider arrivals occurring on the two-dimensional torus, with multiple servers following a greedy strategy to serve the customers by travelling minimally. They find that such a strategy results in servers coalescing making the system inefficient. In contrast, in our work the servers are fixed whereas the arrivals follow nearest neighbour strategies and hence the two works are not directly comparable.

The closest work to ours is [11] where the authors consider load-balancing policies for servers distributed in space. Specifically,  $m$  servers are distributed in a two-dimensional Euclidean space and  $n$  customers arrive distributed uniformly in space. The customers choose the shorter queue among the server they arrived at and a second server sampled uniformly at random from  $k$  geographically nearest servers of the server where the customer arrived. Specifically, the authors observe that such a policy shows a better performance with respect to the maximum load. While these observations are interesting in their own right, the performance metrics of interest and the model setting are considerably different from the current work. Instead of a setting where there are fixed number of arrivals, we consider stationary arrival processes into each of the servers with customers following nearest neighbour mobility strategies. Additionally, instead of the maximum load and the allocation distance, our interest is in the distribution of the number of overloaded servers in the system. These differences render the two works incomparable. However, it should be remarked that there is considerable scope for combining ideas from both the works.

More precisely, in the current work,  $N$  servers are deployed in a Euclidean space and customers arriving into a queue decide whether to stay or to move to a queue nearest to them with a prescribed probability. We call such strategies as nearest neighbour shift (NNS) strategies. The metric of interest is the fraction of overloaded servers in the system which we characterize for users following an NNS strategy.

We begin by describing our system model and state the main results in Section 2. Section 3 provides the required background on nearest neighbour graphs which will be used to characterize the overloaded servers for the NNS strategies. This is followed by the analysis of NNS strategies for general dimension in Section 4 and additional results for one dimension in Section 5. Section 6 provides numerical simulations justifying our results and Section 7 highlights some future directions.

## 2. System model and main results

### 2.1. Model description

Consider  $N$  service stations with locations  $\{X_i\}_{i=1}^N$  uniformly and independently distributed within  $[0, 1]^d$ ,  $d \geq 1$ , equipped with the standard Euclidean metric. Each service station is associated with a stationary arrival process with an *exogenous arrival rate* of  $\lambda$ , a server operating at a service rate  $\mu$ , and a queue of infinite capacity. We use the terms queue  $i$  and server  $i$  to refer to the service station located at  $X_i$ . Customers arriving at a queue can adopt different strategies. In this work, we consider all customers to be identical who make independent decisions. Our interest is in *nearest neighbour shift* (NNS) strategies which we describe next.

**Definition 2.1.** Under the  $(k, p)$ -NNS strategy with activity range  $k \geq 1$  and a shift rate  $p \in [0, 1]$ , customers arriving at queue  $i$  join queue  $i$  with probability  $1 - p$ , and join queue  $j \in \mathcal{N}_k(i)$  with probability  $\frac{p}{k}$ , where  $\mathcal{N}_k(i)$  denotes the set of  $k$  nearest neighbours of  $i$ .

We emphasize the difference between a customer *arriving* at a queue and *joining* a queue. Whereas each queue  $i$  has exogenous arrival rate  $\lambda$ , the rate at which customers join queue  $i$  may be smaller or larger than  $\lambda$ .

**Definition 2.2.** The *effective arrival rate*  $\lambda_{\text{eff}}(i)$  at a server  $i$  is the rate at which customers join queue  $i$ .

Our goal is to characterize the expected fraction of overloaded servers in the system. Because each queue is processed at service rate  $\mu$ , the load at queue  $i$  is equal to  $\rho_i = \frac{\lambda_{\text{eff}}(i)}{\mu}$ , and the fraction of overloaded servers (with  $\rho_i > 1$ ) can be written as

$$\mathcal{O}_N = \frac{1}{N} \left| \left\{ i : \lambda_{\text{eff}}(i) > \mu \right\} \right|, \quad (1)$$

where  $|\cdot|$  denotes the cardinality. Note that  $\mathcal{O}_N$  is a random variable where the randomness is due to the spatial distribution of the servers.

## 2.2. Main results for general dimension

Our main results characterize the fraction of overloaded servers for the  $(k, p)$ -NNS strategy for  $N$  servers randomly distributed within  $[0, 1]^d$ . The results are described in terms of a geometric constant  $\alpha_d$ , defined as the maximum number of points that can be placed on the unit sphere of  $\mathbb{R}^d$  so that the distance between all distinct points is strictly larger than one. It is known [12] that  $\alpha_2 = 5$  and  $\alpha_3 = 12$ , and that  $\alpha_d$  is bounded from above by the kissing number in  $\mathbb{R}^d$ .

**Theorem 2.1.** *Consider  $N$  servers with service rate  $\mu$  that are uniformly distributed in  $[0, 1]^d$  and receive exogenous arrivals at rate  $\lambda$  following the  $(k, p)$ -NNS strategy.*

(i) *If  $\frac{1}{1-p+\alpha_d p} < \frac{\lambda}{\mu} \leq 1$ , then there exist constants  $q_{d,k,n}$  and  $\sigma_{d,k}$  such that for  $N \rightarrow \infty$ , the fraction of overloaded servers satisfies*

$$\mathcal{O}_N \xrightarrow{\text{a.s.}} \sum_{n=[\theta]+1}^{\alpha_d k} q_{d,k,n} \quad \text{and} \quad \sqrt{N}(\mathcal{O}_N - \mathbb{E}\mathcal{O}_N) \xrightarrow{d} \mathcal{N}(0, \sigma_{d,k}^2)$$

$$\text{where } \theta = k + \frac{k}{p} \left( \frac{\mu}{\lambda} - 1 \right).$$

(ii) *If  $\frac{\lambda}{\mu} \leq \frac{1}{1-p+\alpha_d p}$ , then  $\mathcal{O}_N = 0$  a.s. for all  $N$ .*

**Theorem 2.1** confirms that the number of overloaded servers has a normal distribution in the regime when  $N \rightarrow \infty$ . This is illustrated in our numerical results in Section 6. The constants  $q_{d,k,n}, \sigma_{d,k}$  are discussed in Section 3.2. Note that the almost sure ( $\xrightarrow{\text{a.s.}}$ ) and distributional ( $\xrightarrow{d}$ ) limits above refer to the randomness induced by the spatial distribution of the servers.

## 2.3. Main results for one dimension

As a corollary of **Theorem 2.1**, we obtain the following explicit result for dimension  $d = 1$  in which the arrivals follow the  $(1, p)$ -NNS strategy.

