On the $k$ Nearest-Neighbor Path Distance from the Typical Intersection in the Manhattan Poisson Line Cox Process

Konstantinos Koufos, Harpreet S. Dhillon, Mehrdad Dianati and Carl P. Dettmann

Abstract—In this paper we calculate the exact cumulative distribution function (CDF) of the path distance (L1 norm) between a randomly selected intersection and the $k$-th nearest node of the Cox point process driven by the Manhattan Poisson line process. The CDF is expressed as a sum over the integer partition function $p(k)$, which allows us to numerically evaluate the CDF in a simple manner. The distance distributions can be used to study the $k$-coverage of broadcast signals in intelligent transportation systems (ITS) transmitted from a road side unit (RSU) that is located at an intersection. They can also be insightful for network dimensioning in urban vehicle-to-everything (V2X) systems, because they can yield the exact distribution of network load within a cell, provided that the RSU is located at an intersection. Finally, they can find useful applications in other branches of science like spatial databases, emergency response planning, and districting. We corroborate the applicability of the distance distribution model using the map of an urban area.

Index Terms—Manhattan Poisson line Cox process, spatial databases, stochastic geometry, vehicular networks.

I. INTRODUCTION

The development of road networks is a key component of urban planning because it greatly affects commuting efficiency, districting, emergency response dispatching, and first-aid services, to name but a few. Since the recent advent of mobile broadband connectivity for pedestrians and the ongoing proliferation of connected vehicles through vehicle-to-everything (V2X) systems, the road network is also the setting for several location-based e-services [1]. Exemplar applications could be electric vehicles querying over the internet for the nearest charging stations, and/or pedestrians searching with their smartphones for the closest available taxis [2]–[5].

A. Modeling road networks

The simplest models for urban road networks utilize just a set of vertices and edges [6], [7]. The vertices may represent junctions, the start/end points of roadways, critical locations where the speed limit or the travel direction changes, etc. Naturally, two vertices are connected by an edge if there is a straight link between them, giving rise to the adjacency matrix of the road network. The graph representation can be enhanced by assigning weights to the edges, which might be proportional to the (average or minimum) travel time and fuel cost, along the road segment(s) that the edge represents. Algorithms exploring the graph have been also developed, e.g., the best-first search to identify the nearest neighbors from a vertex and the Dijkstra’s algorithm to find the shortest paths, i.e., the sequence of edges of minimal aggregate cost between two non-adjacent vertices [8]. We will also utilize Dijkstra’s algorithm with edge weights equal to the length of road segments while validating our distance distribution model with the map of an urban area.

Another line of research, particularly useful for emergency response planning, assumes that the edge weights are proportional to the length of the associated streets, and models random events along the edges. If these events represent points of emergency, the graph distance distributions can reveal the intrinsic properties of the response system we need to build to combat all emergencies effectively [9]. For instance, the distance distribution can be used to infer the number of ambulances, medical personnel, etc. we will have to deploy. While certainly important, the graph-based approaches apply to specific road networks. Even though different cities can share common road graph properties [10], the graph-based models provide limited abstraction. Besides, due to the high complexity of graph-based routines, it is often impossible to model the road network very precisely. Also, the graph representation cannot answer questions pertinent to network planning, e.g., what is the minimum required intensity of charging stations, so that two of them are within a driving distance of one kilometer from a randomly chosen intersection, with probability at least 90%? This paper aims to bridge this gap. We argue that the mathematical tools of stochastic geometry, see [11] for an introduction, widely and successfully utilized during the past 15 years in the performance evaluation of random wireless communication networks, see [12]–[18], can also be insightful for urban road planning.

Unlike the graph-based methodology, the stochastic geometry framework does not consider a specific road network. In its most tractable (and hence useful) form, only the intensity of streets is available (or can be estimated). In addition, we assume that: (i) the road layout has a relatively regular structure, hence, an Manhattan Poisson line process (MPLP) is a realistic model for it, and (ii) the locations of points of interest (POI) or facilities, e.g., gas stations, events triggering police action, etc. follow a homogeneous Poisson point process (PPP) along each street. Under these assumptions,
i.e., a Manhattan Poisson line Cox process (MPLCP) for the locations of POI, we will derive the path distance (L1 norm) distribution of the k nearest-neighbors (kNN) (or POI) from a randomly selected intersection. We have identified three potential applications, namely spatial database queries, districting, and urban vehicular ad hoc networks (VANETs), where the kNN path distance distributions could be of use. We elaborate on them next.

B. Motivation and prior art

In the kNN query, a spatial database returns the locations of the k nearest objects, in terms of network distance, to the query point (or agent) [2]. Consider, for instance, a driver looking for the k nearest hotels (static objects) in terms of travel time, or a pedestrian querying for the k nearest vacant cabs (mobile objects). The agent reports its location, and the database solves the query using, e.g., a graph representation for the road network [3, Fig. 2].

Even with static objects, the graph dynamically changes due to varying traffic conditions, and the computational complexity can quickly explode, especially with frequent queries from mobile agents [4]. The server must continuously track and update the locations of the k nearest objects for all agents. Because of that, neglecting the constraints imposed by the road network, and using just Minkowski distances to solve the kNN problem has not been abandoned, especially for group queries [5]. The study in [2] has pointed out that the Euclidean distance is a lower bound to the network distance and could be used to prune the search space in kNN queries. However, pruning based on a lower bound is not always effective. Therefore the calculation of the exact path distance distributions, as we will do in this paper, will be helpful.

The kNN distance distributions can also be used in the planning of dispatching policies for emergency response services and districting [19]. In balanced district design, the road network of a metropolitan area is partitioned into smaller units (territories) which contain about the same expected number of road accidents so that the workload is divided equally among police departments [19]. Given the size of the districts, the kNN distributions can be used to calculate the probability that a police department can successfully cover the k nearest emergencies under some probabilistic time constraint.

The ongoing rollout of 5G networks and the standardization activities for V2X communication, e.g., the transmission of cooperative awareness messages (CAMs) from the vehicles to the infrastructure [20], and the response of the latter with the collective perception message (CPM) [21] have motivated the development of spatial models tailored to vehicular networks. Despite the fact that urban streets have finite lengths, line processes have been extensively used in wireless communication research to gain analytical insights into the network’s performance [13]–[16]. The one- and two-dimensional PPPs are valid models for urban VANETs, in the high- and low-reliability regime respectively [22]. Otherwise, to draw valid conclusions about the network’s performance the road intersections must be explicitly modeled [16]. For motorway VANETs, on the contrary, the superposition of one-dimensional (1D) point processes is sufficient [23]–[25].

The studies in [13]–[15] have used a Poisson line process to capture the random orientation of streets, and stationary 1D PPPs to model the locations of vehicles per street. In the resulting Poisson line process \( \Phi_c \), the study in [13] has evaluated the distance distribution between an arbitrary point in the plane and the nearest point of \( \Phi_c \). This is essentially the serving distance distribution in a cellular vehicular network with nearest base station association, where the locations of base stations follow the two-dimensional PPP [14]. The study in [15] has derived the coverage probability for the typical receiver in a VANET and pointed out the conflicting effect of the intensities of the roads and vehicles. It has also solved for the distance distribution between the typical vehicle and the nearest vehicle of \( \Phi_c \) to model a vehicle-to-vehicle communication scenario. The study in [26] has used the MPLCP to model the locations of base stations in urban street microcells and calculated their distance distribution to the origin to study properties of the interference distribution. Finally, for some recent results on the reliability of inter-vehicle communication in urban streets of finite length, forming intersections and T-junctions, see [27].