**Corollary 2.1.** *Consider  $N$  servers with service rate  $\mu$  that are uniformly distributed in  $[0, 1]$  and receive exogenous arrivals at rate  $\lambda$  following the  $(1, p)$ -NNS strategy.*

(i) *If  $\frac{1}{1+p} < \frac{\lambda}{\mu} \leq 1$ , then the fraction of overloaded servers satisfies*

$$\mathcal{O}_N \xrightarrow{\text{a.s.}} \frac{1}{4} \quad \text{and} \quad \sqrt{N}(\mathcal{O}_N - \mathbb{E}\mathcal{O}_N) \xrightarrow{d} \mathcal{N}\left(0, \frac{19}{240}\right) \quad \text{as } N \rightarrow \infty.$$

(ii) *If  $\frac{\lambda}{\mu} \leq \frac{1}{1+p}$ , then  $\mathcal{O}_N = 0$  a.s. for all  $N$ .*

In one dimension, additional explicit results can be obtained. In fact, for  $d = 1$  and  $k = 1$ , we obtain the exact expression for  $\mathbb{E}\mathcal{O}_N$  for all  $N \geq 1$ . Specifically, for the  $(1, p)$ -NNS strategy, we find that a quarter of the servers get overloaded which is stated in the following theorem.

**Theorem 2.2.** *Consider  $N \geq 4$  servers distributed uniformly in  $[0, 1]$  with exogenous arrivals of rate  $\lambda$  following the  $(1, p)$ -NNS strategy and having a service rate  $\mu$ . If  $\frac{1}{1+p} < \frac{\lambda}{\mu} \leq 1$ , then the expected fraction of overloaded servers is  $\mathbb{E}\mathcal{O}_N = \frac{1}{4}$ .*

In the process of proving this theorem, we also obtain the expected fraction of servers with all possible loads. This is detailed in Section 5.

Often, EV users tend to have a pre-determined destination while joining a highway and choose either to go left (with probability  $\ell$ ) or right (with probability  $r$ ) and proceed to the nearest server in that direction (or stay in the same queue with probability  $1 - r - \ell$ ). Customers at the left (right) boundary, choose the closest neighbour on the right (resp., left) with probability  $r$  (resp.,  $\ell$ ) or stay in the same queue with the remaining probability. For such a *left-right nearest neighbour shift* ( $\ell r$ -NNS) strategy, the following theorem asserts that there are no overloaded servers.

**Theorem 2.3.** *For  $N$  servers distributed uniformly in  $[0, 1]$  with stationary exogenous arrivals of intensity  $\lambda \leq \mu$  following the  $\ell r$ -NNS strategy with parameters  $r$  and  $\ell$ , the fraction of overloaded servers equals  $\mathcal{O}_N = 0$  a.s.*

## 3. Nearest neighbour graphs

In this section, we provide some background knowledge required in order to prove the main results. The proof of **Theorem 2.2** relies on the distribution of order statistics which is detailed in Section 3.1. Section 3.2 provides some results on  $k$ -NN graphs that are required to prove **Theorem 2.1**.

### 3.1. Order statistics

The primary tool that we will use in the 1D case are the order statistics of the server locations which we define below. Let  $X_1, \dots, X_N$  be iid random variables from a distribution  $F(\cdot)$  with density  $f(\cdot)$  supported on  $[0, 1]$ . When arranged in the order of magnitude and then written as  $X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(N)}$ , the random variable  $X_{(i)}$  is called the  $i$ th *order statistic*, and together they are referred to as the *order statistics* of  $X_1, \dots, X_N$ .

The following are some facts about order statistics that will be used in our analysis (see e.g., [13,14]). These are stated for  $f(x) = \mathbf{1}\{x \in [0, 1]\}$  and  $F(x) = \int_0^x f(x)dx$  since in our case the servers are distributed uniformly in  $[0, 1]$ .

- **Fact 1:** Conditioned on  $X_{(i)} = a$ , the random variables  $X_{(1)}, \dots, X_{(i-1)}$  are the order statistics of

$$Y_1, \dots, Y_{i-1} \stackrel{\text{iid}}{\sim} \text{Unif}([0, a]), \tag{2}$$

and  $X_{(i+1)}, \dots, X_{(N)}$  are the order statistics of

$$Y_{i+1}, \dots, Y_N \stackrel{\text{iid}}{\sim} \text{Unif}([a, 1]). \tag{3}$$

Moreover,  $X_{(1)}, \dots, X_{(i-1)}$  and  $X_{(i+1)}, \dots, X_{(N)}$  are conditionally independent given  $X_{(i)}$ .

- **Fact 2:** The joint density of the  $i$ th and the  $j$ th order statistic for  $i < j$  is given by

$$f_{i,j}(r, s) = \frac{N!}{(i-1)!(j-i-1)!(N-j)!} F(r)^{i-1} f(r) [F(s) - F(r)]^{j-i-1} f(s) [1 - F(s)]^{N-j}. \tag{4}$$

### 3.2. $k$ -NN graph

To characterize the  $(k, p)$ -NNS strategy for general dimension, we map the problem onto a graph and use results from stochastic geometry.

**Definition 3.1.** The  $k$ -NN graph  $G_{d,k}^N = (V, E)$  is a directed graph on the vertex set  $V = \{X_1, \dots, X_N\}$  with  $X_i \in [0, 1]^d$ , and the edge set is constructed in the following way: add a directed edge  $i \rightarrow j$ , if node  $j$  is one of the  $k$ -nearest neighbours of node  $i$ .

In this work, the vertex set  $V$  for the  $k$ -NN graph is taken to be  $N$  nodes distributed uniformly within  $[0, 1]^d$ . The *in-degree* of a vertex  $i$  is the number of incoming edges to  $i$ , i.e.,  $|\{j : j \rightarrow i\}|$  and the *out-degree* is the number of outgoing edges from  $i$  i.e.,  $|\{j : i \rightarrow j\}|$ . Some properties of the  $k$ -NN graph (see e.g., [12]) are listed below:

1. Every node  $i \in G_{d,k}^N$  has out-degree  $k$ .
2. Every node  $i \in G_{d,k}^N$  has in-degree at most  $\alpha_d k$ , where  $\alpha_d$  is the geometric constant defined in Section 2.2.

A directed graph  $H_1 = (V_1, E_1)$  is said to be a *subgraph* of  $H_2 = (V_2, E_2)$ , denoted  $H_1 \subseteq H_2$ , if  $V_1 \subseteq V_2$  and  $E_1 \subseteq E_2$ . Two directed graphs  $H_1$  and  $H_2$  are said to be *isomorphic* to each other if there exists a bijection  $f : V_1 \rightarrow V_2$  such that  $(i, j) \in E_1$  if and only if  $(f(i), f(j)) \in E_2$ . A directed graph is said to be *weakly connected* if the graph obtained on replacing the directed edges with undirected edges is connected. A subgraph that is weakly connected is referred to as a *component*.

Let  $I_D^N$  be the number of directed subgraphs of  $G_{d,k}^N$  that are isomorphic to a particular directed graph  $D$ . The convergence of these subgraph counts is characterized in [12] which is reproduced below. Here  $\mathcal{N}(\mu, \sigma^2)$  denotes a normal distribution with mean  $\mu$  and variance  $\sigma^2$ .

**Theorem 3.1** ([12, Theorem 2.1]). For any weakly connected directed graphs  $D_1, \dots, D_m$  and any real numbers  $a_1, \dots, a_m$  there exist constants  $\xi$  and  $\sigma$  such that  $I^N = a_1 I_{D_1}^N + a_2 I_{D_2}^N + \dots + a_m I_{D_m}^N$  satisfies

$$\frac{I^N}{N} \xrightarrow{\text{a.s.}} \xi \quad \text{and} \quad \frac{I^N - \mathbb{E}[I^N]}{\sqrt{N}} \xrightarrow{d} \mathcal{N}(0, \sigma^2) \quad \text{as } N \rightarrow \infty.$$

The subgraph of our interest is the directed star graph  $K_i = (V_i, E_i)$  with  $V_i = \{1, 2, \dots, i+1\}$  and  $E_i = \{(1, i+1), (2, i+1), \dots, (i, i+1)\}$  as shown in Fig. 1. The directed star graph captures the in-degree of a node which is key to our analysis. Let  $Q_{d,k,j}^N$  denote the number of nodes whose in-degree is  $j$  in  $G_{d,k}^N$ . The following proposition relates the quantities  $Q_{d,k,j}^N$  and  $I_{K_i}^N$ . We include a proof of it here since it was omitted in [12].

**Proposition 3.2.** For  $N > \alpha_d k + 1$ ,

$$Q_{d,k,j}^N = \sum_{i=j}^{\alpha_d k} (-1)^{i-j} \binom{i}{j} I_{K_i}^N. \tag{5}$$

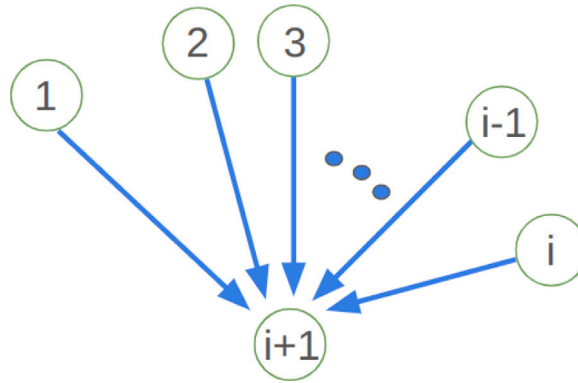


Fig. 1. The star graph  $K_i$  formed by a node with in-degree  $i$ .

**Proof.** Since  $K_i \subseteq K_{i+1}$ , the number of copies of  $K_i$  in  $G_{d,k}^N$  can be written as

$$I_{K_i}^N = Q_{d,k,i}^N + \binom{i+1}{i} Q_{d,k,i+1}^N + \dots + \binom{\alpha_d k}{i} Q_{d,k,\alpha_d k}^N.$$

Stacking these equations into a matrix form for  $1 \leq i \leq \alpha_d k$ , we obtain

$$\begin{bmatrix} I_{K_1}^N \\ I_{K_2}^N \\ \vdots \\ I_{K_{\alpha_d k-1}}^N \\ I_{K_{\alpha_d k}}^N \end{bmatrix} = \begin{bmatrix} 1 & \binom{2}{1} & \dots & \binom{\alpha_d k-1}{1} & \binom{\alpha_d k}{1} \\ 0 & 1 & \dots & \binom{\alpha_d k-1}{2} & \binom{\alpha_d k}{2} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & \binom{\alpha_d k}{\alpha_d k-1} \\ 0 & 0 & \dots & 0 & 1 \end{bmatrix} \begin{bmatrix} Q_{d,k,1}^N \\ Q_{d,k,2}^N \\ \vdots \\ Q_{d,k,\alpha_d k-1}^N \\ Q_{d,k,\alpha_d k}^N \end{bmatrix}. \tag{6}$$

The system of linear equations can be solved for the variables  $Q_{d,k,i}^N$  to obtain the statement of the proposition.  $\square$

As a corollary of Proposition 3.2 and Theorem 3.1, we obtain the convergence of the random variables  $Q_{d,k,j}^N$  as stated below.

**Corollary 3.1.** For  $0 \leq j \leq \alpha_d k$ , there exist constants  $q_{d,k,j}$  and  $\sigma_{d,k,j}$  such that as  $N \rightarrow \infty$ ,

$$\frac{Q_{d,k,j}^N}{N} \xrightarrow{\text{a.s.}} q_{d,k,j} \quad \text{and} \quad \frac{Q_{d,k,j}^N - \mathbb{E}[Q_{d,k,j}^N]}{\sqrt{N}} \xrightarrow{d} \mathcal{N}(0, \sigma_{d,k,j}^2).$$

The value of the constant  $q_{d,k,j}$  is hard to compute for general  $k$ . For the case of  $k = 1$ , [15] shows that

$$q_{d,1,j} = \frac{1}{j!} \sum_{i=0}^{\alpha_d-j} \frac{(-1)^i}{i!} C_{i+j}^d, \tag{7}$$

where

$$C_k^d = \int_{A_k^d} e^{-\text{vol}(B(u_1, \|u_1\|) \cup \dots \cup B(u_k, \|u_k\|))} du_1 \dots du_k,$$

with  $B(u, r)$  denoting the ball of radius  $r$  centred at  $u$ ,

$$A_k^d = \left\{ (u_1, \dots, u_k) \in (\mathbb{R}^d)^k : \|u_i\| < \|u_j - u_i\| \text{ for all } i \neq j \right\},$$

and  $\text{vol}(B)$  denoting the Lebesgue measure of set  $B$ .

#### 4. Proof for general dimension

In this section we present the proof of Theorem 2.1 characterizing the fraction of overloaded service stations for the  $(k, p)$ -NNS strategy in  $[0, 1]^d$ .