Unfortunately, the above studies on the performance of VANETs have measured the Euclidean distances (L2 norm), even though the attenuation of wireless signals, especially in millimeter-wave frequencies, is better described by a street canyon model [28, Eq. (1)]. In this regard, the kNN Manhattan distance distributions will be useful in investigating the k-coverage of wireless signals, diffracted around buildings at road intersections as they propagate. We will use them to identify, e.g., how many vehicles within half a kilometer from an intersection can successfully receive broadcast safety messages with probability at least \( q \)%? Thus far, the kNN distributions have been identified for Poisson and binomial processes see [29]–[31], without considering the deployment constraints due to the road layout. For \( k = 1 \), more general convex geometries like the \( n \)-sided polygon have been also investigated [32].

C. Contributions

Let us denote by \( R_k \) the random variable (RV) of the path length (L1 norm) between a randomly chosen intersection of the MPLP and the \( k \)-th nearest neighbor of the MPLCP. The cumulative distribution function (CDF) of the length of the shortest path, \( R_1 \), has been recently calculated in [33]. The authors have related the CDF of \( R_1 \) to the probability \( P_0 \) that all (vertical and horizontal) lines intersecting the Manhattan square around the typical intersection do not contain any POI within the square.

In this paper, we generalize the calculation of the CDF of \( R_k \) for \( k \geq 1 \). The main challenge is to construct a formula, which computes the probabilities \( \{ P_j : j \leq k - 1 \} \) that all lines intersecting the Manhattan square contain exactly \( j \geq 0 \) POI inside the square. The analysis presented in this paper is therefore not a direct extension of [33], where the calculations are limited to \( P_0 \), which can be derived using a null probability argument. While a simple closed-form expression has been derived in [33, Theorem 1] for \( P_0 \), this is not possible for
{P_j : j > 0}, as we will demonstrate in Section V. Instead, we present a low-complexity numerical algorithm to evaluate \( P_k \) using the integer partition function \( p(k) \), which is then used to calculate the CDF of \( \hat{R}_k \).

In Section III, we calculate a closed-form expression for the probability distribution function (PDF) of the total length of line segments inside the Manhattan square, and in Section IV, we compute its moment generating function (MGF). Unfortunately, neither of the expressions can be finally averaged out to obtain a closed-form representation for the probabilities \( \{P_j : j > 0\} \). Nevertheless, these results might be of interest for other applications utilizing the MPLCP. Additionally, they both are intuitive ways to tackle the subject problem, and because of that we have decided to include them in the paper.

We will follow a pedagogical approach where each method occupies a separate section.

Finally, it is noted that the distance between a random intersection and the \( k \)-th nearest point of the MPLCP can serve as a lower bound to the CDF of the distance between a random position of the road and the \( k \)-th nearest neighbor of the MPLCP. For \( k = 1 \) both CDFs are computed in [33].

The remainder of this paper is organized as follows. Section II formally introduces the MPLCP. Section III calculates the PDF of the total length \( L_i \) of line segments inside a Manhattan square, centered at a randomly selected road intersection. In Section IV, we calculate the MGF of the RV \( L_i \), and in Section V we present a numerical algorithm, which can be used to compute the CDF of the path distance between an intersection and the \( k \)-th nearest point of the MPLCP. In Section VI, we validate the suggested algorithm against simulations. We also use a real urban road network and demonstrate improved performance against baseline models using the PPP. Section VII concludes this work.

II. SYSTEM MODEL AND NOTATION

In this section, we first define the MPLP and the MPLCP, then present the construction of the networks for the roads and facilities, along with the adopted assumptions, and finally we conclude this section with the problem formulation.

A. Preliminaries of point and line processes

A line process, in layman’s terms, is just a random collection of lines. If we limit our attention to undirected lines in the Euclidean plane, each line \( \ell_i \) can be uniquely determined by the following parameters: the length \( \rho_i \geq 0 \) and the angle \( \phi_i \in [-\pi/2, \pi/2] \), measured counter-clockwise, of the line segment being perpendicular to line \( \ell_i \) and passing through the origin [11, Chapter 8.2.2]. Therefore a line process can be associated with a point process, and vice versa, where the line \( \ell_i \) is uniquely mapped to the point \( x_i \in \mathbb{R}^2 \) with polar coordinates \( (\rho_i, \phi_i) \). The associated point process is often called the representation space of the line process.

Let us consider the realizations of two independent 1D PPPs of equal intensity \( \lambda \) modeling the road intersections along the \( x \) and \( y \) axes, and construct the associated realization of vertical and horizontal lines. All points on the \( x \) axis give rise to vertical lines \( \phi_i \in \{0, \pi\} \), and all points along the \( y \) axis correspond to horizontal lines \( \phi_i \in \{-\pi/2, \pi/2\} \). The set of horizontal lines is denoted by \( \Phi_{lh} = \{L_{h1}, L_{h2}, \ldots\} \), the set of vertical lines is denoted by \( \Phi_{lv} = \{L_{v1}, L_{v2}, \ldots\} \), and \( \Phi_l = \{\Phi_{lh}, \Phi_{lv}\} \) is the resulting MPLP. The \( \Phi_l \) is a stationary and motion-invariant line process owing to the stationarity and motion-invariance of the PPP in the representation space. Its intensity, defined as the mean total length of lines per unit area, is equal to \( 2\lambda \) [11, Chapter 8.1].

Let us assume that along each line there are facilities, e.g., gas stations, whose locations follow another 1D PPP of intensity \( \lambda_g \). Conditioned on the realization of the line process, the locations of facilities are independent of each other. Under these assumptions, the distribution of facilities becomes a stationary Cox point process, denoted by \( \Phi_g \), and driven by \( \Phi_l \). A Cox point process is in general a doubly-stochastic PPP where the intensity measure is itself random, and it is subsequently constructed in a two-step random mechanism. See [34, Chapters 3 and 4] for an introduction. The intensity of the MPLCP is \( 2\lambda \lambda_g \) [11].

Owing to the stationarity of the MPLP, we can add an intersection point at the origin \( o \) of the \( x\)-\( y \) plane, and two (undirected) lines \( L_x, L_y \) passing through it and aligned with the \( x \) and \( y \) axis respectively. Under Palm probability, the resulting line process becomes \( \Phi_L = \{\Phi_l \cup \{L_x, L_y\}\} \), and the point process of facilities is the superposition of the point process \( \Phi_g \) and the two PPPs of intensity \( \lambda_g \) along \( L_x \) and \( L_y \). See Fig. 1 for an illustration.

B. Modeling of road networks and facilities

In order to generate a realization of the road network, road intersection points are first generated along the typical lines \( L_x \) and \( L_y \) according to a PPP of intensity \( \lambda \). Then, the intersection points along \( L_x / L_y \) give rise to a realization of the line process \( \Phi_{lv} / \Phi_{lh} \). Every line represents a road which is assumed bi-directional. In the subject problem, the width of the road can be safely ignored, because it is expected to be negligible as compared to the expected distance between neighboring road intersections \( \lambda^{-1} \). Note that it is straightforward to extend the calculations for different intensities.
The CDF of the RV $B(r)$

The probability there are $k$ facilities within $B(r)$

$R_k$ The RV describing the Manhattan distance of the $k$-th nearest facility to the origin

$P_0$ A Poisson RV with parameter $z$

$F_X(x)$ The CDF of the RV $X$ with realization $x$

$\mathbb{P}()$ Probability that an event occurs

$\lambda_h, \lambda_v$ between vertical and horizontal streets. Consider, for instance, few main vertical streets traversing the city and many horizontal side streets. Unless otherwise stated, we will use $\lambda_h = \lambda_v = \lambda$ for presentation clarity. In order to generate a random realization for the network of facilities, a PPP of intensity $\lambda_0$ is generated along each line including the typical lines $L_x$ and $L_y$. Naturally, the locations of the facilities are constrained along the road network.