##### Proof of Theorem 2.1.

Assume that  $\lambda < \mu$  and  $0 < p \leq 1$ . We denote by  $d_{\text{in}}(i)$  and  $d_{\text{out}}(i)$  the in-degree and the out-degree of node  $i$  in the  $k$ -NN graph of the servers. Because the server locations are independently and uniformly distributed in  $[0, 1]^d$ , we see that all inter-server distances are unique with probability one. As a consequence,  $0 \leq d_{\text{in}}(i) \leq \alpha_d k$  and  $d_{\text{out}}(i) = k$  for all  $i$  almost surely, as explained in Section 3.2. Recall that customers arriving at location  $i$  join queue  $i$  with probability  $1 - p$ , and customers arriving at location  $j$  with  $i \in \mathcal{N}_k(j)$

join queue  $i$  with probability  $\frac{p}{k}$ . Because the arrival processes are stationary and mutually independent with intensity  $\lambda$ , it follows that the net rate of customers joining queue  $i$  equals  $\lambda_{\text{eff}}(i) = \lambda(1 - p) + \lambda d_{\text{in}}(i) \frac{p}{k}$ . We conclude queue  $i$  is overloaded if and only if

$$\lambda(1 - p) + \lambda d_{\text{in}}(i) \frac{p}{k} > \mu,$$

or equivalently,  $d_{\text{in}}(i) > \theta$  where  $\theta = k + \frac{k}{p} \left( \frac{\mu}{\lambda} - 1 \right)$ . As a consequence, the fraction of overloaded queues in the system can be written as

$$\mathcal{O}_N = \frac{1}{N} \sum_{n: n > \theta} Q_{d,k,n}^N,$$

where  $Q_{d,k,n}^N$  equals the number of servers with in-degree  $n$ . Because  $d_{\text{in}}(i) \leq \alpha_d k$ , we note that  $Q_{d,k,n}^N = 0$  for  $n > \alpha_d k$  almost surely. Therefore, with probability one,

$$\mathcal{O}_N = \begin{cases} \frac{1}{N} \sum_{n=\lfloor \theta \rfloor + 1}^{\alpha_d k} Q_{d,k,n}^N, & \theta < \alpha_d k, \\ 0, & \theta \geq \alpha_d k. \end{cases} \tag{8}$$

We also note that  $\theta \geq \alpha_d k$  if and only if  $\frac{\lambda}{\mu} \leq \frac{1}{1 - p + \alpha_d p}$ . Hence  $\mathcal{O}_N = 0$  a.s. for  $\frac{\lambda}{\mu} \leq \frac{1}{1 - p + \alpha_d p}$ .

Let us now assume that  $\frac{1}{1 - p + \alpha_d p} < \frac{\lambda}{\mu} \leq 1$ . Then  $\theta < \alpha_d k$ . Corollary 3.1 then implies that

$$\mathcal{O}_N \xrightarrow{\text{a.s.}} \sum_{n=\lfloor \theta \rfloor + 1}^{\alpha_d k} q_{d,k,n} \quad \text{as } N \rightarrow \infty.$$

Moreover, Proposition 3.2 implies that

$$\sum_{n=\lfloor \theta \rfloor + 1}^{\alpha_d k} Q_{d,k,n}^N = \sum_{n=\lfloor \theta \rfloor + 1}^{\alpha_d k} \sum_{m=n}^{\alpha_d k} (-1)^{m-n} \binom{m}{n} I_{K_m}^N = \sum_{m=\lfloor \theta \rfloor + 1}^{\alpha_d k} \left( \sum_{n=\lfloor \theta \rfloor + 1}^m (-1)^{m-n} \binom{m}{n} \right) I_{K_m}^N.$$

We conclude that the fraction of overloaded servers can be written as

$$\mathcal{O}_N = \frac{1}{N} \sum_{m=1}^{\alpha_d k} a_m I_{K_m}^N,$$

where

$$a_m = \begin{cases} \sum_{n=\lfloor \theta \rfloor + 1}^m (-1)^{m-n} \binom{m}{n}, & m \geq \lfloor \theta \rfloor + 1, \\ 0, & m \leq \lfloor \theta \rfloor. \end{cases}$$

Theorem 3.1 then asserts that  $\sqrt{N}(\mathcal{O}_N - \mathbb{E}\mathcal{O}_N) \xrightarrow{d} \mathcal{N}(0, \sigma^2)$  for some constant  $\sigma$ .  $\square$

### 5. Proofs for one dimension

The case of dimension  $d = 1$  is interesting in its own right since it can be motivated by numerous applications. With reference to the examples discussed in the introduction, EV charging stations located on a highway and checkout counters at supermarkets correspond to queues on a 1D space.

#### 5.1. In-degree frequencies in the one-dimensional 1-NN graph

In this section we analyse the in-degree frequencies in the 1-NN graph  $G_{1,1}^N$  in dimension  $d = 1$ , which will be useful in characterizing the overloaded servers in one dimension. In  $G_{1,1}^N$ , a particular vertex can have in-degrees  $\{0, 1, 2\}$ . This is because a node  $i$  could be the nearest neighbour to at most two vertices on either side of  $i$ . Recall that  $Q_{d,k,n}^N$  equals the number of nodes with in-degree  $n$  in  $G_{d,k}^N$ . The following lemma describes the expected in-degree frequencies in  $G_{1,1}^N$ .

**Proposition 5.1.** For any  $N \geq 4$ , the expected proportions of nodes with in-degree  $s = 0, 1, 2$  in the 1-NN graph on  $[0, 1]$  are given by

$$\mathbb{E} \left[ \frac{Q_{1,1,0}^N}{N} \right] = \frac{1}{4}, \quad \mathbb{E} \left[ \frac{Q_{1,1,1}^N}{N} \right] = \frac{1}{2}, \quad \text{and} \quad \mathbb{E} \left[ \frac{Q_{1,1,2}^N}{N} \right] = \frac{1}{4}.$$

**Proof.** Let  $X_{(1)}, \dots, X_{(N)}$  denote the order statistics of the locations of the  $N$  servers. In the following discussion, node  $i$  will refer to the  $i$ th order statistic. For  $3 \leq i \leq N - 2$ , node  $i$  has no incoming edges in  $G_{1,1}^N$  if and only if both the events  $E_i^+ := \{X_{(i+2)} - X_{(i+1)} < X_{(i+1)} - X_{(i)}\}$  and  $E_i^- := \{X_{(i-1)} - X_{(i-2)} < X_{(i)} - X_{(i-1)}\}$  occur simultaneously. We first compute the probability of the event  $E_i^+ \cap E_i^-$ .

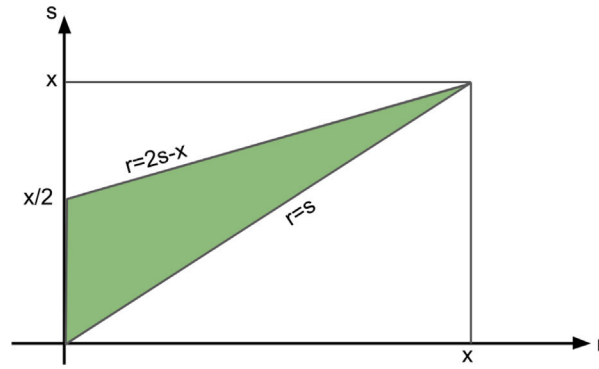


Fig. 2. Region of integration.

From Fact 1 of Section 3.1, conditioned on the value of  $X_{(i)} = x$ , the events  $E_i^+$  and  $E_i^-$  are independent. Thus

$$\mathbb{P}(E_i^+ \cap E_i^-) = \int_0^1 \mathbb{P}(X_{(i+2)} > 2X_{(i+1)} - x | X_{(i)} = x) \mathbb{P}(X_{(i-2)} > 2X_{(i-1)} - x | X_{(i)} = x) f_{X_{(i)}}(x) dx, \tag{9}$$

where  $f_{X_{(i)}}(\cdot)$  is the density of the  $i$ th order statistic. We will now evaluate each of the probabilities in the above expression using (2), (3) and (4). For this, we first write down the joint distribution of the  $j$ th and the  $(j + 1)$ -th order statistic using (4) to be

$$f_{j,j+1}(r, s) = N! \frac{F(r)^{j-1} f(r) f(s) (1 - F(s))^{N-j-1}}{(j - 1)!(N - j - 1)!}. \tag{10}$$

Using (2), (3) with (10) the required integral can be computed as follows: (see Fig. 2)

$$\begin{aligned} \mathbb{P}(X_{(i-2)} > 2X_{(i-1)} - x | X_{(i)} = x) &= \int_0^x \int_0^s \frac{(i-1)!}{(i-3)!} \left(\frac{r^{i-3}}{x^{i-1}}\right) dr ds - \int_{\frac{x}{2}}^x \int_0^{2s-x} \frac{(i-1)!}{(i-3)!} \left(\frac{r^{i-3}}{x^{i-1}}\right) dr ds \\ &= \int_0^x \frac{(i-1)(i-2)}{x^{i-1}} \left(\frac{s^{i-2}}{i-2}\right) ds - \int_{\frac{x}{2}}^x \frac{(i-1)(i-2)}{x^{i-1}} \left(\frac{(2s-x)^{i-2}}{i-2}\right) ds \\ &= \frac{i-1}{x^{i-1}} \int_0^x s^{i-2} ds - \frac{i-1}{x^{i-1}} \int_{\frac{x}{2}}^x (2s-x)^{i-2} ds \\ &= 1 - \frac{i-1}{x^{i-1}} \int_{\frac{x}{2}}^x (2s-x)^{i-2} ds = \frac{1}{2}. \end{aligned}$$

and similarly  $\mathbb{P}(X_{(i+2)} > 2X_{(i+1)} - x | X_{(i)} = x) = \frac{1}{2}$ . Substituting this in (9), we obtain

$$\mathbb{P}(E_i^+ \cap E_i^-) = \int_0^1 \frac{f_{X_{(i)}}(x)}{4} dx = \frac{1}{4}.$$

Nodes 1 and  $N$  have no incoming edges if the event  $E_1^+$  and  $E_N^-$  are true respectively. Moreover, nodes 2 and  $N - 1$  always have an incoming edge from nodes 1 and  $N$  respectively. Thus,

$$\mathbb{E}\left[\frac{Q_{1,1,0}^N}{N}\right] = \frac{1}{N} \mathbb{E}\left[\mathbf{1}\{E_1^+\} + \mathbf{1}\{E_N^-\} + \sum_{i=3}^{N-2} \mathbf{1}\{E_i^+ \cap E_i^-\}\right] = \frac{1}{N} \left[2 \cdot \frac{1}{2} + (N-4) \cdot \frac{1}{4}\right].$$

For  $N \geq 4$ , we obtain  $\mathbb{E}\left[\frac{Q_{1,1,0}^N}{N}\right] = \frac{1}{4}$ . Similar computations for  $\mathbb{E}\left[\frac{Q_{1,1,1}^N}{N}\right]$  and  $\mathbb{E}\left[\frac{Q_{1,1,2}^N}{N}\right]$  prove the proposition.  $\square$

In fact, a stronger statement to Proposition 5.1 holds for  $Q_{1,1,0}^N$  and  $Q_{1,1,2}^N$  which is stated below.

**Proposition 5.2.** *The number of vertices with in-degree 0 and 2 in the 1-NN graph on  $[0, 1]$  are the same, that is,  $Q_{1,1,0}^N = Q_{1,1,2}^N$  a.s.*

**Proof.** A crucial observation for the 1-NN graph is that the only possible cycles are 2-cycles comprising of two nodes that are mutual nearest neighbours (see [16]). If all the components of the 1-NN graph  $G_1^N$  in one dimension are 2-cycles as shown in Fig. 3(a), then  $Q_{1,1,0}^N = Q_{1,1,2}^N = 0$ . On the other hand, every component of more than two vertices contains exactly one 2-cycle. Such components can be of two types: starting at a node of in-degree 0 and culminating in the 2-cycle as shown in Fig. 3(b) or starting and ending at a node of in-degree 0 with a 2-cycle in between as shown in Fig. 3(c). In the former case, since the starting vertex has in-degree 0 and one of the vertices in the 2-cycle associated with the component has in-degree 2, every component of size greater than 2 can be associated with such a unique pair of vertices with in-degrees of 0 and 2. Likewise, in the latter case, the component has two node



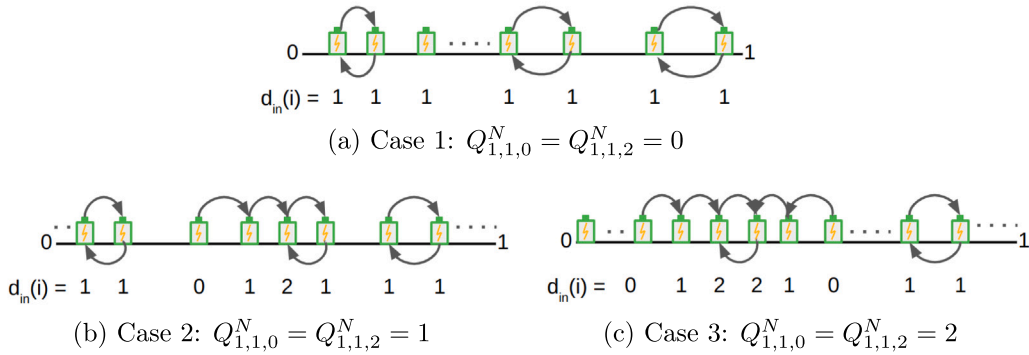


Fig. 3. Illustration of  $Q_{1,1,0}^N = Q_{1,1,2}^N$ . An arrow  $i \rightarrow j$  indicates  $j$  is the nearest neighbour of  $i$ .

in the 2-cycle which have in-degree 2 and the two nodes at the extremes of the component have in-degree 0. Thus, in all the cases we have that  $Q_{1,1,0}^N = Q_{1,1,2}^N$  almost surely which proves the claim.  $\square$

5.2.  $(1, p)$ -NNS strategy

In this section, we characterize the overloaded servers for the  $(1, p)$ -NNS strategy in one-dimension by proving Theorem 2.2.