**C. Problem formulation**

Due to the Slivnyak’s Theorem [11, Chapter 8.2], the distance distribution of a randomly selected intersection of $\Phi_i$ and its $k$-th nearest facility, is equal to the distribution between the origin $o$ of the augmented grid $\Phi_L$ and its $k$-th nearest facility. Let us denote by $R_k$ the RV describing the path distance of the $k$-th nearest facility of the augmented grid $\Phi_L$ to the origin $o$. The CDF can be read as $F_{R_i}(r) = \mathbb{P}(R_k \leq r) = 1 - \mathbb{P}(R_k > r)$. The complementary CDF, $\mathbb{P}(R_k > r)$, is equal to the sum of the probabilities $P_j$, $j \in \{0, 1, 2, \ldots (k-1)\}$, where $P_j$ is the probability that there are exactly $j$ facilities inside the Manhattan square, which is the locus of points with Manhattan distance $r$ to the origin, see Fig. 1. Hence,

$$F_{R_k}(r) = 1 - \sum_{j=0}^{k-1} P_j. \ (1)$$

In this paper, we will derive the $F_{R_k}$ for $k \geq 1$. The CDF of the RV $R_1$ has been already derived in [33, Theorem 1]:

$$F_{R_1}(r) = 1 - P_0 = 1 - e^{-4\lambda r} e^{-4\lambda r(1-a_0)}, \ (2)$$

where $a_0 = \frac{1- e^{-2\lambda r}}{2\lambda r}$.

**D. Notation**

The list of frequently used symbols is included in Table I. In addition, the RV $N_y(\Phi \cap B)$ counts the number of points of the Cox process driven by the line process $\Phi$ within $B$. The total number of lines intersecting $B$, excluding the typical lines $L_x, L_y$ is denoted by the RV $N$. The RV $L_i \geq 0$ describes the random length of the line $\ell_i$ intersecting $B$, and the RV $L_i \geq 4r$ describes the total length of segments in $B$ including the contribution, $4r$, due to the typical lines. Finally, the realizations of the RVs $L_i$ and $L_i$ are denoted by $l$.

**III. Calculating $P_k$ using the distribution of $L_i$**

Given the realization $l$ of the RV $L_i$, the number of facilities in $B$ follows a Poisson distribution with parameter $\lambda_0 l$. As a result, based on the law of total expectation, the probability $P_k$ that there are $k$ facilities in $B$ can be obtained by averaging the Poisson distribution, $\text{Po}(\lambda_0 l)$, over the PDF, $f_{L_i}(l)$, of the total length of line segments $L_i$ in $B$. Hence,

$$P_k = \int_{4r}^{\infty} e^{-\lambda_0 l} \left(\frac{\lambda_0 l}{k!}\right)^k f_{L_i}(l) \, dl, \ (3)$$

where the lower integration limit equals $4r$, because $B$ always contains the segments due to the typical lines $L_x, L_y$.

To derive the PDF $f_{L_i}(l)$, we start with the random number $N$ of line segments intersecting $B$, which follows the Poisson distribution with parameter $4\lambda r$. Recall that $\lambda$ is the density of intersection points along a line, and $2r$ is the length of the diagonal of $B$. The abscissas (ordinates) of the line segments parallel to $L_x$ or $L_y$ are distributed uniformly at random in $(-r, r)$. As a result, the distribution of the RV $L_i$ describing the length of the $i$-th line segment in $B$ is uniform too. In order to derive its CDF, we note that $L_i$ takes values in $(0, 2r)$ with equal probability and thus, $\mathbb{P}(L_i \leq l) = \frac{l}{2r}, l \in (0, 2r)$. For instance, the vertical line with abscissa $z_1$ in Fig. 1 has length $l = 2(r - z_1)$ in $B$.

Conditionally on the realization $n \geq 1$ for the RV $N$, the total length of line segments in $B$, $\sum_{i=1}^{n} L_i$, is equal to the sum of $n$ independent and identically distributed (i.i.d.) uniform RVs in $(0, 2r)$, As a result, the sum $\sum_{i=1}^{n} L_i$ follows the Irwin-Hall distribution with PDF

$$\sum_{i=1}^{n} L_i \sim \frac{1}{2r(n-1)!} \sum_{k=0}^{\lfloor \frac{n}{2} \rfloor} (-1)^k \binom{n}{k} \left(\frac{l}{2r} - k\right)^{n-1}, \ (4)$$

where $n \geq 1$ and $l \geq 0$.

In order to compute the PDF of $L_i$, we need to average equation (4) over the Poisson distributed number $N$ for $n \geq 1$. The case $N = 0$, i.e., no intersections along $L_x \cup L_y$ in $B$, which occurs with probability $e^{-4\lambda r}$, leads to $L_i = 4r$ and it is treated separately below.

$$f_{L_i}(l) = e^{-4\lambda r} \delta_{14r} +$$

$$\mathbb{E}_{n \geq 1} \left\{ \frac{1}{2r(n-1)!} \sum_{k=0}^{\lfloor \frac{n}{2} \rfloor} (-1)^k \binom{n}{k} \left(\frac{l-4r}{2r} - k\right)^{n-1} \right\} \ \ (5)$$

$$= e^{-4\lambda r} \delta_{14r} + \sum_{n=1}^{\infty} \frac{(4\lambda r)^n e^{-4\lambda r}}{n!} \frac{1}{2r(n-1)!} \times \left[ \frac{l-4r}{2r} \sum_{k=0}^{\lfloor \frac{n}{2} \rfloor} (-1)^k \binom{n}{k} \left(\frac{l-4r}{2r} - k\right)^{n-1}, l \geq 4r, \right.$$
carefully setting the lower limit of the sum with respect to \( n \).  

\[
\begin{align*}
  f_{L_t}(l) &= e^{-4\lambda r} \delta_{l,4r} + \frac{-4\lambda r}{2} \sum_{n=\max(1,k)}^\infty \left( \frac{4\lambda r}{n!} \right)^n \left( \frac{1}{(n-1)!} \right) k \left( \frac{1}{2} - k \right)^{n-1} \\
  &= e^{-4\lambda r} \delta_{l,4r} + \frac{-4\lambda r}{2} \left( 4\lambda r \cdot F_1(2,2\lambda (l - 4r)) + \sum_{k=1}^{\infty} \left( -1 \right)^k \frac{(4\lambda r)^n}{n!} \frac{1}{(n-1)!} \left( \frac{1}{2} - k \right)^{n-1} \right) \\
  &= e^{-4\lambda r} \delta_{l,4r} + \frac{-4\lambda r}{2} \left( 4\lambda r \cdot F_1(2,2\lambda (l - 4r)) + \sum_{k=1}^{\infty} \left( -1 \right)^k \frac{(4\lambda r)^n}{n!} \frac{1}{(n-1)!} \left( \frac{1}{2} - k \right)^{n-1} \right),
\end{align*}
\]

where \( l \geq 4r \) and \( F_1(\alpha,z) = \sum_{k=0}^\infty \frac{1}{k!} \frac{(-z)^k}{(\alpha - 1)_k} \) is the regularized hypergeometric function; example validations of equation (6) are depicted in Fig. 2.

The final expression in equation (6) is quite complicated to use in the integral in (3), hence, calling for another approach to evaluate the probabilities \( P_k \).