**Proof of Corollary 2.1.** Let  $d = 1$  and  $k = 1$ . Then  $\alpha_1 = 2$ , and by Theorem 2.1 we find that  $\mathcal{O}_N = 0$  a.s. when  $\frac{\lambda}{\mu} \leq \frac{1}{1+p}$ .

Assume next that  $\frac{1}{1+p} < \frac{\lambda}{\mu} \leq 1$ . Then the constant  $\theta$  in Theorem 2.1 equals  $\theta = 1 + \frac{1}{p} \left( \frac{\mu}{\lambda} - 1 \right)$  and is bounded by  $1 \leq \theta < 2$ . Theorem 2.1 then implies that

$$\mathcal{O}_N \xrightarrow{\text{a.s.}} q_{1,1,2} \quad \text{and} \quad \sqrt{N}(\mathcal{O}_N - \mathbb{E}\mathcal{O}_N) \xrightarrow{d} \mathcal{N}(0, \sigma_{1,1}^2).$$

The claim follows by plugging in the values of the constants  $q_{1,1,2} = \frac{1}{4}$  and  $\sigma_{1,1}^2 = \frac{19}{240}$  obtained in [17].  $\square$

**Proof of Theorem 2.2.** Because  $\frac{1}{1+p} < \frac{\lambda}{\mu} \leq 1$ , we see that the constant  $\theta$  in (8) is bounded by  $1 \leq \theta < 2$ . Because  $\alpha_1 = 2$  for  $d = 1$ , we see that (8) takes the form  $\mathcal{O}_N = \frac{1}{N} Q_{1,1,2}^N$ . Proposition 5.1 now yields the result  $\mathbb{E}[\mathcal{O}_N] = \frac{1}{4}$ .  $\square$

**Remark 5.1.** While Theorem 2.2 characterizes the fraction of overloaded servers for  $N \geq 4$ , it can be inferred exhaustively that  $\mathbb{E}\mathcal{O}_2 = 0$  and  $\mathbb{E}\mathcal{O}_3 = \frac{1}{3}$ .

5.3. Left-right strategy

**Proof of Theorem 2.3.** Consider a service station  $i$  that is not located at the boundary of  $[0, 1]$ . Customers arriving at queue  $i$  join queue  $i$  with probability  $1 - \ell - r$ . Customers arriving to the nearest queue to the left (resp. right) of  $i$  join queue  $i$  with probability  $\ell$  (resp. right). Therefore, the net rate of customers joining queue  $i$  equals

$$\lambda_{\text{eff}}(i) = \lambda(1 - \ell - r) + \lambda\ell + \lambda r = \lambda.$$

A similar computation shows that  $\lambda_{\text{eff}}(i) = \lambda$  also for the service stations located at the boundary of  $[0, 1]$ . Therefore,  $\mathcal{O}_N = 0$  a.s.  $\square$

6. Numerical results

In this section, we present simulation experiments complementing our theoretical results. We restrict to the case of  $k = 1$  since the constants in our results (or their approximation using Monte Carlo simulations) are known only for the  $(1, p)$ -NNS strategy.

Consider  $N = 1000$  servers deployed uniformly in  $[0, 1]$ , each of which is associated with an independent arrival process of intensity  $\lambda = 1$ . The arrival process is simulated for  $t = 1000$  time units with a shift probability  $p$ . The probability parameter  $p$  is chosen from the set  $\{1, 0.75, 0.5, 0.25\}$ . Fig. 4(a) plots the histogram of the observed number of overloaded servers ( $Q_{1,1,2}^N$ ) and the servers with no change in load ( $Q_{1,1,1}^N$ ) over 1000 instantiations of the server locations in the one dimensional case. The number of overloaded servers is concentrated well around the mean of  $N/4 = 250$  irrespective of the probability  $p$  corroborating with Theorem 2.2.



**Table 1**  
 Values of the asymptotic  $k$ -NN in-degree distribution  $q_{d,k,j}$  for  $d = 2$  and  $k = 1$ . Recall that  $q_{2,1,j} = 0$  for  $j \geq 6$ .

$j$	$q_j(2, 1)$
0	$2.84 \times 10^{-1}$
1	$4.63 \times 10^{-1}$
2	$2.22 \times 10^{-1}$
3	$3.04 \times 10^{-2}$
4	$6.56 \times 10^{-4}$
5	$1.90 \times 10^{-7}$

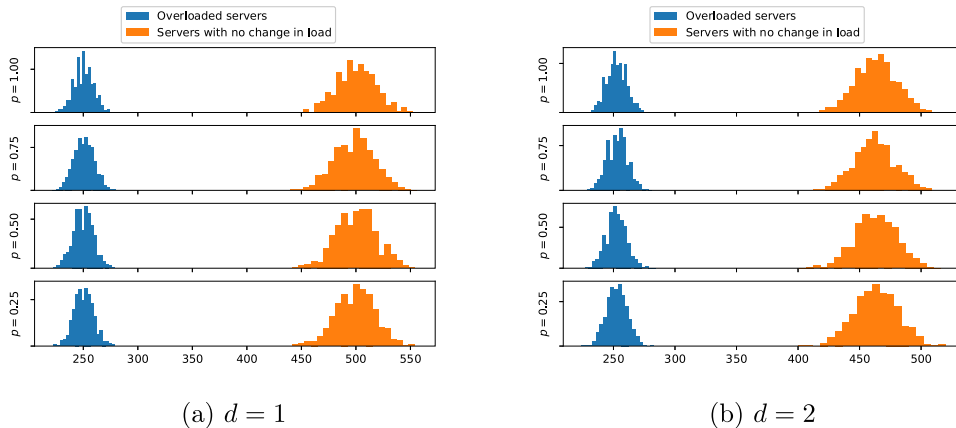


Fig. 4. Histogram of the number of overloaded servers and the servers with no change in load for different probability values of the  $(1, p)$ -NNS strategy in (a) one dimension and (b) two dimension.

Fig. 4(b) shows a similar histogram of the expected fraction of overloaded servers when the servers are distributed in  $[0, 1]^2$ . The values of the constants  $q_{2,1,j}$  for  $j = 0, 1, \dots, 5$  have been computed using Monte Carlo simulations in [18] which is reproduced in Table 1.

When the servers are distributed in 2D and customers follow the  $(1, p)$ -NNS strategy, Theorem 2.1 states that the fraction of overloaded servers converges to  $\sum_{j=2}^5 q_{2,1,j}$  almost surely as  $N \rightarrow \infty$ . Using the values from Table 1, we obtain  $\sum_{j=2}^5 q_{2,1,j} \approx 0.252$  which is where the blue peak is located in Fig. 4(b). Further supporting the result is the peak observed of the fraction of servers with no change in load which is equal to  $q_{2,1,1} = 0.463$ .

## 7. Conclusions and future work

In this paper, we considered multiple servers on a  $d$ -dimensional Euclidean space, where arriving customers follow a probabilistic policy to either remain in the queue or to join a nearest neighbour. In this setting, we provided distributional properties of the fraction of overloaded servers in the stationary regime.

Numerous questions remain yet to be answered in this setting, some of which are discussed below.

### 7.1. Spatial distribution of overloaded servers

In the long run, overloaded servers require higher maintenance and allocating more capacity. This necessitates a characterization of their distribution in space. Figs. 5 and 6 show the distribution of the overloaded servers in space in one and two dimensions respectively. It is reasonable to expect that overloaded servers do not occur close to one another. For example, three adjacent servers in dimension one cannot be overloaded since their total in-degree is upper bounded by 5 in  $G_{1,1}^N$ . This suggests that the spatial point pattern of overloaded servers resembles more a Matérn point process than a standard Poisson point process [19]. The spatial distribution is of particular importance in the context of electric vehicles since it gives the distribution of locations of charging stations requiring maintenance.

### 7.2. State-dependent shift probabilities and adaptive mobility strategies

Exploring state-dependent strategies, where shift probabilities change with transient parameters such as the queue length, is an important research direction. On a related note, [20,21] investigate the steady-state behaviour and system stability when the arrivals are distributed on the circle and a myopic server operates according to a greedy policy serving the nearest arrivals. In our

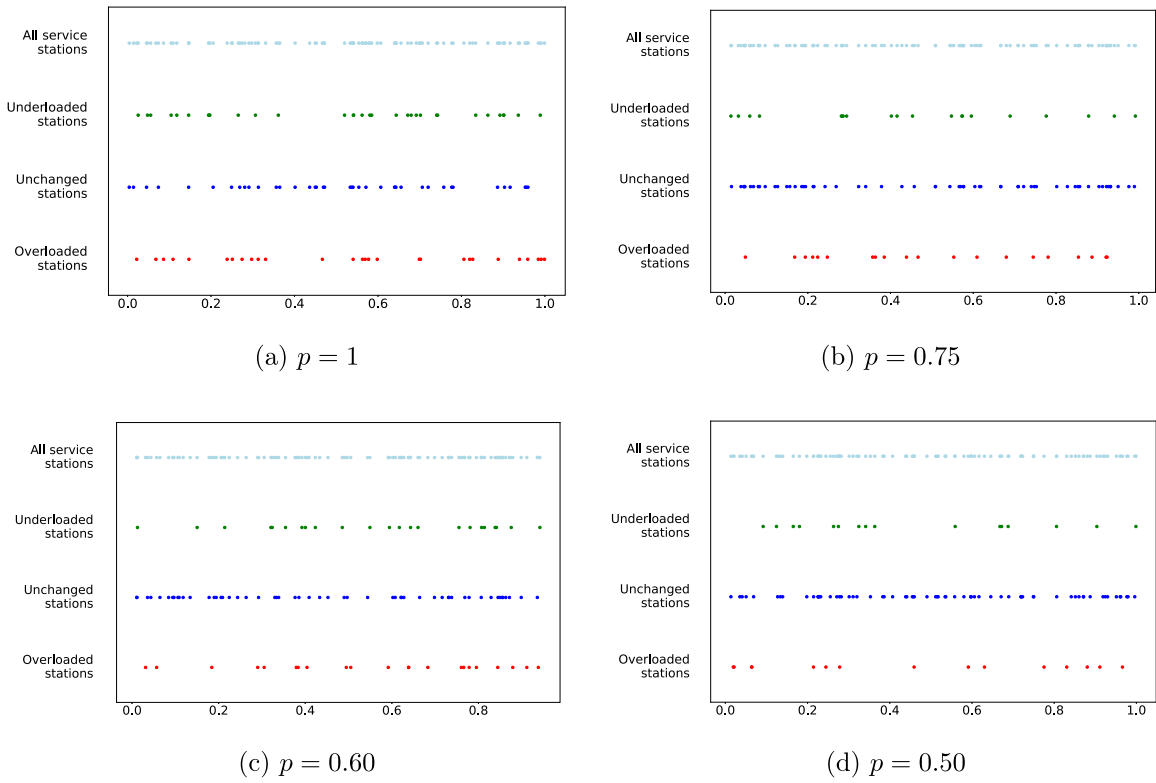


Fig. 5. Spatial distribution of 100 service stations , underloaded stations, stations with no change in load and overloaded stations in  $[0, 1]$  for different probability values.

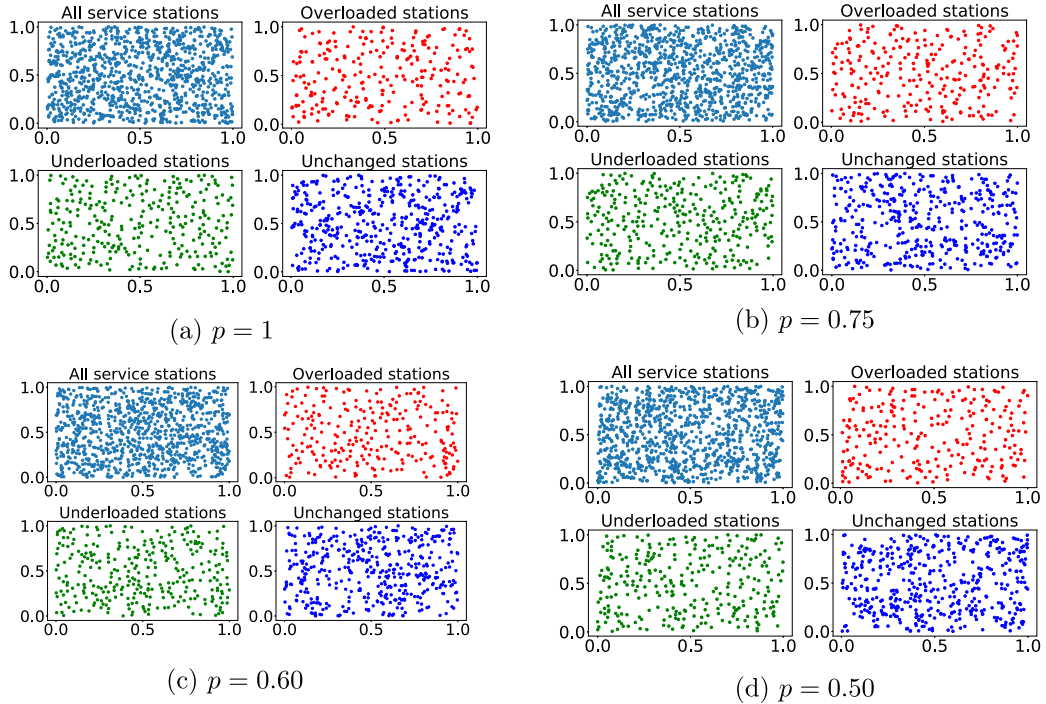


Fig. 6. Spatial distribution of 1000 service stations (top-left), overloaded stations (top-right), underloaded stations (bottom-left) and stations with unchanged load (bottom-right) in  $[0, 1]^2$  for different probability values.