**IV. Calculating \( P_k \) using MGFs**

Since the PDF of the RV \( L_t \) has a complicated form, we may instead work with its MGF, \( M_{L_t}(t) = \mathbb{E}\{e^{tL_t}\}, t \in \mathbb{R} \), which can be computed using the properties of the compound Poisson distribution. Recall that the RV \( L_t \) is equal to the sum of \( N \sim \text{Po}(4\lambda r) \) i.i.d. uniform RVs \( L_i \equiv L_i \) in \((0, 2r)\) plus the constant \( 4r \). Therefore,

\[
\begin{align*}
  \mathbb{E}\{e^{tL_t}\} &= \mathbb{E}_N\{\mathbb{E}\{e^{tL_t|N}\}\} \\
  &= \mathbb{E}_N\{e^{4rt}M_{L_t}(t)^N\} \\
  &\overset{(a)}{=} e^{4rt} \mathbb{E}_N\left( \left( \frac{e^{2\lambda r - 1}}{2r} \right)^N \right) \\
  &\overset{(b)}{=} \exp\left( 4rt + 4\lambda r \left( \frac{e^{2\lambda r - 1}}{2r} - 1 \right) \right),
\end{align*}
\]

where \((a)\) follows from the MGF of a uniform RV, and \((b)\) uses the probability generating function of a Poisson RV.

The limit of the first derivative of \( M_{L_t}(t) \) with respect to \( t \) at \( t \to 0^+ \) in equation (7) yields \((4r + 4\lambda r^2)\), which is the mean of the RV \( L_t \). The first term, \( 4r \), is the fixed length of the typical lines in \( B \). The second term, \( 4\lambda r^2 \), is, as expected, equal to the product of the mean length \( r \) of a randomly selected line segment \( L_i \) multiplied by the expected number \( 4\lambda r \) of line segments in \( B \). Recall that the expected number of line segments is equal to the expected number of intersections along the typical lines.

Conditionally on the realization of the length \( L_t = l \), the number of facilities in \( B \) is Poisson distributed with parameter \( \lambda y_l \). As a result, using the law of total expectation, the MGF of the (discrete) RV of the number of facilities in \( B \), \( \mathbb{E}_L(\Phi_L \cap B) \), can be read as

\[
M_{N_p(\Phi_L \cap B)}(t) = \mathbb{E}_L\{\mathbb{E}_N\{e^{N_p(\Phi_L \cap B) t}\}\} = \mathbb{E}_L\{e^{N_p(\Phi_L \cap B) t L_t}\} = \mathbb{E}_L\{e^{N_p(\Phi_L \cap B) t L_t}\} = \mathbb{E}_L\{e^{N_p(\Phi_L \cap B) t L_t}\} = M_{L_t}(\lambda y_l (e^{t - 1})) ,
\]

where \((a)\) is due to the MGF of a Poisson RV.

After substituting the above argument, \( \lambda y_l (e^{t - 1}) \), into the last equality in (7), we obtain the MGF of the RV \( N_p(\Phi_L \cap B) \). \( M_{N_p(\Phi_L \cap B)}(t) = \exp\left( 4r\lambda y_l (e^{t - 1}) + 4\lambda r \left( \frac{e^{2\lambda r (e^{t - 1}) - 1}}{2r\lambda y_l (e^{t - 1}) - 1} \right) \right) \)

Furthermore, starting from the definition of the MGF of a discrete RV on the natural numbers, \( M_{N_p(\Phi_L \cap B)}(t) = M_{N_p}(t) = \sum_{k=0}^\infty P_k e^{kt} \) we get \( P_k \)

\[
\begin{align*}
  &= \frac{1}{k!} \frac{d^k}{dt^k} \mathbb{E}_N(\log t) \bigg|_{t=0} \\
  &= \frac{1}{k!} \frac{d^k}{dt^k} \exp\left( 4r\lambda y_l (t - 1) + 4\lambda r \left( \frac{e^{2\lambda r (t - 1)} - 1}{2r\lambda y_l (t - 1)} \right) \right) \bigg|_{t=0}.
\end{align*}
\]

After substituting \( k = 0 \) in equation (8), we obtain \( P_0 = \exp(-4\lambda y_l - 4\lambda r (1 - a_0)) \), as expected, see equation (2). For \( k = 1 \) in (8), after some simplification, we have

\[
\begin{align*}
  P_1 &= P_0 \left( 4r\lambda y_l + 4\lambda r (a_0 - e^{-2\lambda r}) \right) .
\end{align*}
\]

The calculation of higher-order derivatives in (8) results in complicated expressions which are difficult to manipulate. For
instance, we list below the expressions we get, after some simplification, for $P_2$ and $P_3$.

\[
P_2 = \frac{1}{2} P_0 \left( 4r_0 \lambda_g + 4r \left( a_0 - e^{-2r \lambda_g} \right) \right)^2 + 4\lambda r P_0 \times \left( a_0 - e^{-2r \lambda_g} - r \lambda_g e^{-2r \lambda_g} \right),
\]

\[
P_3 = \frac{1}{6} P_0 \left( 4r_0 \lambda_g + 4r \left( a_0 - e^{-2r \lambda_g} \right) \right)^3 + 4\lambda r P_0 \times \left( 4r_0 \lambda_g + 4r \left( a_0 - e^{-2r \lambda_g} \right) \times \left( a_0 - e^{-2r \lambda_g} - r \lambda_g e^{-2r \lambda_g} + 4r \lambda P_0 \times \left( a_0 - e^{-2r \lambda_g} - r \lambda_g e^{-2r \lambda_g} - \frac{2r^2 \lambda^2}{e^{-2r \lambda_g}} \right) \right) \right).
\]

(10)

One way to add some structure in the calculation of $P_k$, is to use the Faà di Bruno’s formula for the calculation of the $k$-th derivative of a composite function [35]. Let us define $f(t) = e^t$ and $g(t) = \left( 4r_0 \lambda_g \left( t - 1 \right) + 4\lambda r \left( \frac{e^{2r \lambda_g \left( t - 1 \right)} - 1}{2} \right) \right)$, see equation (8). Leveraging on that $f^{(k)}(t) = e^t$, where $f^{(k)}$ denotes the $k$-th derivative, the Faà di Bruno’s formula is simplified to [35, Eq. (2.2)]:

\[
\frac{d^k f(g(t))}{dt^k} = e^{g(t)} \sum_{m=0}^{k} B_{k,m} \left( g'(t), g''(t), \ldots, g^{(k-m+1)}(t) \right),
\]

(11)

where $B_{k,m}$ and $B_k$ are the partial and the complete exponential Bell polynomials respectively.

The Bell polynomials emerge in set partitions. For instance, $B_{4,2}(x_1, x_2, x_3, x_4) = 3x_2^2 + 4x_1x_3$, indicates that there are three ways to separate the set $\{x_1, x_2, x_3, x_4\}$ into two subsets of size two and four ways to separate it into a block of size three and another of size one. Note that the total number of partitions, i.e., seven, is the Stirling number of second kind which, in general, counts the ways to separate an $m$-element set into $k$ disjoint and non-empty subsets, e.g., $S(4,2) = 7$.

The calculation of the Bell polynomials is widely available in today’s numerical software packages like Mathematica.

Recall from equation (8) that the $k$-th derivative of the MGF has to be evaluated in the limit $t \to 0$. It is also noted that $P_0 = e^{g(0)}$. Combining equations (8) and (11) yields

\[
P_k = P_0 \frac{k!}{k!} \sum_{m=1}^{k} B_{k,m} \left( g'(0), g''(0), \ldots, g^{(k-m+1)}(0) \right).
\]

(12)

Equation (12) is insightful to understand why the calculation of $P_3$ in equation (10) consists of three terms. The first term over there, $(4r_0 \lambda_g + 4r \left( a_0 - e^{-2r \lambda_g} \right))^3$, is essentially equal to the partial Bell polynomial $B_{3,3}(g'(0)) = g'(0)^3$. It is also straightforward to verify that the remaining two terms in (10) are equal to $B_{3,2}(g'(0), g''(0)) = 3g'(0)g''(0)$ and $B_{3,1}(g'(0), g''(0), g'''(0)) = g'''(0)$.

The Faà di Bruno’s formula can provide some insight into the calculation of the MGF of the RV describing the total number of points in $B$, $N_p(\Phi_L \cap B)$, however, the cost of computing the $k$-order partial derivatives might be high, especially for large values of $k$. In the next section, we will use enumerative combinatorics, revealing a simple numerical algorithm to evaluate $P_k$ without involving higher-order derivatives as in equation (12).

V. Calculating $P_k$ Using Integer Partitions

Let us assume there are $k$ facilities inside the Manhattan square $B(r)$ and separate their allocation into two sets: along the typical segments $L_x, L_y$ and in the rest of $B(r)$. The probability $P_k = \mathbb{P}(N_p(\Phi_L \cap B) = k)$ can be read as $P_k = \sum_{i \leq k} \mathbb{P}(N_p(L_x \cup L_y) = i) \cdot \mathbb{P}(N_p(\Phi_I \cap B) = m)$, (13) where $m = k - i$, and the product of probabilities follows from the independent locations of intersections along $L_x$ and $L_y$, and the independent locations of facilities along each line.

Next, we detail the calculation of the two probability terms in (13) followed by a simplified procedure to calculate $P_k$ named after Algorithm 1. We also generalize the calculation to unequal intensities $\lambda_h, \lambda_v$ before concluding this section by presenting a numerical illustration for the computation of $P_k$ for $k = 5$.

A. Calculating $P_k$

The first probability term in (13) can be read as

\[
\mathbb{P}(N_p(\{L_x \cup L_y\} \cap B) = i)
\]

\[
= \mathbb{P}(N_p(L_x \cap B) + N_p(L_y \cap B) = i)
\]

\[
= \frac{(4\lambda r)^i e^{-4\lambda r}}{i!} \mathbb{P}(N_p(\Phi_I \cap B) = m),
\]

(14)

where (a) uses the fact that the superimposed PPPs of facilities along $L_x$ and $L_y$ is another PPP with twice the intensity $2\lambda_y$.

The calculation of the second probability term in (13) is more involved, but similar to (14), it helps to consider just a single PPP of intersection points with twice the intensity, $2\lambda$, along $L_x$ instead of two line processes $\Phi_{Lh}, \Phi_{Lv}$. Let us denote the resulting distribution of vertical lines by $\Phi'_I$. Obviously, $\mathbb{P}(N_p(\Phi_I \cap B) = m) = \mathbb{P}(N_p(\Phi'_I \cap B) = m) = \mathbb{P}(N_p(\Phi'_I \cap B) = m)$. The latter can be written as $\mathbb{P}(N_p(\Phi'_I \cap B) = m)$

\[
= \sum_{n=0}^{\infty} \mathbb{P}(N_p(\Phi'_I \cap B) = m|N = n) \cdot \mathbb{P}(N = n)
\]

\[
= \sum_{n=0}^{\infty} \frac{(4\lambda r)^n e^{-4\lambda r}}{n!} \mathbb{P}(N_p(\Phi'_I \cap B) = m|N = n).
\]

(15)

In order to evaluate the conditional probability in (15), we enumerate the number of ways of allocating $m$ facilities (or objects) into $n$ lines (or urns), with both objects and urns being indistinct. For each allocation, we need to obtain its probability of occurrence, and finally we will sum over all obtained values.

For $n \geq m$, the number of ways to allocate $m$ objects into $n$ urns is equal to the number of integer partitions of $m$, denoted by $p(m)$, because only the number of objects going to each urn is relevant. For $n < m$, the restricted integer partitions of size at most $n$, denoted by $p_n(m)$; are considered. Empty urns are obviously allowed. Next, we sum over the probabilities of all partitions, $\mathbb{P}(N_p(\Phi'_I \cap B) = m|N = n) = \sum_{\xi \in p_n(m)} \mathbb{P}(N_p(\Phi'_I \cap B) = m|N = n, \xi)$, where $p_n(m) = p(m)$ for $n \geq m$ and $\xi$ is the set associated with a partition, e.g., $p(3) = \{3\}, \{2, 1\}, \{1, 1, 1\}$.
After substituting the above equality in the last line of (15), and interchanging the orders of summations, we end up with

$$\mathbb{P}(N_p(F_{i\xi}, B) = m) = \sum_{\xi \in \mathbb{P}(m)} \sum_{n=|\xi|}^{\infty} \frac{(4\lambda r)^n e^{-4\lambda r}}{n!} \mathbb{P}(N_p(F_{i\xi}, B) = m | N = n, \xi),$$

where $|\xi|$ denotes the set cardinality, e.g., $|\xi| = 2$ for $\xi = \{2, 1\}$.

At this point, it helps to define the parameter $a_k, k \in \mathbb{N}$ describing the probability that a vertical line with abscissa $z > 0$, uniformly distributed between the origin and the point $(r, 0)$, contains exactly $k$ facilities in $B$, e.g., in Fig. 1, the line with abscissa $z_1$ does not contain any.

$$a_k = \frac{1}{r} \int_0^r \left(2\lambda g (r - z)\right)^k e^{-2\lambda g (r - z)} dz = \frac{\Gamma(k+1)}{2\lambda g r^k},$$

where $\Gamma(k+1) = k!$, $\Gamma(\alpha, x) = \int_0^x r^{\alpha-1} e^{-r} dr$ is the lower incomplete Gamma function, and for $k = 0$ we obtain the parameter $a_0$ defined under equation (2).

Let us consider the partition $\xi = \{1, 1, \ldots, 1\}$ with $m$ 1’s. The inner sum in equation (16), conditioned on this partition, yields $\mathbb{P}(N_p(F_{i\xi}, B) = m | \xi)$

$$= \sum_{n=m}^{\infty} \frac{(4\lambda r)^n e^{-4\lambda r}}{n!} \mathbb{P}(N_p(F_{i\xi}, B) = m | N = n, \xi) = \sum_{n=m}^{\infty} \frac{(4\lambda r)^n e^{-4\lambda r}}{n!} (\frac{n}{m}) a_m a_0^{n-m} = \frac{(4\lambda r)^m e^{-4\lambda r} (1-a_0)}{m!},$$

where the binomial coefficient $\binom{n}{m}$ represents the number of ways to select the $m$ lines containing just one facility in $B$.

Substituting the above equation and equation (14) into (13) yields the conditional probability, $P_{k|\xi}$, of $k$ facilities in $\{F_{1\xi}, \ldots, F_{k\xi}\} \in B$ given the partition $\xi$.

$$P_{k|\xi} = \frac{\sum_{i=0}^{k} (4\lambda r)^i e^{-4\lambda r} (4\lambda r)^{k-i} a_0^{m-q-i} e^{-4\lambda r} (1-a_0)}{(k-i)!} \quad \text{(18)}$$

In a similar manner, we can evaluate $P_{k|\xi}$ for all $\xi \in p(k)$ and complete the calculation of $P_k = \sum_{\xi \in p(k)} P_{k|\xi}$ in (13).