case, however, we have multiple servers. When customers change queues based on the queue length they observe when they arrive, characterization of the steady-state behaviour and stability are compelling future directions. As a first step in this direction, one could study an adaptive  $(k, p)$ -NNS strategy where customers join empty queues or, with probability  $p$ , join a non-empty queue; otherwise, they select one of the  $k$  nearest neighbouring queues. Such strategies likely improve server utilization and reduce idling. Techniques similar to those in [22] might be feasible for showing that overloaded queues under adaptive strategies would also be overloaded under non-adaptive ones, potentially using non-adaptive results as upper bounds for adaptive cases. Understanding these dynamics under state-dependent and adaptive strategies remains an interesting area for future research.

### 7.3. Impact of ambient space

How does the ambient space impact  $\mathcal{O}_N$ ? If factors from the Euclidean space such as traversal times to other queues are considered, several classical queueing problems (delay, wait times etc.) can be formulated which might be interesting in their own right, see for example [23] and references therein.

### CRedit authorship contribution statement

**B.R. Vinay Kumar:** Writing – review & editing, Writing – original draft, Validation, Methodology, Investigation, Conceptualization. **Lasse Leskelä:** Writing – review & editing, Writing – original draft, Validation, Supervision, Methodology, Investigation, Formal analysis.

### Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: B. R. Vinay Kumar reports financial support was provided by French government through RISE Academy of UCA JEDI Investments in Future project managed by the National Research Agency (ANR) with reference number ANR15 IDEX 0001. B. R. Vinay Kumar reports financial support was provided by Aalto Science Institute (ASCI) visiting researcher fellowship, Aalto University, Finland and the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement, Grant Agreement No 101034253. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

### Acknowledgments

The authors would like to thank the anonymous reviewers and the guest editor for helping to improve the manuscript significantly. The authors would also like to thank Sem Borst and Jamol Pender for pointing out relevant references for the present work.

### References

- [1] G. Dong, J. Ma, R. Wei, J. Haycox, Electric vehicle charging point placement optimisation by exploiting spatial statistics and maximal coverage location models, *Transp. Res. D: Transp. Environ.* 67 (2019) 77–88.
- [2] C. Ren, Y. Hou, A novel evaluation method of electric vehicles charging network based on stochastic geometry, 2020, p. 468.
- [3] R. Atat, M. Ismail, E. Serpedin, Stochastic geometry planning of electric vehicles charging stations, in: *Proc. International Conference on Acoustics, Speech and Signal Processing, ICASSP*, 2020, pp. 3062–3066.
- [4] E. Altman, H. Levy, Queueing in space, *Adv. in Appl. Probab.* 26 (4) (1994) 1095–1116.
- [5] V. Kavitha, E. Altman, Queueing in space: Design of message ferry routes in static ad hoc networks, in: *2009 21st International Teletraffic Congress*, 2009, pp. 1–8.
- [6] G.D. Çelik, E. Modiano, Dynamic vehicle routing for data gathering in wireless networks, in: *49th IEEE Conference on Decision and Control, CDC*, 2010, pp. 2372–2377.
- [7] E. Koenigsberg, On jockeying in queues, *Manage. Sci.* (1966).
- [8] S.A. Dudin, O.S. Dudina, O.I. Kostyukova, Analysis of a queueing system with possibility of waiting customers jockeying between two groups of servers, *Mathematics* 11 (6) (2023) 1475.
- [9] B. Lin, Y. Lin, R. Bhatnagar, Optimal policy for controlling two-server queueing systems with jockeying, *J. Syst. Eng. Electron.* 33 (1) (2022) 144–155.
- [10] K.W. Stacey, D.P. Kroese, Greedy servers on a torus, in: *Proceedings of the 2011 Winter Simulation Conference, WSC*, 2011, pp. 369–380.
- [11] N.K. Panigrahy, T. Vasantam, P. Basu, D. Towsley, A. Swami, K.K. Leung, On the analysis and evaluation of proximity-based load-balancing policies, *ACM Trans. Model. Perform. Evaluat. Comput. Syst.* 7 (2–4) (2022) 1–27.
- [12] S. Bahadır, E. Ceyhan, On the number of weakly connected subdigraphs in random kNN digraphs, *Discrete Comput. Geom.* 65 (1) (2021) 116–142.
- [13] H.A. David, H.N. Nagaraja, *Order statistics*, John Wiley & Sons, 2004.
- [14] S. Ross, *A first course in probability*, Pearson, 2010.
- [15] C.M. Newman, Y. Rinott, A. Tversky, Nearest neighbors and Voronoi regions in certain point processes, *Adv. in Appl. Probab.* 15 (4) (1983) 726–751.
- [16] D. Eppstein, M.S. Paterson, F.F. Yao, On nearest-neighbor graphs, *Discrete Comput. Geom.* 17 (1997) 263–282.
- [17] S. Bahadır, E. Ceyhan, On the number of reflexive and shared nearest neighbor pairs in one-dimensional uniform data, 2016, arXiv preprint [arXiv:1605.01940](https://arxiv.org/abs/1605.01940).
- [18] J. Cuzick, R. Edwards, Spatial clustering for inhomogeneous populations, *J. R. Stat. Soc. Ser. B Stat. Methodol.* 52 (1) (1990) 73–96.
- [19] M. Karonen, L. Leskelä, Hard-core thinnings of germ–grain models with power-law grain sizes, *Adv. in Appl. Probab.* 45 (3) (2013) 595–625.
- [20] L. Leskelä, F. Unger, Stability of a spatial polling system with greedy myopic service, *Ann. Oper. Res.* 198 (1) (2012) 165–183.
- [21] L.T. Rolla, V. Sidoravicius, Stability of the greedy algorithm on the circle, *Comm. Pure Appl. Math.* 70 (10) (2017) 1961–1986.
- [22] S. Borst, M. Jonckheere, L. Leskelä, Stability of parallel queueing systems with coupled service rates, *Discrete Event Dyn. Syst.* 18 (4) (2008) 447–472.
- [23] J. Pender, k-Nearest neighbor queues with delayed information, *Int. J. Bifurcation Chaos* 32 (12) (2022) 2250174.