B. Simplifying the calculation of $P_k$

Evaluating the conditional probabilities $P_{k|\xi}$ as in (18) might be cumbersome, unless a pattern is identified. Next, we derive a simple expression for $P_{k|\xi}$, depending on the number of times an integer appears in the partition $\xi$. Let us consider $\xi = \{q, q, \ldots, q, 1, \ldots, 1\}$, where the $q$ appears $f$ times and there are also $(m-qf)$ 1’s. The inner sum in equation (16), conditioned on the partition $\xi$ with cardinality $|\xi| = (m-qf)+f$, yields $\mathbb{P}(N_p(F_{i\xi}, B) = m | \xi)$

$$= \sum_{n=(m-qf)+f}^{\infty} \frac{(4\lambda r)^n e^{-4\lambda r}}{n!} a_m a_{m-qf} a_0^{n-(m-qf)} \times$$

$$\left(\frac{a_0^{(m-qf)} e^{-4\lambda r} (1-a_0)}{(m-qf)!} \cdot \frac{a_0^{(m-qf)} e^{-4\lambda r} (1-a_0)}{(m-qf)!}\right) \cdot \frac{\Gamma(k+1)}{2\lambda g r^k},$$

where $|\xi| = (m-(q-1)f)$ is the number of lines containing facilities in $B$. They are selected with $\binom{n}{m-(q-1)f}$ ways from the available $n$ lines, and $\binom{m-(q-1)f}{k}$ is the number of ways to select $f$ out of the segments containing facilities in $B$, and allocate $q$ facilities to each one of them.

Keeping in mind that due to the existence of $f$ replicas of the integer $q$ in the partition, only up to $(k-qf)$ facilities may be located along $L_x$, we substitute the above equation and (14) into (13), ending up with

$$P_{k|\xi} = \frac{\sum_{i=0}^{k-qf} (4\lambda r)^i e^{-4\lambda r} (4\lambda r)^{k-i} a_0^{m-qf} a_0^{k-i-qf} \times}{{(k-i-qf)!}} \frac{(4\lambda r)^k e^{-4\lambda r} (1-a_0)}{(k-fq)!} \frac{(4\lambda r a_0)^f}{f!} \frac{\Gamma(k+1)}{2\lambda g r^k} \frac{P_0}{f!}.$$

It is straightforward to generalize the above calculation to include partitions with more than one $q > 1$. Let us assume that the integer $q_i > 1$ appears $f_i \geq 1$ times in the partition $\xi$. Equation (19) can be generalized as

$$P_{k|\xi} = \frac{\sum_{k=1}^{f_i} (4\lambda r g_i + \lambda a_0)^k \sum_{i=0}^{f_i} f_i! \frac{(4\lambda r a_0)^f}{f!} \frac{\Gamma(k+1)}{2\lambda g r^k} \frac{P_0}{f!} \prod_{q_i \neq 1} \frac{(4\lambda r a_0)^f}{f!} \frac{\Gamma(k+1)}{2\lambda g r^k} \frac{P_0}{f!}. \quad \text{(20)}$$

To sum up, in order to evaluate $P_{k|\xi}$, we start with $P_{k|\xi} = P_0$. Given that the integer $q > 1$ appears in the partition $f \geq 1$ times, we set $P_{k|\xi} = \frac{P_{k|\xi} (4\lambda r a_0)^f}{f!} \frac{\Gamma(k+1)}{2\lambda g r^k} \frac{P_0}{f!}$. For $q = 1$, we set $P_{k|\xi} = \frac{P_{k|\xi} (4\lambda r g + \lambda a_0)^f}{f!} \frac{\Gamma(k+1)}{2\lambda g r^k} \frac{P_0}{f!}$. We update $P_{k|\xi}$ for all integers $q \in \xi$. Next, we repeat the same procedure for all partitions $\xi$, and we compute the probability $P_k = \sum_{\xi \in p(k)} P_{k|\xi}$. For completeness, the pseudocode used to calculate $P_k$ is provided as Algorithm 1. Given $P_f, \forall f \in \{0, 1, \ldots, (k-1)\}$, the CDF of the distance distribution is evaluated from (1).

Algorithm 1 Compute $P_k$

Input: $\lambda_0, \lambda, k$

Output: $P_k$

1: $a_q \leftarrow \frac{\Gamma(q+1, 2\lambda_0 r)}{2\lambda_0 r}$ for $q = 0, 1, \ldots, (k-1)$

2: $P_0 = e^{-4\lambda g} - \frac{4\lambda g}{r} \mathbf{I}$

3: $\Xi = \text{IntegerPartitions}(k)$

4: for all $\xi \in \Xi$ do

5: $P_{k|\xi} \leftarrow P_0$

6: for all $q \in \xi$ do

7: $f_q \leftarrow \text{card}(q)$

8: if $q = 1$ then

9: $P_{k|\xi} \leftarrow P_{k|\xi} \frac{(4\lambda g a_0 + \lambda a_0)^f}{f!} \frac{\Gamma(k+1)}{2\lambda g r^k} \frac{P_0}{f!}$

10: else

11: $P_{k|\xi} \leftarrow P_{k|\xi} \frac{(4\lambda g + \lambda a_0)^f}{f!} \frac{\Gamma(k+1)}{2\lambda g r^k} \frac{P_0}{f!}$

12: end if

13: end for

14: $P_k \leftarrow P_k + P_{k|\xi}$

15: end for

C. Complexity analysis

The computational complexity for evaluating the parameter $P_k$ using Algorithm 1 is proportional to the number of integer partitions of $k$. The growth of the integer partition function
Fig. 3. The time required to generate $P_k$ using Algorithm 1 is fitted to the function $f(k) = \frac{1}{c} e^{d/k}$ for positive constants $c$ and $d$. The red markers include also the time required to generate the partitions $p(k)$.

$p(k)$ is known only asymptotically, i.e., $p(k) \sim \frac{1}{k\sqrt{2\pi}\lambda^2} e^{-k/\lambda^2}$ as $k \to \infty$ [36]. In Fig. 3, it is illustrated that the asymptotic formula, with fitted preconstants using least-squares, gives a good approximation for $k \geq 20$.

D. Unequal intensities $\lambda_v, \lambda_h$

Thus far, we have considered equal intensity of intersections $\lambda_v = \lambda_h = \lambda$ along the typical lines $L_x, L_y$. The calculations in Algorithm 1 can be easily generalized to $\lambda_v \neq \lambda_h$. Specifically, one will have to use

$$P_0 \leftarrow e^{-4\lambda_r r - 2(\lambda_v + \lambda_h)r(1 - a_0)},$$

in line 2, and

$$P_{k|\xi} \leftarrow P_{k|\xi} \frac{(2(2\lambda_v + (\lambda_v + \lambda_h)a_0)r_0)^{a_0}}{f_0^{a_0}},$$

$$P_{k|\xi} \leftarrow P_{k|\xi} \frac{(2(\lambda_v + \lambda_h)a_2 r)^{a_2}}{f_2^{a_2}},$$

in lines 9 and 11 respectively. Also, the intensity of intersections for the full Manhattan grid can be read as $\lambda = \frac{2\lambda_v \lambda_h}{\lambda_v + \lambda_h}$. We will have to use unequal intensities once Algorithm 1 is evaluated on a real urban map.

E. Partitioning example

For illustration purposes, in Table II, we list the contributions of the seven different terms involved in the calculation of $P_3$. The inputs in the rightmost column, which is equal to the product of the terms in the middle column, are generated based on equation (20). For instance, for the partition $\{3, 2\}$, we have $f_1 = f_2 = 1$, because each of the numbers $q_1 = 2$, $q_2 = 3$ appears only once in the partition. Furthermore, $(f_1 q_1 + f_2 q_2 = 5)$ and the exponent of the term $(\lambda_g + \lambda_{a1})$ is zero. Therefore, equation (20) degenerates to the product of just two terms, $4\lambda_{a2}$ and $4\lambda_{a3}$, scaled by $P_0$, and the result for the probability $P_{\xi}(3, 2)$ directly follows.

Now, it becomes clear in the calculation of $P_3$ in equation (10) that the first term corresponds to the partition $\{1, 1, 1\}$ with $a_1 = a_0 - e^{-2\lambda_r}$, the second term to the partition $\{2, 1\}$ with $a_2 = a_0 - e^{-2\lambda_g} - r\lambda_g e^{-2\lambda_g}$, and the last term to the partition $\{3\}$ with $a_3 = \frac{1}{2} \left( a_0 - e^{-2\lambda_r} - r\lambda_g e^{-2\lambda_g} - \frac{2}{3} \pi^2 \lambda_g^2 e^{-2\lambda_g} \right)$.

VI. NUMERICAL ILLUSTRATIONS & APPLICATIONS

In this section, we will start by validating Algorithm 1 with simulations, and we will proceed with the study of three example scenarios where this algorithm might be of use. Finally, we will show that Algorithm 1 can give a quick and accurate estimate about the distance distributions in real urban road networks with an approximate regular street layout. For that, we will use the map of an area near Manhattan in New York with the map data retrieved using OpenStreetMap [37], [38]. In Fig. 4 we have validated the calculation of the path distance distribution using Algorithm 1 for the ten nearest neighbors of a MPLCP. Fig. 5 and Fig. 6 illustrate that the Euclidean distance (L2 norm) is a bad approximation to the Manhattan distance (L1 norm) distribution, unless the deployment scenario is associated with a sparse road network and densely populated facilities. The approximation quality deteriorates for larger k’s. The planar PPP of equal intensity, $\mu = 2\lambda_\gamma$, where the locations of facilities are not constrained by the road network, is not a better approximation either. The distance to the k-th nearest neighbor for the planar PPP follows
The curves are generated using Algorithm 1 for the Manhattan distance, simulations for the Euclidean distance and equation (23) for the planar PPP.

the generalized gamma distribution with PDF [39, Theorem 1]:

\[ f_{R_k}(r) = \frac{2 e^{-\mu \pi r^2} (\mu \pi r^2)^{k}}{r \Gamma(k)}. \]  

(23)

Having verified that the PPP and the Euclidean distance give inadequate approximations, we present below case studies where the path distance distributions are useful.

### A. Distance distributions in spatial database queries

Let us consider an electric vehicle at an intersection querying for the nearest charging station. The charging stations might be closed or fully occupied, depending on the time of day and the road traffic conditions. In that case, the spatial database should return the nearest available charging station to the vehicle. Given that any facility is available with probability \( q \), independently of other facilities, and the average travel speed is \( v \), the CDF of the average travel time to the nearest available facility follows from the geometric distribution:

\[ P(t \leq \tau) = \sum_{i=1}^{\infty} q (1 - q)^{i-1} F_{R_i}(\tau), \]  

(24)

where \( \tau = rv^{-1} \).

See Fig. 7a for the validation of (24). The underlying assumption is that the delay at the intersection and the traffic-related delays are not modeled explicitly but incorporated into the model through the average velocity \( v \).

### B. Planning the network of facilities in a city

Before starting to build charging stations for electric vehicles in a city, it is important to identify their required density, i.e., the minimum number of stations per square kilometer, so that certain design constraints are satisfied. This process resembles network dimensioning in wireless communications. Given the intensity of roads \( \lambda \), we would like to identify the minimum required intensity of facilities \( \lambda_y \), so that a vehicle at a randomly selected intersection can arrive at the nearest available facility within the target time, e.g., 100 s with probability larger than 90%. Due to the low computational complexity of the path distance distributions using Algorithm 1, we can obtain the required intensity \( \lambda_y \) numerically. To give an example, for the parameter settings used to generate Fig. 7b, the above target can be safely met for \( \lambda_y \geq 1 \text{ km}^{-1} \).

### C. Urban vehicular communication networks

Turning our interest to wireless communications applications, we assume that an RSU is deployed at the typical intersection and the locations of vehicles follow an MPLCP. The RSU broadcasts messages to the vehicles. For wireless propagation along urban street micro cells, the pathloss model is a function of the Manhattan distance [40, Fig. 5]. Also, the vehicles with non-light-of-sight (NLoS) connection suffer from serious diffraction losses due to the propagation of wireless signals around the corner. In Fig. 8 the distribution of the signal-to-noise ratio (SNR) for the 10 nearest vehicles with NLoS connection to the RSU is depicted. Assuming a distance-based propagation pathloss \( r^{-\eta} \), where \( r \) stands for the Manhattan distance, \( \eta \geq 2 \) denotes the pathloss exponent, and a diffraction loss \( L \), it is straightforward to convert the distance distributions into received signal level distributions. Then, it also remains to scale the obtained CDFs by the noise power level \( N_0 \). Specifically, for the \( k \)-th nearest vehicle with NLoS connection we have

\[ P(\text{SNR}_k \leq \theta) = P \left( LR_k^{-\eta} \leq \theta N_0 \right) = 1 - P \left( R_k \leq \left( \theta N_0 / L \right)^{-1/\eta} \right) \]  

(25)

where the vehicles along \( L_x, L_y \) with a line-of-sight connection to the RSU are neglected, hence, in the above equation, equation (20) degenerates to

\[ P_k[i] = e^{-4\lambda (1-a_0)} \prod_{j=1}^{\infty} \frac{(4\lambda r a_j)^{f_i}}{f_i!}. \]

Given the size of the Manhattan cell \( B \), we can compute the distribution of the number of NLoS vehicles within \( B \), i.e., the network load distribution for NLoS connectivity, using Algorithm 1. Since the NLoS vehicles have much lower link gains than the vehicles along the typical lines \( L_x, L_y \), the RSU must allocate to them much more spectral resources under some fair scheduling scheme. Therefore, our ability to quickly characterize the network load distribution for NLoS vehicles, see Fig. 9 for an example illustration, is important. To further justify the importance of network load statistics, the associated CDF of the network load for NLoS vehicles is depicted in the inset. For instance, in the downlink, the CDF indicates, e.g., that the number of NLoS vehicles requesting resource blocks exceeds 10 with probability less than 10%.
The $k$-th nearest neighbor distance distributions, $k \in \{1, 10\}$, from a random intersection of a MPLCP using the Manhattan (blue lines) and the Euclidean (red lines) distance. The distance distributions for a planar PPP of equal intensity are also demonstrated (black lines). Dense roads (DR) correspond to $\lambda = 10 \text{ km}^{-1}$ and sparse roads (SR) to $\lambda = 1 \text{ km}^{-1}$. Dense facilities (DF) are associated with $\lambda_g = 4 \text{ km}^{-1}$ and sparse facilities with $\lambda_g = 0.5 \text{ km}^{-1}$. For sparse roads and dense facilities (SR-DF) the nearest neighbor is mostly located along the typical lines, hence, the L2 norm degenerates to the L1 norm. Because of that, the 1D PPP with intensity $2\lambda_g$ provides also a good fit for $k = 1$, see the green dashed line.

Fig. 8. The SNR distribution at the $k$ nearest vehicles $k \leq 10$ with NLoS connection to the RSU. The locations of vehicles follow a MPLCP with $\lambda = 5 \text{ km}^{-1}$ and $\lambda_g = 10 \text{ km}^{-1}$. Distance-based pathloss $r^{-\eta}$ with $\eta = 3$, diffraction loss around the corner 20 dB and noise power level $N_0 = 10^{-8}$. The simulations are depicted in ‘red’ and the (exact) calculations in ‘blue’.

D. Model validation with a real map

In this section, we investigate the effectiveness of the proposed approach in a practical setting, where the layout of the road network does not precisely follow the MPLP. For...
this purpose, we have extracted the road network of an urban area of size $1.8 \text{ km}^2$ using [37], [38], and selected 15 points, forming approximately a grid, where the distance distributions of the shortest paths are simulated. The map of the area can be seen in Fig. 10a and the associated road network is extracted in Fig. 10b. The road network consists of approximately 500 linear segments. We assume that the locations of facilities follow a one-dimensional PPP of intensity $\lambda_g$ along each segment. Therefore, the only difference between the model and the studied scenario is the underlying road layout. In order to parameterize Algorithm 1, we have estimated the intensity of intersections $\lambda_h \approx 5.9 \text{ km}^{-1}$ equal to the average of horizontal distances between neighboring points across the rows of the 15-point grid. Similarly, the intensity $\lambda_v \approx 12.5 \text{ km}^{-1}$ is the average of vertical distances across the columns of the grid. Then, the intensity of intersections $\lambda$, which is the input to Algorithm 1, is calculated based on the harmonic average of $\lambda_h$ and $\lambda_v$, i.e., $\lambda = \frac{2\lambda_h \lambda_v}{\lambda_h + \lambda_v}$, yielding $\lambda \approx 7.9 \text{ km}^{-1}$. The optimal setting of the intensity parameters given the underlying road infrastructure has been left as a future research topic.

In Fig. 11a, we can see that the performance accuracy of Algorithm 1 improves for larger $k$, while, on the other hand, the prediction accuracy of the two-dimensional PPP degrades. In addition, increasing the density of facilities to $\lambda_g = 1 \text{ km}^{-1}$, compare Fig. 11a with Fig. 11b for the same value of $k$, is associated with worse model performance. That is because, for denser facilities, the $k$ nearest neighbors are likely to come closer to the intersection point. As a result, the local characteristics of the road network near the intersection, which do not follow exactly the MPLP model, start to affect more.

We conclude this section by evaluating the model performance for a very low density of facilities. In this case, we use only the intersection point at the top-right corner of the grid located near the center of the total area. We can see in Fig. 12 that the model performance further improves. The facilities spread out far from the intersection point, hence, the underlying road structure starts to have less effect on the path distance separation. It is mostly the vertical and horizontal distance to the intersection that matter. That is a promising finding because it indicates that, for sparse facilities, the distance distributions derived using the MPLCP are fairly valid independently of the underlying urban road layout. However, further data collection is required to ascertain this claim.
Fig. 12. Path distance distributions for the intersection point at the top-right corner of the grid of the 15 intersection points, see Fig. 10a. Intensity of facilities $\lambda_g = 0.25 \text{km}^{-1}$. For the rest of the parameter settings, see the caption of the previous figure.

VII. CONCLUSIONS

In this paper, we have devised a low-complexity numerical algorithm to calculate the distribution of the path distance between a randomly selected road intersection and the $k$-th nearest node of a Cox point process driven by the Manhattan Poisson line process. This algorithm can be used to identify the minimum required density of facilities (modeled by a MPLCP), e.g., charge stations for electric vehicles, to ensure that a vehicle at an intersection can reach the nearest available facility within a target time under a probability constraint. The distance distributions derived in this paper can also be used to calculate the distribution of network load for V2X systems. Finally, using real road network data, we illustrated the enhanced performance of the MPLP as compared to the state-of-art distance distribution model using the two-dimensional PPP.

In the future, it would be interesting to extend our analysis and data collection with larger maps and different urban areas, investigating the independence of the path distance distributions from the underlying road infrastructure, when the network of facilities is sparse.

ACKNOWLEDGMENT

K. Koufos and M. Dianati were partially supported by Jaguar Land Rover and the United Kingdom-Engineering and Physical Sciences Research Council (UK-EPSRC) as part of the jointly funded Towards Autonomy: Smart and Connected Control (TASCC) Programme under Grant EP/N01300X/1 (CARMA). H.S. Dhillon’s work was supported in part by the US National Science Foundation (Grant IIS-1633363).

REFERENCES


Professor Mehrdad Dianati is the Head of Intelligent Vehicles Research Department and the Technical Research Lead in the area of Networked Intelligent Systems at the Warwick Manufacturing Group (WMG), the University of Warwick, UK. The focus of his research is on the application of Digital Technologies (Information and Communication Technologies and Artificial Intelligent) for the development of future mobility and transport systems. He has over 29 years of combined industrial and academic experience, with 20 years in various leadership roles in multi-disciplinary collaborative R&D projects. He works closely with the Automotive and ICT industries as the primary application domains of his research. He is also the Director of Warwick’s Centre for Doctoral Training on Future Mobility Technologies, training doctoral researchers in the areas of intelligent and electrified mobility systems in collaboration with the experts in the field of electrification from the Department of Engineering of the University of Warwick. In the past, he has served as an Editor for the IEEE Transactions on Vehicular Technology and several other international journals, including IET Communications. Currently, he is the Field Chief Editor of Frontiers in Future Transportation.

Konstantinos Koufos obtained the diploma in electrical and computer engineering from Aristotle University of Thessaloniki, Greece, in 2003, and the M.Sc. and D.Sc. in radio communications from Aalto University, Finland, in 2007 and 2013. After working as a post-doctoral researcher in Aalto University, and as a senior research associate in spatially embedded networks in the School of Mathematics in the University of Bristol, U.K., he is currently a senior research fellow of cooperative autonomy within the Warwick Manufacturing Group (WMG) in the University of Warwick, U.K. His current research interests include stochastic geometry, interference and mobility modeling for 5G.

Harpreet S. Dhillon (S’11–M’13–SM’19) received the B.Tech. degree in electronics and communication engineering from IIT Guwahati in 2008, the M.S. degree in electrical engineering from Virginia Tech in 2010, and the Ph.D. degree in electrical engineering from the University of Texas at Austin in 2013. After serving as a Viterbi Postdoctoral Fellow at the University of Southern California for a year, he joined Virginia Tech in 2014, where he is currently an Associate Professor of electrical and computer engineering and the Elizabeth and James E. Turner Jr. ’56 Faculty Fellow. His research interests include communication theory, wireless networks, stochastic geometry, and machine learning. He is a Clarivate Analytics Highly Cited Researcher and has coauthored five best paper award recipients including the 2014 IEEE Leonard G. Abraham Prize, the 2015 IEEE ComSoc Young Author Best Paper Award, and the 2016 IEEE Heinrich Hertz Award. In 2020, he received Early Achievement Awards from the IEEE Communication Theory Technical Committee (CTTC) and the IEEE Radio Communications Committee (RCC). He was named the 2017 Outstanding New Assistant Professor, the 2018 Steven O. Lane Junior Faculty Fellow, the 2018 College of Engineering Faculty Fellow, and the recipient of the 2020 Dean’s Award for Excellence in Research by Virginia Tech. His other academic honors include the 2008 Agilent Engineering and Technology Award, the UT Austin MCD Fellowship, the 2013 UT Austin WNCG leadership award, and the inaugural IIT Guwahati Young Alumni Achiever Award 2020. He currently serves as a Senior Editor for the IEEE WIRELESS COMMUNICATIONS LETTERS and an Editor for the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS and the IEEE TRANSACTIONS ON GREEN COMMUNICATIONS AND NETWORKING.